CODEN: JASMAN

The Journal of the Acoustical Society of America

ISSN: 0001-4966

Vol. 117, No. 3 Pt. 1 March 2005

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 117, NO. 3, MARCH 2005

CONTENTS—*Continued from preceding page*

ULTRASONICS, QUANTUM ACOUSTICS, AND PHYSICAL EFFECTS OF SOUND [35]

STRUCTURAL ACOUSTICS AND VIBRATION [40]

NOISE: ITS EFFECTS AND CONTROL [50]

ARCHITECTURAL ACOUSTICS [55]

ACOUSTIC SIGNAL PROCESSING [60]

PHYSIOLOGICAL ACOUSTICS [64]

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 117, NO. 3, MARCH 2005

CONTENTS—*Continued from preceding page*

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 117, NO. 3, MARCH 2005

CONTENTS—*Continued from preceding page*

CODEN: JASMAN

The Journal of the Acoustical Society of America

ISSN: 0001-4966

Vol. 117, No. 3, Pt. 2 of 2 March 2005

THE NORTH PACIFIC ACOUSTIC LABORATORY

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 117, NO. 3, MARCH 2005

CONTENTS—*Continued from preceding page*

Subscription Prices, 2005

			Air Freight	
	U.S.A & Poss.	North, Central, & S. America	Europe	Mideast, Africa, Asia & Oceania
ASA Members	(on membership)		$ 125.00	$ 125.00
Regular rate	$1545.00	$1615.00	$1670.00	$1670.00

Single copies: $125.00 each.

Microfiche editions: *The Journal of the Acoustical Society of America* is published simultaneously in microfiche and hard copy. Prices of current microfiche editions are the same as hard copy.

Back Numbers: All back issues of the *Journal* are available on microfiche and some are available on CD ROM. Some, but not all, print issues are also available. Prices will be supplied upon request to Elaine Moran, ASA Office Manager, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502. Telephone: (516) 576-2360; FAX: (516) 576-2377; E-mail: asa@aip.org.

The *Journal* has been indexed for ready reference and twelve cumulative indexes, covering five-year periods, have been published. The Acoustical News—USA section of this issue lists prices and availability.

The Journal of the Acoustical Society of America (ISSN: 0001-4966) is published monthly by the Acoustical Society of America through the American Institute of Physics. POSTMASTER: Send address changes to *The Journal of the Acoustical Society of America,* Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502. Periodicals postage paid at Huntington Station, NY 11746 and additional mailing offices.

Subscription, renewals, and address changes should be addressed to AIP Circulation and Fulfillment Division (CFD), Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502. Allow at least six weeks advance notice. For address changes please send both old and new addresses and, if possible, include a mailing label from a recent issue.

Claims, Single Copy Replacement and Back Volumes: Missing issue requests will be honored only if received within six months of publication date (nine months for Australia and Asia). Single copies of a journal may be ordered and back volumes are available in print or microfilm. Members—contact AIP Member Services at (516) 576-2288; (800) 344-6901. Nonmember subscribers—contact AIP Subscriber Services at (516) 576-2270; (800) 344-6902; E-mail: subs@aip.org.

Page Charge and Reprint Billing: Contact: AIP Publication Page Charge and Reprints—CFD, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; (516) 576-2234; (800) 344-6909.

Document Delivery: Copies of journal articles can be ordered for online delivery from DocumentStore, our online document delivery service (URL: http://documentstore.org/).

Online Availability: Abstracts of journal articles published by the AIP and Member Societies (and several other physics publishers) are available from AIP's SPIN database, via AIP's Online Journal Publishing Service (OJPS) (http://ojps.aip.org).

Microfilm editions of complete volumes of *The Journal of the Acoustical Society of America* are available on 16 and 35 mm. A *Microfilm Catalog* is available on request. This *Journal* is indexed quarterly in *Current Physics Index,* a subject and author index (with abstracts) to all journals published by AIP and its member societies.

ACOUSTICAL NEWS—USA

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

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

Leo L. Beranek Receives Awards

On 6 October 2004, the Institute of Acoustics awarded Leo L. Beranek an Honorary Fellowship with the citation ''Leo Beranek has inspired generations of acousticians through his publications. As co-founder and President of Bolt Beranek and Newman, he created an acoustical consulting company that became synonymous with acoustical excellence throughout the world. For his outstanding contributions to leadership, publications and inspirational work in acoustics, The Institute of Acoustics is proud to award to him an Honorary Fellowship.'' The Institute of Acoustics, with about 2500 members, is the United Kingdom's professional body for those working in acoustics, noise, and vibration. It was formed in 1974 from the amalgamation of the Acoustics Group of the Institute of Physics and the British Acoustical Society.

On 15 November 2004, during the ASME International Mechanical Engineering Congress, the 2004 Per Bruel Gold Medal for Noise Control and Acoustics was awarded to Leo L. Beranek ''for exemplar leadership in formulating and disseminating practical and useful applications of today's most advanced technologies in noise control and acoustics, and for authoring a series of seminal textbooks that provide educators and practitioners with methods of application for these technologies.'' Founded in 1880 as the American Society of Mechanical Engineers, today's ASME is a 120 000-member professional organization focused on technical, educational, and research issues of the engineering and technology community.

USA Meetings Calendar

Listed below is a summary of meetings related to acoustics to be held in the U.S. in the near future. The month/year notation refers to the issue in which a complete meeting announcement appeared.

2005

16–20 May	149th Meeting joint with the Canadian Acoustical Association, Vancouver, Canada [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: http://asa.aip.org].
16–19 May	Society of Automotive Engineering Noise & Vibration Conference, Traverse City, MI [Patti Kreh, SAE International, 755 W. Big Beaver Rd., Ste. 1600, Troy, MI 48084; Tel.: 248-273-2474; E-mail: pkreh@sae.org].
18–22 July	17th International Symposium on Nonlinear Acoustics, State College, PA [Anthony Atchley, The Pennsylvania State University, 217 Applied Research Lab Building, University Park, PA 16802; Tel.: 814-865-6364; E-mail: ISNA17@outreach.psu.edu; WWW: http://www.outreach.psu.edu/c&i/isna17/].
17–21 October	150th Meeting joint with Noise-Con, Minneapolis, MN [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: http://asa.aip.org].

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

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

Some indexes are out of print as noted below.

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

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

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

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

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

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

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

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

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

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

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

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

Volumes 105–114, 1999–2003: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 616. Price: ASA members $50; nonmembers $90 (paperbound).

BOOK REVIEWS

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

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

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

Concert Halls and Opera Houses: Music, Acoustics, and Architecture (Second Edition)

Leo L. Beranek

Springer-Verlag, New York, 2003.
712 pp. Price: $69.00 (hardcover), ISBN: 0387955240.

This is the third in a trilogy of books that document the enormous contributions of Leo Beranek to the understanding of the acoustics of halls for music performance. All three have been written in language understandable not only to acousticians but to musicians, music critics, architects, engineers, administrators and managers of orchestras, and all involved in the decision-making process in the building, renovation, and day-to-day management of music performance facilities. The late Editor-in-Chief of the *Journal of the Acoustical Society of America*, Dan Martin, was effusive in his praise of both previous books: *Music, Acoustics and Architecture* (Wiley, 1962), and *Concert and Opera Halls—How they sound* (Acoustical Society of America, 1996). In his review of the 1996 volume, Martin quotes from his 1962 review: ''No previous treatment of architectural acoustics has approached the level of musical understanding and appreciation that the author (Beranek) has revealed.'' Martin's words have proven prophetic, and no less can be said about this concluding volume of the trilogy.

The 1962 book, *Music, Acoustics, and Architecture* (MAA), provided at least the following four ground-breaking features that were further developed and enhanced in the subsequent 1996 and 2003 books: First, a common vocabulary was established so that acousticians might better communicate with musicians about the complex subjective and objective attributes of a hall's acoustics. Second, Beranek first identified the many interdependent acoustical metrics, beside a hall's reverberance, that go toward a listener's or performer's subjective evaluation of a hall's acoustics. Third (and in this reviewer's judgment, most significantly), Beranek provided acoustical and architectural data on some 54 halls from around the world in a consistent format that permitted meaningful comparative studies by researchers and hall designers alike. Finally, MAA provided model guidelines for evaluating or predicting the acoustical performance of both existing and new halls. Chief among these attributes related to a hall's ''acoustical intimacy,'' initial-time-delay-gap was introduced and remains a significant indicator of a hall's acoustical quality.

The 1996 book, *Concert and Opera Halls—How They Sound* (C&OH-HTS), resulted largely from concerted pressure by Beranek's colleagues to produce a revised and updated version of the 1962 book. In a JASA tutorial ''Concert Hall Acoustics 1992'' [J. Acoust. Soc. Am. **92**, 1–39 (1992)], Beranek assessed the enormous progress that had taken place in the previous three decades since the publication of MAA and included these results in a chapter of C&OH-HTS. Also included were many of the new halls built since the 1960s, bringing the total number of halls to 76. The numerical hall evaluation scales first proposed in MAA were also refined and transformed in new chapters reflecting current research results and applications experience, all of which was continuing at a quickening pace in many countries throughout the world. Beranek was deeply involved in the Japanese research in particular, as well as in the design of new halls in Japan. He, along with his colleagues, produced numerous professional papers on their work in both the Japanese and American Acoustical Societies. When copies of C&OH-HTS were exhausted in less than four years, the author concluded that a new revised edition was appropriate. Thus, the third book in the trilogy, *Concert Halls and Opera Houses: Music, Acoustics and Architecture (Second Edition)*.

A listing of the chapter titles of the new 2003 book reveals that it has all the desirable features of the 1962 and 1996 versions plus some: Chap. 1, ''Music and Acoustics;'' Chap. 2, ''The Language of Acoustics;'' Chap. 3, ''100 Concert Halls and Opera Houses;'' Chap. 4, ''Acoustics of Concert Halls;'' and Chap. 5, ''Acoustics of Opera Houses.'' In addition, there are three useful appendices: Appendix 1, ''Terms, Definitions and Conversion Factors;'' Appendix 2, ''Acoustical Data for Concert Halls and Opera House;'' and Appendix 3, ''Equations, Technical Data and Sound Absorption Coefficients.'' Twenty-four new halls have been added since the 1996 book and, as can be inferred from Table I below, it covers important halls throughout the world. Beranek has maintained his previous rules to include only halls on which he has complete and credible acoustical and physical data, as

TABLE I. Geographical distribution of halls included in studies for 1962, 1996, and 2003 books.

	MAA (1962)	C&OH/HTS (1996)	CH&OH/MAA (2003)
Argentina	1	1	1
Australia	0	1	1
Austria	3	3	4
Belgium	1	1	1
Brazil	0	0	1
Canada	2	3	2
China	0	0	2
Denmark	2	2	2
Finland	2	1	1
France	1	3	3
Germany	8	9	11
Great Britain	8	11	12
Greece	0	0	1
Hungary	0	1	2
Israel	2	2	2
Italy	1	1	2
Japan	0	7	12
Malaysia	0	0	1
Mexico	0	1	1
Netherlands	1	2	2
New Zealand	0	1	1
Norway	0	0	1
Spain	0	0	2
Sweden	1	1	1
Switzerland	3	2	3
Taiwan	0	1	1
USA	17	21	26
Venezuela	1	1	1
Total halls	54	76	100

well as halls that he has personally visited and has heard musical performances in. The fact that all of the architectural drawings are easily scalable facilitates comparative studies and makes the book an invaluable resource for architects and acousticians alike.

As with Beranek's previous two books, *Concert Halls and Opera Houses: Music, Acoustics, and Architecture* is not intended to be read in detail by all who may be involved in the decision-making roles involved with designing, renovating, and managing concert halls and opera houses. Rather, CH&OH:MAA may be likened to an encyclopedia that can be referred to repeatedly as needed. For example, musicians will find Chaps. 1 and 2 of most benefit in helping to understand the relationships between their art form and the forms of performance spaces most suitable to them. They also address the dilemma that a single hall may not always be ideal for all types of music that may be performed there. Architects and hall designers will find Chap. 3 the most valuable, and will probably refer to many of the halls over and over again as their new hall design takes shape. Acousticians and musicians alike will be interested in Beranek's rank ordering of 58 of the halls in Chap. 4 that was developed from interviews and questionnaires from experienced listeners, including well over 100 conductors and music critics.

In all, this book is destined to become the premier reference book for concert halls and opera house acoustics. It belongs in the libraries of everyone involved in their design and use.

WILLIAM J. CAVANAUGH
Natick, MA

Acoustic Absorbers and Diffusers, Theory, design and application

T. J. Cox and P. D'Antonio

Spon Press, London, 2004.
405 pp. Price: $159.95 (hardcover), ISBN: 0-415-29649-8.

After Marshall, Hyde, and Barron first successfully installed number-theoretic (''Schroeder'') diffusors in Wellington Town Hall in New Zealand, Peter D'Antonio was the first entrepreneur to make reflection phase gratings (RPGs) based on quadratic residues widely available. The present volume is the gist of his and his collaborators' experience with these structures and their cousins in the field of architectural acoustics during the last quarter century.

After general introduction, separate chapters focus on basic principles of absorbers and diffusors and the measurement of their properties. Porous and resonant absorbers are discussed in detail, as are random geometric and number-theoretic diffusors both in one and two dimensions. Of particular interest is a final chapter on active absorption and diffusion, as well as hybrid (active–passive) systems.

The book is profusely and exceedingly well illustrated. There are numerous helpful references and several appendices on absorption coefficients (from draperies to wooden pews), seven MATLAB programs, and extensive diffusion-coefficient tables (from semicylinders and triangles to quadratic-residue and primitive-root diffusors).

Diffusor applications to roadside noise barriers, ''street canyons'' as well as recording studios and music practice rooms are likewise treated.

Of particular interest are the authors' fractal diffusors (''diffractals''), which embed high-frequency diffusors within a low-frequency diffusor. I was also intrigued by their 2D hybrid diffusors which make use of perforated masks of binary maximum-length sequences whose period-length can be factored. Thus, for example, $2^{10}-1=1023=31\times 33$. (Such masks have also been used in x-ray astronomy for imaging distant x-ray emitting stars. There seems to be no end to the useful applications of number theory to practical problems.)

The book can be highly recommended to workers in the field as well as scientists interested in a broad range of problems and solutions for acoustical absorbers and diffusors for all kinds of waves (sonar, radar, light).

MANFRED SCHROEDER
University of Goettingen
Drittes Physikalisches Institut
37073 Goettingen, Germany

BOOK REVIEWS

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

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

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

Acoustic Absorbers and Diffusers, Theory, design and application

T. J. Cox and P. D'Antonio

Spon Press, London, 2004.
405 pp. Price: $159.95 (hardcover), ISBN: 0-415-29649-8.

After Marshall, Hyde, and Barron first successfully installed number-theoretic (''Schroeder'') diffusors in Wellington Town Hall in New Zealand, Peter D'Antonio was the first entrepreneur to make reflection phase gratings (RPGs) based on quadratic residues widely available. The present volume is the gist of his and his collaborators' experience with these structures and their cousins in the field of architectural acoustics during the last quarter century.

After general introduction, separate chapters focus on basic principles of absorbers and diffusors and the measurement of their properties. Porous and resonant absorbers are discussed in detail, as are random geometric and number-theoretic diffusors both in one and two dimensions. Of particular interest is a final chapter on active absorption and diffusion, as well as hybrid (active–passive) systems.

The book is profusely and exceedingly well illustrated. There are numerous helpful references and several appendices on absorption coefficients (from draperies to wooden pews), seven MATLAB programs, and extensive diffusion-coefficient tables (from semicylinders and triangles to quadratic-residue and primitive-root diffusors).

Diffusor applications to roadside noise barriers, ''street canyons'' as well as recording studios and music practice rooms are likewise treated.

Of particular interest are the authors' fractal diffusors (''diffractals''), which embed high-frequency diffusors within a low-frequency diffusor. I was also intrigued by their 2D hybrid diffusors which make use of perforated masks of binary maximum-length sequences whose period-length can be factored. Thus, for example, $2^{10}-1=1023=31\times 33$. (Such masks have also been used in x-ray astronomy for imaging distant x-ray emitting stars. There seems to be no end to the useful applications of number theory to practical problems.)

The book can be highly recommended to workers in the field as well as scientists interested in a broad range of problems and solutions for acoustical absorbers and diffusors for all kinds of waves (sonar, radar, light).

MANFRED SCHROEDER
University of Goettingen
Drittes Physikalisches Institut
37073 Goettingen, Germany

 0001-4966/2005/117(1)/988/1/$20.00

REVIEWS OF ACOUSTICAL PATENTS

Lloyd Rice
11222 Flatiron Drive, Lafayette, Colorado 80026

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

Reviewers for this issue:

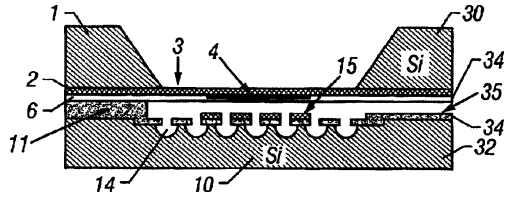
GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
JOHN M. EARGLE, *JME Consulting Corporation, 7034 Macapa Drive, Los Angeles, California 90068*
MARK KAHRS, *Department of Electrical Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania 15261*
DAVID PREVES, *Starkey Laboratories, 6600 Washington Ave. S., Eden Prairie, Minnesota 55344*
ROBERT C. WAAG, *Univ. of Rochester, Department of Electrical and Computer Engineering, Rochester, New York 14627*

6,806,593

43.38.Ar THIN FILM ELECTRET MICROPHONE

Yu-Chong Tai *et al.*, assignors to California Institute of Technology
19 October 2004 (Class 307/400); filed 15 May 2001

The authors describe in detail a specific manufacturing process for MEMS (microelectromechanical system) electret microphones to meet ever-



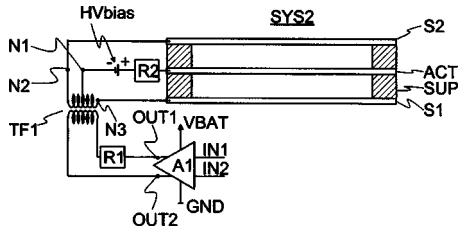
growing needs in today's communications activities. Manufacturers in this field will want to study the patent in detail.—JME

6,785,393

43.38.Bs AUDIO TRANSDUCERS

Markku Lipponen *et al.*, assignors to Nokia Mobile Phones, Limited
31 August 2004 (Class 381/191); filed in Finland 2 December 1999

The illustration shows what appears to be a conventional electrostatic loudspeaker. In this case, however, stator **S2** and diaphragm **ACT** are trans-



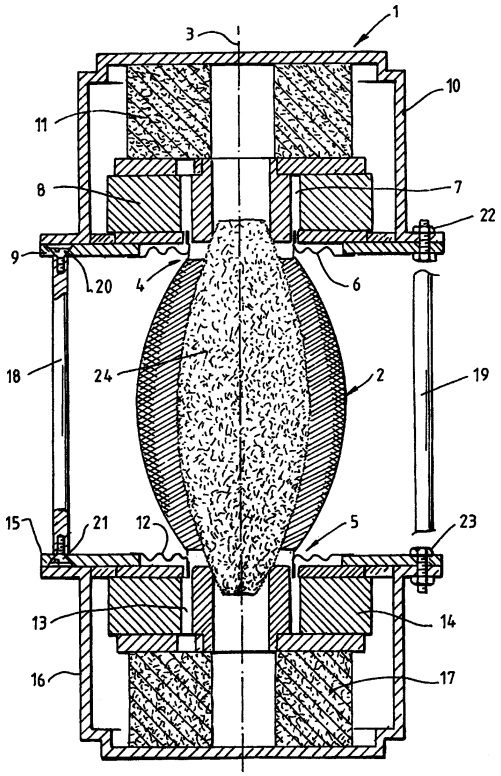
parent, allowing the user to see stator **S1**, which is also an electrical display panel. The arrangement is intended to save space in a cellular telephone.—GLA

6,785,397

43.38.Ja ELECTRO-ACOUSTIC CONVERTER

Barry Arnstein, Ivanhoe, Australia
31 August 2004 (Class 381/423); filed in Australia 16 February 2001

Flextensional transducers are well-known prior art, both as conventional loudspeakers and in underwater sound. There is nothing in this patent to indicate familiarity with either field. Most of the text is taken up with a long explanation of how a flextensional diaphragm works, yet there is no



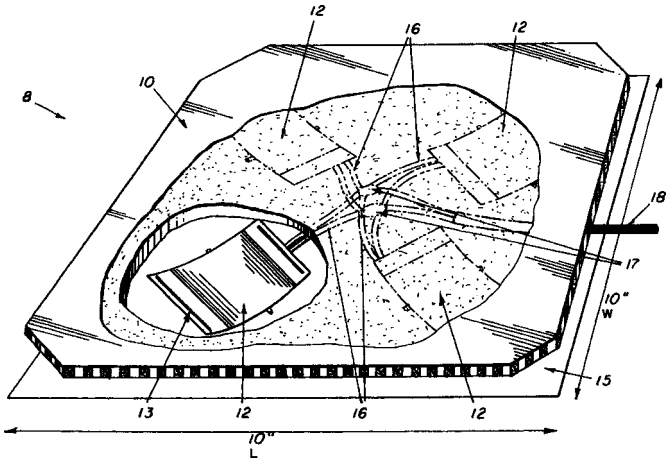
mention of the tricky pleated geometry required to make it work. The smooth, barrel-shaped diaphragm shown probably functions more like a bending wave transducer. The one novel feature of the invention appears to be the use of dual driving motors. Has no one thought of this before?—GLA

6,788,794

43.38.Ja THIN, LIGHTWEIGHT ACOUSTIC ACTUATOR TILE

Robert Corsaro *et al.*, assignors to The United States of America as represented by the Secretary of the Navy
7 September 2004 (Class 381/152); filed 1 October 2002

This invention is a low-frequency transducer intended for use in active noise control applications. What appears to be a free-edge diaphragm **10** is driven by several PZT bender elements **12** which are supported by backing structure **15**. Movement of the bender elements is transferred to the diaphragm through flexible strips **13**, thus allowing relatively long excursions with relatively little restraint. In use, the assembly might be attached to the interior of an aircraft cabin, presumably with no air seal between the edges of the diaphragm and the mounting surface, yet useful output down to 30 Hz is emphasized.—GLA

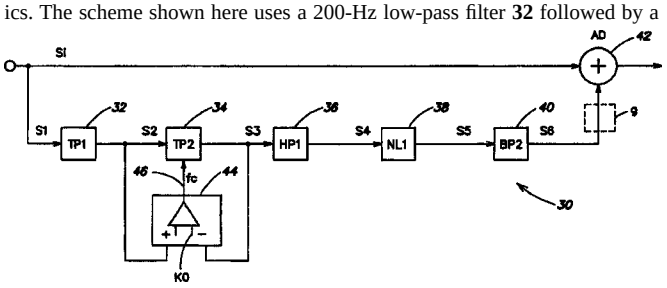


6,792,115

43.38.Ja APPARATUS FOR GENERATING HARMONICS IN AN AUDIO SIGNAL

Matthias Vierthaler, assignor to Micronas GmbH
14 September 2004 (Class 381/61); filed in Germany 18 November 1999

To compensate for the frequency characteristics of loudspeaker drivers, nonlinear harmonics can be used to generate the missing lower harmonics. The scheme shown here uses a 200-Hz low-pass filter **32** followed by a variable low-pass filter with a corner frequency $\mathbf{f_c}$. The nonlinear generator **NL1 38** is filtered by the bandpass filter **BP2 40** and added to the original signal at the summer **42**.—MK

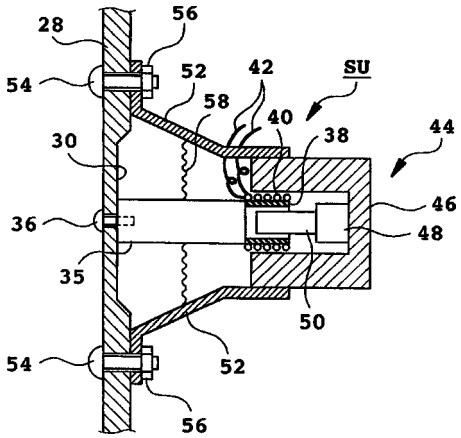


6,792,126

43.38.Ja INFORMATION PROCESSING DEVICE AND SPEAKER UNIT APPLICABLE THERETO

Hideyuki Okuno and Wataru Tanaka, assignors to Fujitsu Limited
14 September 2004 (Class 381/412); filed in Japan 15 May 1998

The case of your portable computer probably includes small perforated areas in front of tiny loudspeakers. However, prior art includes several methods of using an inertia transducer to drive the case itself. In this variant, a special portion **30** of the case is utilized as a clamped-edge diaphragm driven by a conventional moving-coil motor.—GLA

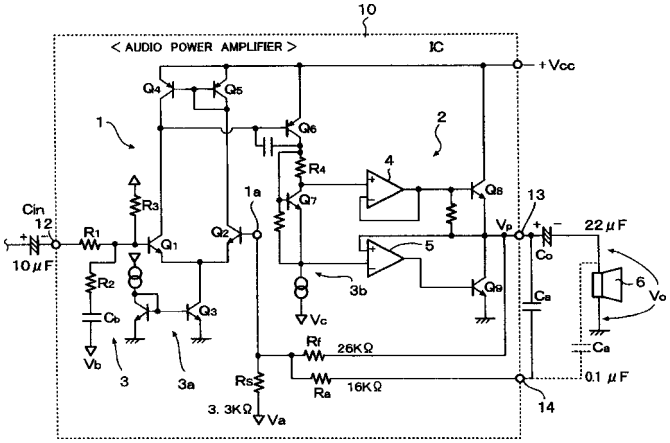


6,792,121

43.38.Lc AUDIO SIGNAL AMPLIFIER CIRCUIT AND PORTABLE TELEPHONE SET AND PORTABLE ELECTRONIC DEVICE USING THE SAME AUDIO SIGNAL AMPLIFIER CIRCUIT

Yasuyuki Koyama and Masanori Fujisawa, assignors to Rohm Company, Limited
14 September 2004 (Class 381/120); filed in Japan 17 April 2000

The circuit shown is a low-voltage, integrated circuit audio power amplifier that might be used in a mobile telephone or portable CD player. The values of input and output coupling capacitors **Cin** and **Co** require these to be outboard components, taking up valuable space. The output capacitor presents the biggest problem since it must be 2000 mf or larger to drive a low-impedance loudspeaker. Well, why not use a small capacitor and then equalize the amplifier to make up for the resulting droop in bass response? The invention takes this approach by adding components **Ca** and **Ra** to the feedback loop. The question of how an already underpowered amplifier can produce the required voltage swing is not discussed.—GLA



6,462,264

43.38.Md METHOD AND APPARATUS FOR AUDIO BROADCAST OF ENHANCED MUSICAL INSTRUMENT DIGITAL INTERFACE (MIDI) DATA FORMATS FOR CONTROL OF A SOUND GENERATOR TO CREATE MUSIC, LYRICS, AND SPEECH

Carl Elam, Baltimore, Maryland
8 October 2002 (Class 84/645); filed 26 July 1999

The author of this patent wishes to encode MIDI events for transmission over a rf channel such as television or radio. It is a pity that no mechanism was even discussed that would make it possible for these completely incompatible receivers to process the MIDI stream. Furthermore, the writer is so fervent in his MIDI belief, that he also wants to encode speech using MIDI.—MK

6,473,019

43.38.Md LOW CAPACITANCE, LOW KICKBACK NOISE INPUT STAGE OF A MULTI-LEVEL QUANTIZER WITH DITHERING AND MULTI-THRESHOLD GENERATION FOR A MULTI-BIT SIGMA-DELTA MODULATOR

Antti Ruha *et al.*, assignors to Nokia Corporation
29 October 2002 (Class 341/143); filed 21 June 2001

The use of dither in a sigma delta converter serves at least three purposes: (1) redistribution of quantization noise; (2) decorrelation of quantization; and (3) reduction of idle tones. In this invention, the authors address specific concerns related to switched-capacitor converters, specifically the ''kickback noise,'' which is the transient switching noise generated at comparator inputs when changing from latching to tracking. Circuit heads will like the schematics.—MK

6,760,635

43.38.Md AUTOMATIC SOUND REPRODUCTION SETTING ADJUSTMENT

Cary Lee Bates *et al.*, assignors to International Business Machines Corporation
6 July 2004 (Class 700/94); filed 12 May 2000

This sound control system operates on individual components of the sound system to produce a desired overall acoustic effect, for example, in an automobile. Individual volume and tone control settings may be stored for specific content, such as individual tracks from a CD or a specific news broadcast. A sentence in the patent says that ''the current song information is identified by the CPU.'' How this is to be done is not further discussed.—DLR

6,785,656

43.38.Md METHOD AND APPARATUS FOR DIGITAL AUDIO PLAYBACK USING LOCAL STORED CONTENT

Stelios Patsiokas *et al.*, assignors to XM Satellite Radio, Incorporated
31 August 2004 (Class 704/500); filed 5 June 2001

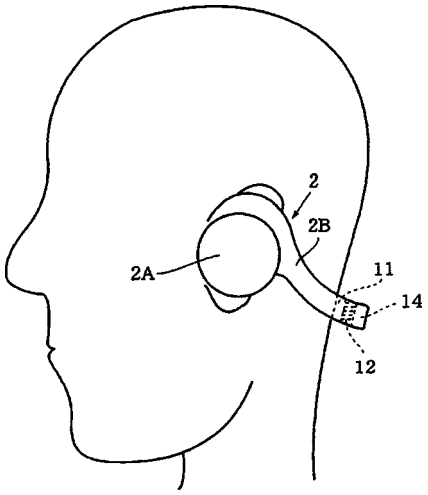
You can now patent a business plan—or so it appears. Here, XM Satellite Radio describes in the broadest terms the fundamentals behind their system. Details of even the vaguest kind are conspicuously absent.—MK

6,792,122

43.38.Si ACOUSTIC DEVICE

Haruo Okada *et al.*, assignors to Pioneer Corporation
14 September 2004 (Class 381/151); filed in Japan 28 December 1998

Tactile sensations are sometimes included with airborne sound reproduction for added impact and enjoyment. At one time or another, inventors



have designed vibrating chairs, headrests, dance floors, and lollipops. This new Pioneer headset includes a vibrating appendage that contacts the user's cervix.—GLA

6,785,563

43.38.Si MOBILE TERMINAL OPERATING IN TELEPHONIC AND TACTILE MODES

Tadashi Tsukamoto, assignor to NEC Corporation
31 August 2004 (Class 455/567); filed in Japan 12 May 2000

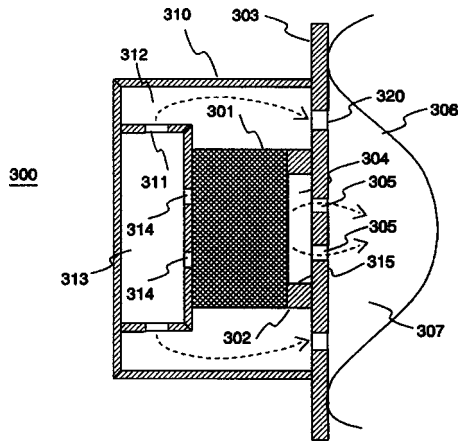
Most cellular telephones allow the user to switch between an audible mode and a silent vibration signal. This patent explains how to make the vibrator serve as an interrupt signal during normal voice operation.—GLA

6,788,798

43.38.Si METHOD AND ARRANGEMENT FOR IMPROVING LEAK TOLERANCE OF AN EARPIECE

Juha Backman, assignor to Nokia Mobile Phones Limited
7 September 2004 (Class 381/372); filed in Finland 6 October 1997

Headphones are normally designed and tested with the assumption that the earpiece will be sealed against the user's ear. In the case of a mobile telephone, however, an uncontrolled air leak is almost certain to be present, causing substantial low-frequency attenuation. To improve air leak tolerance, energy from the rear of the diaphragm can be conducted through one or more tuned chambers and used to load the front of the diaphragm, like a



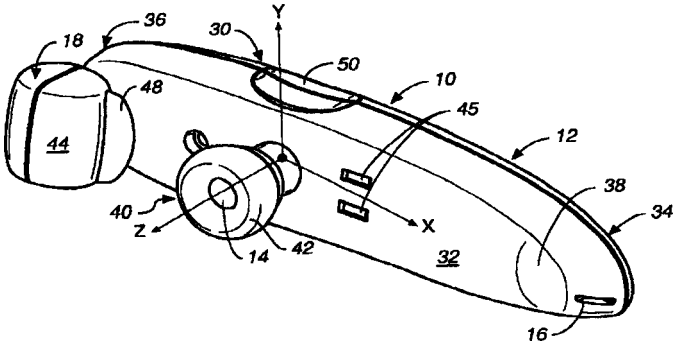
miniature R-J speaker enclosure. In this improved variant, front and rear energy are combined at the user's ear through openings **305** and **320** rather than in front chamber **304**. With this arrangement, the various acoustical impedances can be optimized such that a change in air volume **307** has only a minor effect on sound quality.—GLA

6,795,718

43.38.Si HEADSET COMMUNICATION DEVICE

Hyon S. Bae, assignor to Youngbo Engineering, Incorporated
21 September 2004 (Class 455/575.2); filed 15 February 2002

This very small, lightweight, all-in-one wireless headset is the aural equivalent of pince-nez eyeglasses—it is entirely supported by earbud **40**. ''The battery pack and the ear bud are further positioned on the casing to



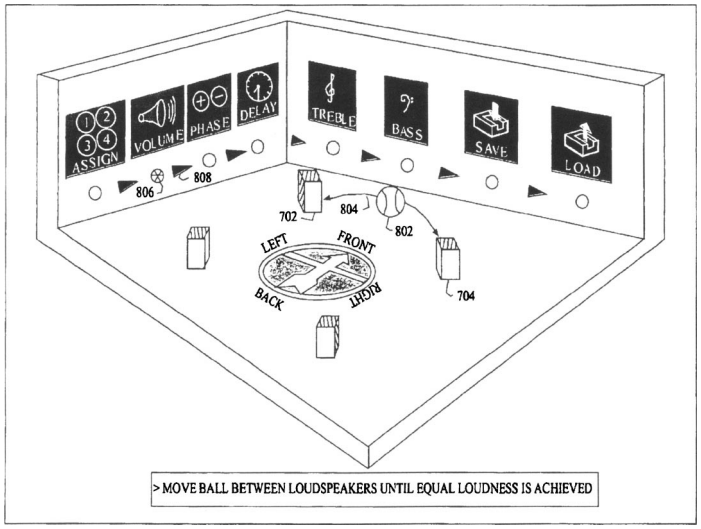
equally distribute the weight and ... prevent unnecessary movement of the headset ... during operation.''—GLA

6,798,889

43.38.Vk METHOD AND APPARATUS FOR MULTI-CHANNEL SOUND SYSTEM CALIBRATION

Sam Dicker *et al.*, assignors to Creative Technology Limited
28 September 2004 (Class 381/303); filed 13 November 2000

Setting up a modern surround-sound system in the typical home environment can be both a daunting and tiresome task if the standard ITU playback array is to be duplicated. But wait; there is an easier way to achieve this goal. You don't really have to position each loudspeaker exactly at its target position when an approximation will probably suffice. What you then do is ''collimate'' the system by individually adjusting the outputs of



the loudspeakers in terms of level, relative delay, polarity, and frequency response until all these factors coalesce and have been justified at the target listening position. If all of this were to be done manually it would take a very long time. What the authors have done here is to automate the process via a graphic user interface that appears to be interesting, innovative, and intuitive.—JME

6,804,358

43.38.Vk SOUND IMAGE LOCALIZING PROCESSOR

Seiji Kawano, assignor to Sanyo Electric Company, Limited
12 October 2004 (Class 381/17); filed in Japan 8 January 1998

The patent deals with stereo (dual-loudspeaker) playback strategies for creating the semblance of a surround-sound field, with source specificity, enveloping a listener positioned along the median plane of the playback array. The typical program input to such a system would be the five discrete channels from a Dolby digital, DTS, or one of several MPEG audio sources. The basic requirements for doing this have been visited and revisited many times over the last two decades, most notably by Cooper and Bauck (JAES, 37, 1/2), yet the patents keep coming. The author does not appear to have broken any new ground here, but those in this field may want to review the patent nevertheless.—JME

6,756,700

43.60.Bf SOUND-ACTIVATED WAKE-UP DEVICE FOR ELECTRONIC INPUT DEVICES HAVING A SLEEP-MODE

Yu-Wen Zeng, assignor to Kye Systems Corporation
29 June 2004 (Class 307/112); filed 13 March 2002

As portable and hand-held devices become more widespread, there is increasing interest in methods of conserving battery power as much as possible. One way of doing this, which has only been partially successful, is the use of a sleep mode. The solution offered here is to use a wake-up signal with significant low-frequency content, such as would be produced by touching or moving the device. Such a signal is more akin to vibration than audible sound, is easier to filter, and will typically have a greater amplitude difference between true and false candidate signals.—DLR

6,763,116

43.66.Ts HEARING AID AND OPERATING METHOD THEREFOR WITH CONTROL DEPENDENT ON THE NOISE CONTENT OF THE INCOMING AUDIO SIGNAL

Roland Barthel and Torsten Niederdränk, assignors to Siemens Audiologische Technik GmbH
13 July 2004 (Class 381/312); filed in Germany 24 September 2001

The input signal to a hearing aid is analyzed for the presence of predefined, unwanted disturbances using, if necessary, antennas in additional to the hearing aid microphone. If an interference signal, such as that from a cellular telephone, is thought to be present, the hearing aid switches parameter settings to minimize the interference.—DAP

6,755,657

43.70.Dn READING AND SPELLING SKILL DIAGNOSIS AND TRAINING SYSTEM AND METHOD

Janet M. Wasowicz, assignor to Cognitive Concepts, Incorporated
29 June 2004 (Class 434/167); filed 8 November 2000

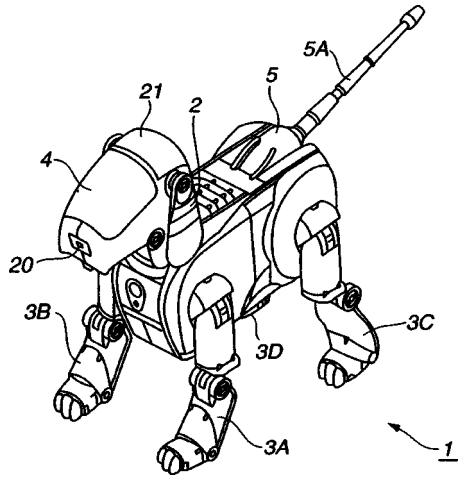
This patent represents another entry into the field of computer game-based speech training. Various game scenarios are available for teaching improvements in auditory processing, phonological awareness, morphological awareness, visual orthographic memory, and reading and spelling skills. Individual scenarios may be combined in various ways to provide a comprehensive training regimen customized for a particular required skill set.—DLR

6,754,560

43.71.An ROBOT DEVICE, ROBOT DEVICE ACTION CONTROL METHOD, EXTERNAL FORCE DETECTING DEVICE AND EXTERNAL FORCE DETECTING METHOD

Masahiro Fujita *et al.*, assignors to Sony Corporation
22 June 2004 (Class 700/245); filed in Japan 31 March 2000

The idea here is that a touch sensation and a perceived word would become associated in the robotic device's memory. According to this patent, prior robotic systems have been able to form separate concept ''nodes'' for certain sounds and for certain touch sensations and then able to link these two nodes. Here, the common occurrence of events in the two modalities



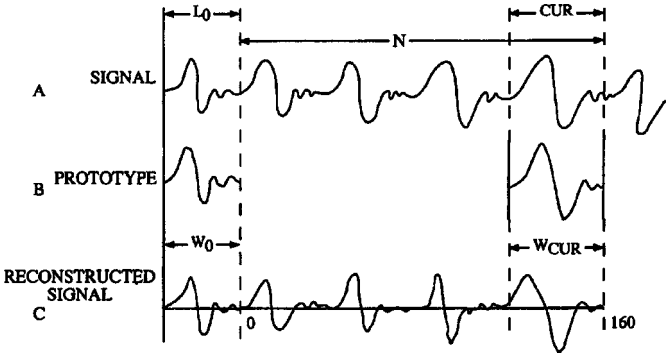
leads to the formation of a single concept ''node'' with links to the percepts in both modalities, as well as to corresponding motor commands. The result is said to provide a more effective way of learning a postural command, such as ''sit.''—DLR

6,754,630

43.72.Gy SYNTHESIS OF SPEECH FROM PITCH PROTOTYPE WAVEFORMS BY TIME-SYNCHRONOUS WAVEFORM INTERPOLATION

Amitava Das and Eddie L. T. Choy, assignors to Qualcomm, Incorporated
22 June 2004 (Class 704/268); filed 13 November 1998

This patent expands on a prior-art method of speech coding known as time-synchronous waveform interpolation. The appropriate references are cited. In this version, additional flexibility has been included for shifting the



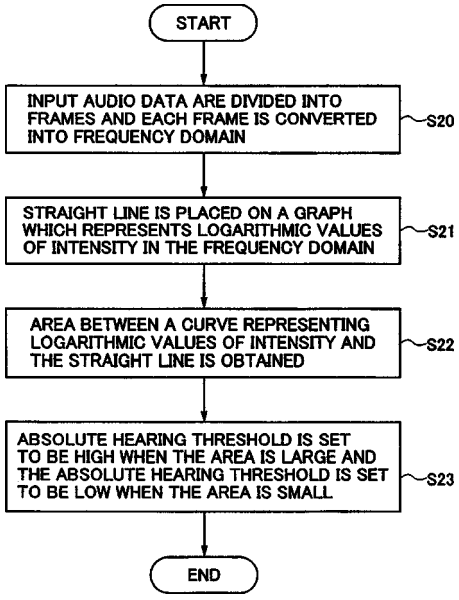
relative phases of the various component ''pitch prototype'' waveforms, which are then resampled to construct the output speech signal.—DLR

6,772,111

43.72.Gy DIGITAL AUDIO CODING APPARATUS, METHOD AND COMPUTER READABLE MEDIUM

Tadashi Araki, assignor to Ricoh Company, Limited
3 August 2004 (Class 704/200.1); filed in Japan 30 May 2000

A digital audio coding system is described in which the subband absolute hearing thresholds are changed adaptively as a function of the



frequency-domain intensity distribution of the audio data. Coding bits are assigned for each subband from allowable distortion levels determined from the thresholds.—DAP

6,791,481

43.72.Gy PORTABLE CD-ROM/ISO TO HDD/MP3 RECORDER WITH SIMULTANEOUS CD-READ/MP3-ENCODE/HDD-WRITE, OR HDD-READ/MP3-DECODE, TO PLAY, POWER SAVING BUFFER, AND ENHANCED SOUND OUTPUT

William Christopher Altare and Anton N. Handal, assignors to Echo Mobile Music, LLC
14 September 2004 (Class 341/55); filed 18 May 2001

A combination portable MP3 and CD recorder/player decodes words into audio waveforms while encoding and digitizing the audio in real time into 24-bit MP3 digital data. The MP3 data are stored temporarily in memory and then written out to a hard disk.—DAP

6,757,653

43.72.Ja REASSEMBLING SPEECH SENTENCE FRAGMENTS USING ASSOCIATED PHONETIC PROPERTY

Peter Buth *et al.*, assignors to Nokia Mobile Phones, Limited
29 June 2004 (Class 704/258); filed in Germany 30 June 2000

This speech synthesis system would construct output sentences by joining prerecorded speech fragments, each typically several words in length. In order to improve the resulting delivery, phrase fragments are selected according to codes which represent the style and rate of speech of the fragments. There is no discussion of the well-known disadvantages of this type of synthesis system, such as the requirement that all present and future phrases be recorded in a uniform manner by the same speaker. Examples are given of German-language synthesis for automobile navigation.—DLR

6,760,703

43.72.Ja SPEECH SYNTHESIS METHOD

Takehiko Kagoshima and Masami Akamine, assignors to Kabushiki Kaisha Toshiba
6 July 2004 (Class 704/262); filed in Japan 4 December 1995

This is essentially a speech synthesizer based on the concatenation of stored syllable-sized units, but with several additional schemes. The stored syllable units are indexed according to duration, stress, up to five contextual phonetic units, and a few other parameters. This much is prior art. The selected units are then filtered to improve the long-term spectral smoothness. A pitch profile is introduced by FFT, inverse FFT, and LPC, although the pitch profile source is not clearly described. The patent includes a fairly thorough and detailed treatment of how the spectral structure is to be constructed.—DLR

6,760,704

43.72.Ja SYSTEM FOR GENERATING SPEECH AND NON-SPEECH AUDIO MESSAGES

Steven M. Bennett, assignor to Intel Corporation
6 July 2004 (Class 704/270); filed 29 September 2000

This system allows the user of a PDA or cell phone to set up a system of audio cues to signal waiting messages or other events which need attention. Designed to appeal to the latest generation of turned-on, hyperactive users, audio cues may include either speech or nonspeech sounds and may overlap in certain combinations. Thus, a beep meaning one thing may happen in the middle of the playback of something else.—DLR

6,760,702

43.72.Ne METHOD FOR GENERATING CANDIDATE WORD STRINGS IN SPEECH RECOGNITION

Shih-Chieh Chien and Sen-Chia Chang, assignors to Industrial Technology Research Institute
6 July 2004 (Class 704/251); filed in Taiwan, Province of China 21 February 2001

The patent describes what is said to be an efficient method of generating candidate word sequences in a speech recognition system. Roughly, it seems to work as follows. Each candidate word (or ''node'') is rated by the best overall match which could include that word. Nodes with high scores are grouped. String start and end times are determined for sets of nodes with high scores. Sets with close end-point times are selected. Chinese-language examples are given.—DLR

6,760,705

43.72.Ne VIRTUAL SPEECH INTERFACE SYSTEM AND METHOD OF USING SAME

Joseph L. Dvorak, assignor to Motorola, Incorporated
6 July 2004 (Class 704/270.1); filed 31 May 2001

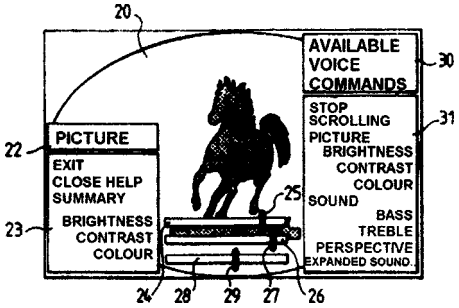
What is meant here by a virtual speech interface is that an extra layer of speech recognition server is provided in a network environment to allow control of other network-connected devices or software systems which do not themselves have a voice interface capability. An example is suggested in which the user adjusts a camera by voice control while sighting through the viewfinder. The camera could have a Bluetooth or other wireless interface, but probably would not have a speech recognition system of its own.—DLR

6,762,692

43.72.Ne SYSTEM COMPRISING A REMOTE CONTROLLED APPARATUS AND VOICE-OPERATED REMOTE CONTROL DEVICE FOR THE APPARATUS

Jacques Mingot *et al.*, assignors to Thomson Licensing S.A.
13 July 2004 (Class 340/825.69); filed in France 21 September 1998

This patent says that if you have a device with a display screen capability and you want to set it up with a speech recognition controller, then



you can use the display to show the menu prompts. Is that nonobvious, or what?—DLR

6,763,331

43.72.Ne SENTENCE RECOGNITION APPARATUS, SENTENCE RECOGNITION METHOD, PROGRAM, AND MEDIUM

Yumi Wakita and Kenji Matsui, assignors to Matsushita Electric Industrial Company, Limited
13 July 2004 (Class 704/251); filed in Japan 1 February 2001

This sentence understander is set up to work from either spoken or text input. It begins with a typical phonetic match, leading to an acoustic score, followed by a word match, which produces a language score. If the best sum of the two scores exceeds the second best by a suitable margin, success is declared. Otherwise, the system uses a low-budget version of *n*-word statistics, where $n=2$. It basically just keeps a list of word pairs it has seen and if any of the individual words from a stored pair is found in the input, then a system of linguistic word classes is used to guess other words from the database which might be found in the input. This provides a narrowed field of possible word candidates, allowing a faster search for a sentence match. Some examples are given in untranslated Japanese.—DLR

6,766,290

43.72.Ne VOICE RESPONSIVE AUDIO SYSTEM

Iwan R. Grau, assignor to Intel Corporation
20 July 2004 (Class 704/211); filed 30 March 2001

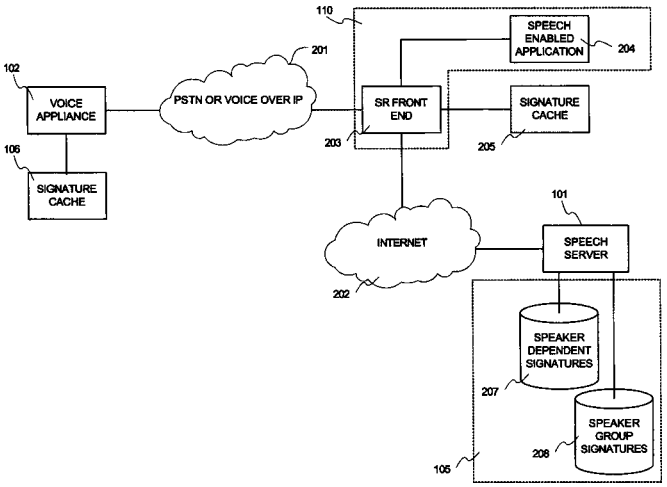
This patent says that if you have a sound system, which may provide either live or replayed audio material, and you want to set up a speech recognition system for voice control, then the system can subtract the audio material from the microphone signal to get a cleaner copy of your speech input. And, Intel paid for the patent.—DLR

6,785,647

43.72.Ne SPEECH RECOGNITION SYSTEM WITH NETWORK ACCESSIBLE SPEECH PROCESSING RESOURCES

William R. Hutchison, Boulder, Colorado
31 August 2004 (Class 704/231); filed 20 April 2001

Speaker-dependent voice signatures are developed for several speakers and stored in a network repository. A configurable speech-processing engine identifies a particular speaker and uses the speaker-dependent voice signature associated with the identified speaker to perform speech recognition. The voice signatures are compatible with different types of computer hardware and can be used with several different voice-enabled applications.—DAP

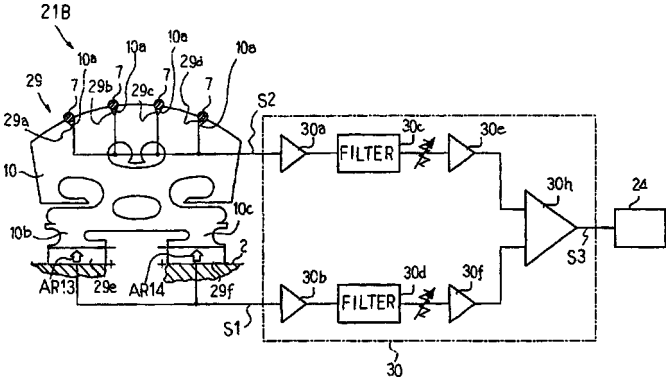


6,791,023

43.75.De BOWED STRINGED MUSICAL INSTRUMENT FOR GENERATING ELECTRIC TONES CLOSE TO ACOUSTIC TONES

Hiroshi Nakaya, assignor to Yamaha Corporation
14 September 2004 (Class 84/736); filed in Japan 16 November 2001

This patent begins by enumerating problems with existing electric violins: (1) ''poor prompitude in following up the bowing''—i.e., cancellation of the piezoelectric transducers; (2) ''poor fidelity of the tones''—thin transducers with good high-frequency response are expensive, so thick transducers are preferred from a cost standpoint; and (3) poor imaging of the synthetic tones by the player. As shown in the figure, string pickup units **29** are amplified and filtered by the low-pass filter **30c**. The signals generated by piezoelectric transducers at the foot of the bridge **29e**, **29f** are amplified and filtered. As you might imagine, the patent writing is stilted and awkwardly translated.—MK

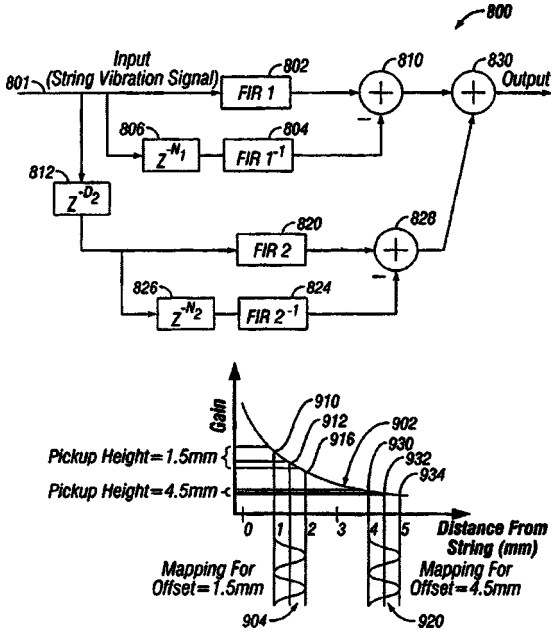


6,787,690

43.75.Fg STRINGED INSTRUMENT WITH EMBEDDED DSP MODELING

Peter J. Celi *et al.*, assignors to Line 6
7 September 2004 (Class 84/723); filed 16 July 2002

This simple DSP solution as shown in the figure includes a delay for multiple transducers z^{-D} and the delay from the nut to the pickups z^{-N}. The



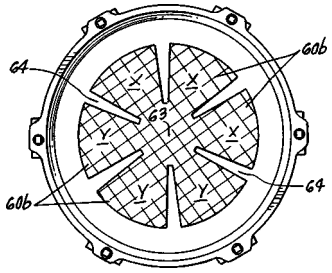
response of the magnetic pickup can be equalized as well. Note that this scheme does not add body response or any other effects.—MK

6,518,490

43.75.Hi DRUM HEAD WITH SOUND ATTENUATING CENTER COATING

John J. Good, assignor to Drum Workshop, Incorporated
11 February 2003 (Class 84/414); filed 23 January 2001

By adding a heavier material **63** to the surface of the drum head, a



drum can provide local attenuation regions. The patent specifically mentions the use of epoxy resin with abrasion filler particles.—MK

6,784,352

43.75.Hi DRUMHEAD AND MUTING STRUCTURE FOR ACOUSTIC AND ELECTRONIC PERCUSSION INSTRUMENTS

Yuichiro Suenaga, assignor to Yamaha Corporation
31 August 2004 (Class 84/411 R); filed in Japan 15 November 1999

It is argued here that by producing a porous drumhead surface, the loudness of the drum will be diminished. An additional embodiment illustrates the use of such a drumhead in an electronic drum.—MK

6,791,019

43.75.St STAFF SHEET PRINTER

Toshio Hayakawa, Los Angeles, California
14 September 2004 (Class 84/483.1); filed 29 January 2002

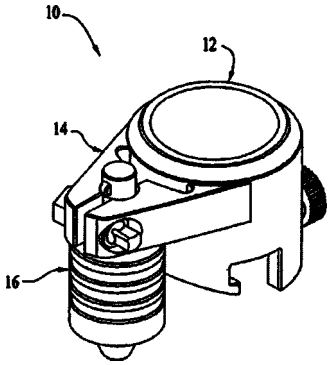
In spite of the title, what is proposed is a music transcription device. But, the details of the signal-processing module are completely and utterly absent. The remaining material and claims are vacuous.—MK

6,794,569

43.75.Hi ACOUSTIC INSTRUMENT TRIGGERING DEVICE AND METHOD

Yuujirou Kamijima and Kiyoshi Yoshino, assignors to Roland Corporation
21 September 2004 (Class 84/746); filed 14 January 2003

This attachable sensor is designed to attach to drum rims. As shown,



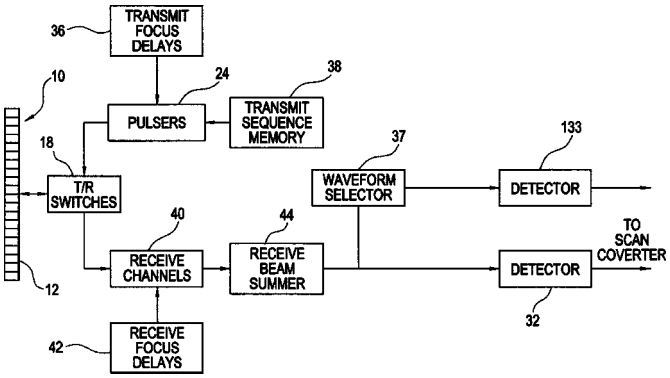
the sensor **16** is attached to the arm **14** and post attachment. The piezoelectric sensor is mounted at the end of the arm.—MK

6,796,944

43.80.Vj DISPLAY FOR SUBTRACTION IMAGING TECHNIQUES

Anne Hall, assignor to GE Medical Systems Global Technology, LLC
28 September 2004 (Class 600/443); filed 17 May 2002

Echoes received from two ultrasound pulse transmissions are processed to obtain a flow or contrast signal component from which an image is



formed. Echoes from one or the other or both transmissions are processed to obtain a B-mode background or reference signal from which another image is formed. The flow or contrast image is displayed along with the B-mode reference image.—RCW

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.

Embedding uncertainty into ocean acoustic propagation models (L)

Steven Finette[a)]
Acoustics Division, Naval Research Laboratory, Washington, DC 20375

(Received 9 September 2004; revised 8 December 2004; accepted 13 December 2004)

A probabilistic formalism is proposed for the direct inclusion of environmental uncertainty into an acoustic model that describes propagation in an ocean waveguide. Incomplete environmental knowledge is characterized by a spectral representation of uncertainty using expansions of random processes in terms of orthogonal random polynomials. A brief summary of the method is presented and a set of coupled differential equations describing the propagation of both the acoustic field and its associated uncertainty is derived for the case where the uncertain environment is attributed to a lack of complete information concerning the waveguide's sound speed distribution.
[DOI: 10.1121/1.1855811]

PACS numbers: 43.20.Bi, 43.20.Mv, 43.30.Re [DRD] Pages: 997–1000

I. INTRODUCTION

It is assumed implicitly, though rarely stated explicitly, that results obtained through numerical simulation of acoustic wave propagation in ocean waveguides have some predictive capability for estimating, e.g., transmission loss, spatial coherence, detection range, or sonar system performance. The significance of the results depends critically on the nature of environmental uncertainty and how (or if) this issue is addressed in the context of interpreting the results of a numerical simulation. The term uncertainty is defined here as some quantitative measure of our lack of complete knowledge or understanding concerning the description of both the sound speed field and boundary conditions that constitute necessary waveguide information for simulation based prediction of the acoustic field. This ''knowledge gap'' introduces errors within the model formulation that can interact, amplify, and propagate along with the acoustic field and this paper proposes a method to incorporate uncertainty within the propagation physics. It is important to note that these errors, due to fixed, but poorly known and/or under-sampled values for the physical inputs, are quite distinct from errors related to numerical discretization of either the wave equation or its associated boundary conditions.

The predictive capability would be significantly enhanced through the specification of an objective measure of uncertainty associated with the characterization of the environment through which the acoustic field propagates. However, numerical algorithms describing wave propagation in ocean acoustics are most often based on a *deterministic* wave equation, i.e., one whose boundary conditions, parameters, etc., are fixed and implicitly assumed to be known *exactly*. The pressure field that results from applying a numerical algorithm to solve this equation can be interpreted as a function having zero variance (i.e., no uncertainty). Since real world environments are neither known with infinite accuracy nor properly sampled in space and time this deterministic approach, by itself, is incompatible with the direct inclusion of environmental uncertainty within the simulation. An indirect and commonly used method to include uncertainty in the dynamics is based on Monte Carlo sampling. For example, realizations of random processes describing the sound speed field or bathymetry can be estimated by sampling probability distributions that describe uncertainty about the environment. The deterministic wave equation is then solved separately over many realizations and the resulting set of pressure fields is used to estimate statistical moments of the field.[1,2] While generally valid, proper sampling can involve the computation of many thousands of realizations to ensure convergence of the statistics and it is computationally prohibitive under a number of reasonable environmental scenarios. An example of such a scenario involves three-dimensional acoustic propagation where the sound speed field is determined through the solution of the equations of motion governing fluid flow. The approach described in Sec. II effectively computes all realizations of the pressure field simultaneously by solving a set of coupled differential equations. This allows for a significant computational speed advantage relative to

[a)]Electronic mail: finette@wave.nrl.navy.mil

Monte Carlo based methods, where realizations are computed individually. The new approach also provides a much more compact description of uncertainty in complex systems comprised of subsystems for which uncertainty can be transferred among the subsystems. In ocean acoustic modeling, a two-component system might consist of the ocean environment and the pressure field propagating within that environment. The uncertainty in the environment subsystem couples into the second subsystem through the sound speed field, and a complete description of the uncertainty must include that introduced in both subsystems and propagated between the subsystems. The compact formulation facilitates quantitative analysis of the effects of acoustic field uncertainty due to different subsystem components without explicitly dealing with a large set of separate field realizations. An alternative method for the incorporation of uncertainty considers the propagation of the field moments by the use of moment equations.[3,4] This random media approach often involves a number of simplifying mathematical and physical assumptions (e.g., closure, stationarity, the Markov assumption) that are either difficult to justify in practice or tend to limit applications to relatively benign environments. In addition, a separate equation must be introduced and solved for each moment, and these equations become increasingly complex for higher moments. The probabilistic method outlined here allows for a unified formulation that includes information on all moments and fewer basic assumptions than the moment equation methods.

A probabilistic framework that incorporates uncertainty as an integral component of the model dynamics was developed in the field of stochastic finite elements.[5] The method has quite general validity and is extended here to include wave propagation in ocean waveguides where uncertainty related to the waveguide properties is naturally embedded directly into the propagation physics. This extension involves a generalization of the deterministic formulation of wave propagation to a stochastic framework, where the original deterministic parameters and fields are now treated as random variables or random fields.[6] A stochastic differential equation that incorporates both the wave dynamics and environmental uncertainty is then derived and projected onto a subspace associated with an uncertainty ''dimension,'' where it can be solved to obtain statistics of the field as well as relationships between uncertainty linked to different components of the modeled environment. The statistics represent a probabilistic measure of the effect of environmental uncertainty on the acoustic field. As an illustration, Sec. II presents a derivation of a set of coupled differential equations describing the propagation of both the acoustic field and its corresponding uncertainty in which the latter quantity represents a lack of complete environmental knowledge of the sound speed distribution.

II. DERIVATION OF UNCERTAINTY EQUATIONS

In this section, uncertainty related to the sound speed field is directly incorporated into the narrow angle parabolic equation.[6] It should be emphasized that the approach discussed in the following is not limited to the narrow angle formulation, but this approximation is used here to illustrate the method in a simple setting. In addition to the usual space variables, (r,z), representing two-dimensional propagation, an additional dimension, θ, is introduced to include uncertainty. In a probabilistic sense, a value of θ can be interpreted as representing an element of an event space containing, for example, the set of all possible realizations of the sound speed field. This dimension is associated with both the sound speed distribution and acoustic pressure field since they are functionally related. The (deterministic) narrow angle parabolic equation can be written in the form[7]

$$2jk_0\frac{\partial\psi(r,z)}{\partial r}+\frac{\partial^2\psi(r,z)}{\partial z^2}+k_0^2(n^2(r,z)-1)\psi(r,z)=0, \tag{1}$$

where the acoustic pressure $P(r,z)$ is related to ψ by the relation $P(r,z)\approx(2/\sqrt{\pi k_0 r})e^{j(k_0 r-\pi/4)}\psi(r,z)$ in the far field approximation, $n(r,z)$ is the index of refraction, $c_0/c(r,z)$, relative to a reference wave number k_0 and $j=\sqrt{-1}$. Incomplete knowledge of the sound speed distribution is now included by expressing $c(r,z)$ as a random field comprised of a dominant, deterministic component, $\overline{c}(z)$, and a small stochastic contribution, δc, using the ansatz $c(r,z)\Rightarrow c(r,z;\theta)=\overline{c}(z)+\delta c(r,z;\theta)$. For ocean acoustics problems, it is assumed that $\langle(\delta c/\overline{c})^2\rangle\ll 1$, where $\langle\ \rangle$ denote ensemble average. A significant amount of complexity is embedded in δc, which might include randomness related to an internal wave field or other oceanographic structures that are incompletely known for a particular environment. This ansatz implies that the function ψ is also a random field, $\psi\Rightarrow\psi(r,z;\theta)$, since it is a nonlinear functional of the sound speed distribution. Therefore, Eq. (1) can be generalized to the following stochastic differential equation:

$$2jk_0\frac{\partial\psi(r,z;\theta)}{\partial r}+\frac{\partial^2\psi(r,z;\theta)}{\partial z^2}+k_0^2\left(\frac{c_0^2}{\overline{c}^2(z)}-1\right)\psi(r,z;\theta)$$
$$-\frac{2c_0^2k_0^2\delta c(r,z;\theta)}{\overline{c}^3(z)}\psi(r,z;\theta)=0. \tag{2}$$

Assuming that ψ and δc are second-order random processes (i.e., they have finite variance or ''energy''), the key step is to apply a theorem by Cameron and Martin,[8] which states that such processes can always be expressed as stochastic functional expansions in terms of a complete set of orthogonal polynomials $\{\Phi_i\}$ in the random variables $\xi_i(\theta)$:

$$\mu(\theta)=b_0\Phi_0+\sum_{i_1=1}^{\infty}b_{i_1}\Phi_1(\xi_{i_1}(\theta))+\sum_{i_1=1}^{\infty}\sum_{i_2=1}^{i_1}b_{i_1i_2}\Phi_2(\xi_{i_1}(\theta),\xi_{i_2}(\theta))+\cdots. \tag{3}$$

In this expansion, Φ_q represents a member of a complete basis set in the Hilbert space of second-order random variables and is known as the polynomial chaos[5] of order q. The series is mean-square convergent and the basis functions can be explicitly identified from the Wiener–Askey class of multivariate orthogonal functionals;[9] these include (but are not limited to) the familiar Hermite, Laguerre, and Legendre polynomials. Once the deterministic expansion coefficients,

$b_{i_1 i_2 \ldots}$, are computed, Eq. (3) contains a complete probabilistic description of sound speed or acoustic pressure along with their respective uncertainties, depending on whether μ represents δc or ψ. Note that the coefficients may be functions of space and time and they weight the relative contributions of uncertainty from different terms in the expansion. If the process represented by Eq. (3) has a "short" correlation length (approaching white noise) more terms are needed to capture the uncertainty, while only a relatively few terms are needed to provide a compact description of the uncertainty when δc and ψ are smoothly varying over the spatial and temporal scales of interest ("long" correlation length).[5,9] The family of stochastic basis functions $\{\Phi_q\}$ form a complete orthogonal basis set in this Hilbert space in the sense that the ensemble average $\langle \Phi_i \Phi_k \rangle \equiv \int \Phi_i(\boldsymbol{\xi}) \Phi_k(\boldsymbol{\xi}) W(\boldsymbol{\xi}) d\boldsymbol{\xi} = \langle \Phi_i^2 \rangle \delta_{ik}$, where $W(\boldsymbol{\xi})$ is a weighting function corresponding to the particular choice of basis.[5,9] The weighting factor represents the multivariate probability density function of the independent random variables $\boldsymbol{\xi}$ that characterize uncertainty in the system. For example, if $\boldsymbol{\xi}$ represents a multidimensional Gaussian random variable, then the corresponding polynomial is the Hermite polynomial and the weighting function is a multidimensional Gaussian density function. Uncertainty is introduced through the presence of the density function, the spread of which is interpreted as a probabilistic measure of our lack of knowledge concerning the waveguide environment. It has been shown[5] that Monte Carlo simulation represents a special case of Eq. (3) where a particular realization or outcome $\theta = \theta_i$ is selected. In effect, polynomial chaos expansions allow for all possible realizations to be computed simultaneously and described in a compact fashion, while an equivalent simulation using Monte Carlo sampling computes each realization separately.

For notational convenience the series in Eq. (3) is written in the simplified notation $\mu(\theta) = \Sigma_{i=0}^{\infty} c_i \Omega_i(\boldsymbol{\xi}(\theta))$, where a one-to-one mapping relates the terms in this expression with those of Eq. (3) and the Ω_i also satisfy the orthogonality relation discussed above.[5] Substituting into Eq. (2) the truncated expansions $\delta c(r,z;\theta) = \Sigma_{i=0}^{Q} \alpha_i(r,z) \Omega_i(\boldsymbol{\xi}(\theta))$ and $\psi(r,z;\theta) = \Sigma_{m=0}^{P} \gamma_m(r,z) \Omega_m(\boldsymbol{\xi}(\theta))$ for the random processes representing the sound speed perturbations and field, respectively, multiplying the result by Ω_l, and ensemble averaging the result yields

$$2jk_0 \sum_{i=0}^{P} \frac{\partial \gamma_i(r,z)}{\partial r} \langle \Omega_i \Omega_l \rangle + \sum_{i=0}^{P} \frac{\partial^2 \gamma_i(r,z)}{\partial z^2} \langle \Omega_i \Omega_l \rangle + k_0^2 \left(\frac{c_0^2}{\overline{c^2}} - 1 \right) \sum_{i=0}^{P} \gamma_i(r,z) \langle \Omega_i \Omega_l \rangle - \frac{2 c_0^2 k_0^2}{\overline{c^3}(z)} \sum_{i=0}^{Q} \sum_{m=0}^{P} \alpha_i(r,z) \gamma_m(r,z) \langle \Omega_i \Omega_m \Omega_l \rangle = 0. \quad (4)$$

Applying the orthogonality relation to Eq. (4) yields a set of coupled, deterministic, partial differential equations for the coefficients $\gamma_l(r,z)$ linking the uncertainty in the acoustic field to that induced by incomplete knowledge in the sound speed distribution:

$$2jk_0 \frac{\partial \gamma_l(r,z)}{\partial r} + \frac{\partial^2 \gamma_l(r,z)}{\partial z^2} + k_0^2 \left(\frac{c_0^2}{\overline{c^2}(z)} - 1 \right) \gamma_l(r,z) - \frac{2 c_0^2 k_0^2}{\overline{c^3}(z) \langle \Omega_l^2 \rangle} \sum_{i=0}^{Q} \sum_{m=0}^{P} \alpha_i(r,z) \gamma_m(r,z) \langle \Omega_i \Omega_m \Omega_l \rangle = 0. \quad (5)$$

The last term in this expression describes the coupling of uncertainty between the input sound speed distribution and the output pressure field. Uncertainty in the pressure field, represented by the values of the expansion coefficients γ_m, is therefore constrained by both the physics of wave propagation and uncertainty in our knowledge of the sound speed field. The expansion coefficients $\alpha_i(r,z)$ can be estimated from either a model of the sound speed perturbations or through experimentally acquired sound speed data by using the orthogonality relation for the basis functions, while ensemble averages over products of known basis functions can be computed analytically or through numerical integration. After solving Eq. (5) for $\gamma_l(r,z)$, the harmonic pressure field is determined by

$$P(r,z;\theta) \approx \sqrt{\frac{2}{\pi k_0 r}} \exp\left[j \left(k_0 r - \frac{\pi}{4} \right) \right] \sum_{i=0}^{P} \gamma_i(r,z) \Omega_i(\boldsymbol{\xi}(\theta)) \quad (6)$$

and field moments to all orders can be computed by algebraic manipulation of Eq. (6). For example, the second spatial moment or spatial coherence of the pressure field, R_{12}, can be expressed in the form

$$R_{12}(r_1,z_1;r_2,z_2) = A \sum_{i=1}^{P} [\gamma_i^*(r_1,z_1) \gamma_i(r_2,z_2) \langle \Omega_i^2 \rangle], \quad (7)$$

where A is a constant factor and the asterisk denotes complex conjugation. When $r_2 = r_1$, $z_2 = z_1$, this expression reduces to the variance of the field at a particular point in the waveguide.

Potential limitations and related implementation issues will have to be carefully addressed for practical applications. Processes characterized by a short correlation length need many terms in the spectral expansions, increasing the complexity and computational load. A sound speed field may contain several dominant correlation length scales; the inclusion of a correlation cutoff, for example, would depend on the acoustic frequency. The sensitivity to errors in computing the coefficients is a measure of the robustness of the approach and needs to be assessed.[10] Also, the inclusion of multiple sources of uncertainty, described by an increased dimensionality of the basis functions, may affect the convergence rate. While the series representing the random processes are guaranteed to converge, they must be truncated for practical applications and numerical issues related to truncation need to be addressed with respect to the particulars of the environment. Uncertainty in the boundary conditions can also be included in this formalism, though this issue and those related to the numerical solution of Eq. (5) are beyond the scope of this note. The solution of Eq. (5) for a particular choice of basis function, including validation of the approach

by a direct comparison with field moments computed through Monte Carlo simulations, is the subject of future work.

ACKNOWLEDGMENT

The author would like to thank the Office of Naval Research for support through NRL Base Funding.

[1] D. Tielburger, S. Finette, and S. Wolf, ''Acoustic propagation through an internal wave field in a shallow water waveguide,'' J. Acoust. Soc. Am. **101**, 787–808 (1997).
[2] D. Rouseff and T. E. Ewart, ''Effect of random sea surface and bottom roughness on propagation in shallow water,'' J. Acoust. Soc. Am. **98**, 3397–3404 (1995).
[3] L. Dozier and F. Tappert, ''Statistics of normal mode amplitudes in a random ocean. I. Theory,'' J. Acoust. Soc. Am. **63**, 353–365 (1978).
[4] B. J. Uscinski, C. Macaskill, and T. E. Ewart, ''Intensity fluctuations. I. Theory,'' J. Acoust. Soc. Am. **74**, 1474–1483 (1983).
[5] R. Ghanem and P. Spanos, *Stochastic Finite Elements: A Spectral Approach* (Springer, Berlin, 1991).
[6] S. Finette, ''Propagation of uncertainty for acoustic fields using polynomial chaos expansions,'' J. Acoust. Soc. Am. **114**, 2461 (2003).
[7] F. D. Tappert, ''The parabolic approximation method,'' in *Wave Propagation and Underwater Acoustics*, edited by J. B. Keller and J. S. Papadakis, Lecture Notes in Physics, Vol. 70 (Springer, New York, 1977), pp. 224–287.
[8] R. Cameron and W. Martin, ''The orthogonal development of nonlinear functionals in series of Fourier-Hermite functionals,'' Ann. Math. **48**, 385–392 (1947).
[9] D. Xiu and G. Karniadakis, ''The Wiener-Askey polynomial chaos for stochastic differential equations,'' SIAM J. Sci. Comput. (USA) **24**, 619–644 (2002).
[10] R. Ghanem, ''Probabilistic characterization of transport in heterogeneous media,'' Comput. Methods Appl. Mech. Eng. **158**, 199–220 (1998).

Spectral subtraction-based speech enhancement for cochlear implant patients in background noise (L)

Li-Ping Yang and Qian-Jie Fu[a)]
Department of Auditory Implants and Perception, House Ear Institute, 2100 West Third Street, Los Angeles, California 90057

(Received 20 February 2004; revised 14 September 2004; accepted for publication 2 December 2004)

A single-channel speech enhancement algorithm utilizing speech pause detection and nonlinear spectral subtraction is proposed for cochlear implant patients in the present study. The spectral subtraction algorithm estimates the short-time spectral magnitude of speech by subtracting the estimated noise spectral magnitude from the noisy speech spectral magnitude. The artifacts produced by spectral subtraction (such as ''musical noise'') were significantly reduced by combining variance-reduced gain function and spectral flooring. Sentence recognition by seven cochlear implant subjects was tested under different noisy listening conditions (speech-shaped noise and 6-talker speech babble at +9, +6, +3, and 0 dB SNR) with and without the speech enhancement algorithm. For speech-shaped noise, performance for all subjects at all SNRs was significantly improved by the speech enhancement algorithm; for speech babble, performance was only modestly improved. The results suggest that the proposed speech enhancement algorithm may be beneficial for implant users in noisy listening. © *2005 Acoustical Society of America.*
[DOI: 10.1121/1.1852873]

PACS numbers: 43.71.Es, 43.71.Ky, 43.66.Ts Pages: 1001–1004

I. INTRODUCTION

Cochlear implants (CIs) have provided profoundly deaf individuals with hearing sensation. Many postlingually deafened adults are capable of high degrees of speech understanding in quiet listening conditions. However, CI users' speech recognition falls sharply in the presence of background noise or competing speakers (Kiefer *et al.*, 1996; Müller-Deile *et al.*, 1995; Fetterman and Domico, 2002; Hochberg *et al.*, 1992).

To reduce the effects of background noise, some speech enhancement algorithms originally developed for normal hearing (NH) persons have been applied to CI speech processing (Müller-Deile *et al.*, 1995; Hochberg *et al.*, 1992); these algorithms were able to somewhat improve CI users' performance in noisy listening conditions. However, many NH speech enhancement algorithms have not been tested in a CI speech-processing context. Our purpose in this study is to develop a speech enhancement algorithm for CI patients by modifying some of the most recent and most powerful speech enhancement algorithms developed for NH persons. The algorithm is designed to preprocess the input speech before CI speech processing.

This study focused on single-channel speech enhancement. Spectral subtraction was used in the algorithm because of the simplicity implementation and the flexibility of parametric adjustments to the subtraction technique. In general, spectral subtraction estimates the short-time spectral magnitude of speech by subtracting the estimated noise spectral magnitude from the noisy speech spectral magnitude (Boll, 1979). The phase of the noisy speech spectrum is preserved based on the best estimate of the clean speech in a least mean square sense. The main disadvantage of the spectral subtraction algorithm is the resultant ''musical noise'' artifact, which is mainly due to variance in the spectrum estimation. Several approaches have been proposed to overcome the musical noise problem (Ephraim and Malah, 1984; Gustafsson *et al.*, 2001; Lockwood and Boudy, 1992; Berouti *et al.*, 1979). In the present study, some previously proposed methods were combined to provide the optimal performance.

[a)]Send correspondence to Qian-Jie Fu, Ph.D., Department of Auditory Implants and Perception, House Ear Institute, 2100 West Third Street, Los Angeles, California 90057. Phone: (213) 273-8036; fax: (213) 413-0950; electronic mail: qfu@hei.org

II. EXPERIMENT DESIGN

A. Algorithm

The proposed speech enhancement algorithm consists of two major parts: speech pause detection and spectral subtraction.

1. Speech pause detection

A speech pause detection algorithm proposed by Marzinzik and Kollmeier (2002) was used in the present study to detect when speech was absent from the input signal. The algorithm detects speech pauses by adaptively tracking minima in a noisy signal's power envelope, both for the broadband signal and for the high-pass and low-pass filtered signal. Some constant values used in this study were different from those by Marzinzik and Kollmeier (2002); time constant τ_E (envelope smoothing)=12 ms, and both time constants τ_{decay} and $\tau_{\text{raise}}=7.3$ s.

2. Spectral subtraction

The spectral subtraction algorithm used in the present study was based on an algorithm proposed by Gustafsson *et al.* (2001) with some modifications.

Let $s(n)$, $d(n)$, and $y(n)$ represent the speech signal, noise signal, and noisy speech signal, respectively. Therefore,

$$y(n)=s(n)+d(n). \tag{1}$$

Signal processing is done on a frame-by-frame basis in the frequency domain. Speech and noise are assumed to be uncorrelated.

Let $|\hat{S}_N(f,i)|$, $|\hat{D}_N(f,i)|$, and $|\hat{Y}_N(f,i)|$ represent the estimated spectral magnitude of short-term speech, noise, and noisy speech, respectively; N is the frame length, $f \in [0,N-1]$ is a discrete variable enumerating the frequency bins, and i is a time frame index.

The noise spectral magnitude can be estimated by

$$|\overline{D}_N(f,i)|=\begin{cases}\mu|\overline{D}_N(f,i-1)|+(1-\mu)|\hat{D}_N(f,i)|, & \text{noise only;}\\ |\overline{D}_N(f,i-1)|, & \text{speech and noise;}\end{cases} \tag{2}$$

where μ is an exponential averaging time constant. $|\hat{D}_N(f,i)|$ is computed during speech pauses.

In general, the spectral subtraction operation corresponds to a time varying filtering operation,

$$|\hat{S}_N(f,i)|=G_N(f,i)|\hat{Y}_N(f,i)|. \tag{3}$$

The gain of the operation, $G_N(f,i)$, can be shown as

$$G_N(f,i)=\left(1-k\frac{|\overline{D}_N^a(f,i)|}{|\hat{Y}_N^a(f,i)|}\right)^{1/a}, \tag{4}$$

where k is the subtraction factor and a determines the sharpness of the transition from when $G(f,i)=1$ (the spectral component is not modified) to when $G(f,i)=0$ (the spectral component is suppressed).

The gain function contains a division. Therefore, an averaged lower resolution gain function, $G_M(f,i)$, should be used in order to avoid extreme variability. Both the averaged background noise spectrum estimate, $\overline{D}_M(f,i)$, and the recent frame noisy speech spectrum estimate, $\hat{Y}_M(f,i)$, employed in the computation of the gain function have block length $M \ll N$,

$$G_M(f,i)=\left(1-k\frac{|\overline{D}_M^a(f,i)|}{|\hat{Y}_M^a(f,i)|}\right)^{1/a}. \tag{5}$$

The shorter periodogram estimates are computed by using sub-blocks of the input frame, $y_L(i)$, combined with a Barlett method (Oppenheim and Schafer, 1989), which is used to decrease the variance of the estimated spectrum. This averaged gain function, $G_M(f,i)$, has a corresponding impulse response, $g_M(n,i)$, of length $M<N-L$; so the time–domain aliasing problem inherited from periodic circular convolution could be solved. In order to reconstruct a gain function that matches the number of FFT bins, N, the gain function is interpolated from the shorter gain function, $G_M(f,i)$, to form $G_{M\uparrow N}(f,i)$. The resulting output is obtained by using overlap-add and an inverse FFT of

$$|S_N(f,i)|=G_{M\uparrow N}(f,i)|\hat{Y}_{L\uparrow N}(f,i)|. \tag{6}$$

However, the gain function $G(f,i)$ with fixed subtraction factor k is unable to adapt to varying noise levels and characteristics. The nonlinear spectral subtraction algorithm (NSS) (Lockwood and Boudy, 1992), which adapts the subtraction factor k in time and frequency based on the SNR, was used in this study:

$$k(f,i)=k_c\times\frac{\max\limits_{i-20\leqslant\tau\leqslant i}(|\hat{D}_M(f,\tau)|)}{|D_M(f,i)|}}{1+\gamma\left(\frac{|\hat{Y}_M(f,i)|}{|\overline{D}_M(f,i)|}\right)}, \tag{7}$$

where k_c is a subtraction constant and γ is a scaling factor.

Then the gain function $G(f,i)$ with adaptive subtraction factor $k(f,i)$ can be shown as

$$G_M(f,i)=\left(1-k(f,i)\frac{|\overline{D}_M^a(f,i)|}{|\hat{Y}_M^a(f,i)|}\right)^{1/a}. \tag{8}$$

The reduced resolution has been used to reduce the variations of the spectrum estimates originating from the gain function, $G_M(f,i)$. The variations may be further decreased by using an adaptive exponential averaging of the gain function:

$$\overline{G}_{M,1}(f,i)=\alpha_1(i)\times\overline{G}_{M,1}(f,i-1)+(1-\alpha_1(i))\times G_M(f,i), \tag{9}$$

where $\alpha_1(i)$ is an adaptive averaging time parameter derived from a spectral discrepancy measure and $\overline{G}_{M,1}(f,i)$ is the adaptively averaged gain function. $\alpha_1(i)$ is derived from a spectral discrepancy measure, $\beta(i)$, where

$$\alpha_1(i)=1-\beta(i). \tag{10}$$

The spectral discrepancy measure, $\beta(i)$, depends on the relation between the current block spectrum, $\hat{Y}_M(f,i)$, and the current averaged noise spectrum, $\overline{D}_M(f,i)$,

$$\beta(i)=\min\left\{\frac{\Sigma_{f=0}^{M-1}||\hat{Y}_M(f,i)|-|\overline{D}_M(f,i)||}{\Sigma_{f=0}^{M-1}|\overline{D}_M(f,i)|},1\right\}. \tag{11}$$

A small spectral discrepancy yields a large $\alpha_1(i)$ and, thus, a longer averaging time of the gain function, $G_M(f,i)$. This corresponds to a stationary background noise situation. A large spectral discrepancy results in a shorter averaging time, or no averaging of the gain function. This corresponds to situations where speech or highly varying background noise is present.

When the input signal transitions from a noisy speech period to a background noise period, the averaging time parameter, $\alpha_1(i)$, must increase so that the effective averaging time of the gain function, $\overline{G}_{M,1}(f,i)$, also increases. However, directly increasing the averaging time would result in an audible "shadow voice," since the gain function suited for the speech spectrum would persist for a longer period. Accordingly, the averaging time should only be allowed to increase slowly, allowing the gain function to adapt to the stationary input (Lockwood and Boudy, 1992). Thus, by inserting a conditional averaging of the adaptive averaging time constant, $\alpha_2(i)$, of the gain function, this effect can be achieved:

$$\alpha_2(i)=\begin{cases}\gamma_c\alpha_2(i-1)+(1-\gamma_c)\alpha_1(i), & \alpha_2(i-1)<\alpha_1(i);\\ \alpha_1(i), & \text{otherwise.}\end{cases} \tag{12}$$

The variance-reduced gain function is thus given by

$$\overline{G}_{M,2}(f,i)=\alpha_2(i)\times\overline{G}_{M,2}(f,i-1)+(1-\alpha_2(i))\times G_M(f,i). \tag{13}$$

Finally, spectral flooring (Hochberg *et al.*, 1992) was used to further reduce the "musical noise":

$$\overline{G}_{M,3}(f,i)=\max(\beta_{\text{floor}},\overline{G}_{M,2}(f,i)). \tag{14}$$

Therefore,

$$|S_N(f,i)|=\overline{G}_{M\uparrow N,3}(f,i)|\hat{Y}_{L\uparrow N}(f,i)| \tag{15}$$

and

$$S_N(f,i)=|S_N(f,i)|\exp(j\times\arg(\hat{Y}_{L\uparrow N}(f,i))). \tag{16}$$

Constants used in this study include frame length $L=320$ with 160-point overlap, periodogram length $M=64$, and the FFT length $N=512$; $a=1$ in (8), $k_c=1.8$ in (7), $\gamma=0.3$ in (7), $\gamma_c=0.8$ in (12), and $\beta_{\text{floor}}=0.1$ in (14).

B. Performance evaluation

To evaluate the performance of the proposed algorithm, 7 post-lingually deafened CI subjects were tested using Hearing in Noise Test (HINT) sentences (Nilsson *et al.*, 1994) under different conditions of background noise. Table I shows relevant information for the CI subjects.

In the HINT corpus, there are 26 sets of sentences, with 10 sentences in each set; subjects were presented with 3 sets per condition. The percent correct of words-in-sentences was measured as a function of SNR for each subject, for all noise conditions; subjects were tested with and without the speech enhancement algorithm.

Two types of noise—speech-shaped noise and 6-talker speech babble—were tested in this study. CI users are generally more susceptible to fluctuating maskers (such as competing speech) than steady-state noise. The speech-shaped noise was obtained by filtering white noise with a second-order Butterworth filter (1100 Hz cutoff frequency). Noise was added to speech to obtain signal-to-noise ratios (SNRs) of +9, +6, +3, and 0 dB. The noisy speech sentences were then processed by the speech enhancement algorithm.

Sentences were presented to CI subjects in the free field via single loudspeaker (Tannoy Reveal) at 65 dBA. Subjects were tested using their clinically assigned speech processor; subjects were asked to use their normal, everyday volume/sensitivity settings.

TABLE I. Subject information for seven cochlear implant listeners who participated in the present study.

Subject	Age	Implant	CI use (yr)
S1	46	Nuc22	11
S2	62	Nuc22	13
S3	60	Nuc22	11
S4	52	Nuc22	9
S5	39	Med-El	3
S6	50	Clarion	4
S7	66	Clarion	1

III. RESULTS AND DISCUSSION

Figure 1 shows word-in-sentence recognition in the presence of speech-shaped noise, as a function of the SNR. Individual and mean CI data are shown with (open circles) and without (filled circles) the speech enhancement algorithm. For all subjects, at all SNRs, performance was better with the speech enhancement algorithm. Mean recognition scores, averaged across all subjects and all SNRs, increased by 21 percentage points. Two-way ANOVA tests revealed significant main effects of noise levels [$F(3,48)=8.186$, $p<0.001$] and algorithms [$F(1,48)=11.672$, $p=0.001$] but no significant interaction between noise levels and algorithms [$F(3,48)=0.141$, $p=0.935$].

Figure 2 shows word-in-sentence recognition in the presence of interfering speech babble, as a function of the SNR. Individual and mean CI data are shown with (open circles) and without (filled circles) the speech enhancement algorithm. The speech enhancement algorithm showed only a

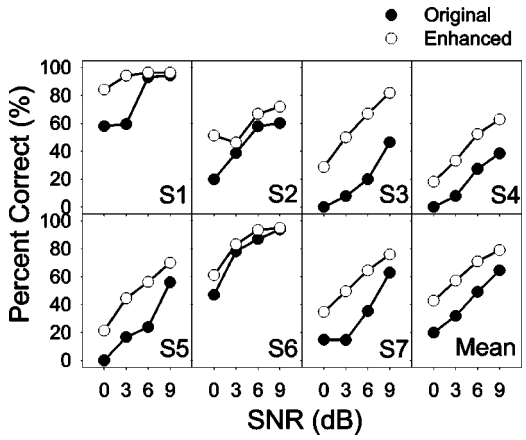


FIG. 1. Word correct percentage as a function of SNR for speech-shaped noise.

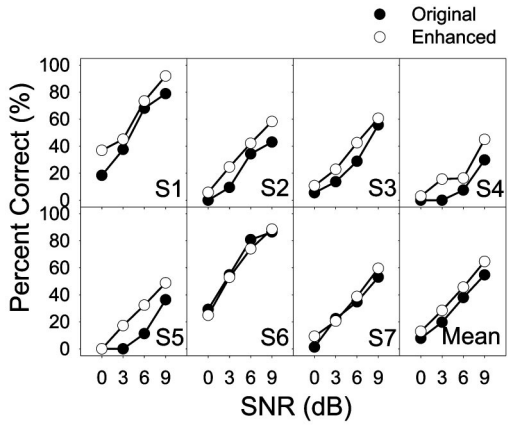


FIG. 2. Word correct percentage as a function of SNR for speech babble noise.

modest improvement in performance in the presence of speech babble. For some subjects (subjects 6 and 7), performance even slightly decreased. Mean recognition scores, averaged across all subjects and all SNRs, increased by 7.75 percentage points. Two-way ANOVA tests revealed significant main effects of noise levels [$F(3,48)=17.65$, $p<0.001$] but no significant effects of algorithms [$F(1,48)=2.387$, $p=0.129$] as well as no significant interaction between noise levels and algorithms [$F(3,48)=0.039$, $p=0.989$]. However, the paired student one-tail t-tests revealed that sentence recognition scores were significantly better with the speech enhancement algorithm at +9 dB [$t(6)=4.93$, $p=0.001$], +6 dB [$t(6)=2.37$, $p=0.028$] and +3 dB [$t(6)=2.88$, $p=0.013$].

In general, the speech enhancement algorithm improved the CI listeners' speech understanding in the presence of noise. Two possible reasons may explain why the algorithm worked much better for speech-shaped noise than for speech babble. One reason is that the time constant values in the speech pause detection algorithm were optimized according to the speech-shaped noise. The hit rate of speech pause detection was observed to be much lower when these time constant values were used with speech babble. Although the hit rate could have been increased by modifying the time constant values, the false hit rate would have also increased, causing some of the speech signal to be cancelled as if it were noise. The other reason is that, due to the rapid varying characteristics of speech babble, the noise spectrum estimated in previous signal blocks could not accurately represent the noise spectrum in the current block, where the spectral subtraction is performed. Subtracting the inaccurately estimated noise spectrum from the noisy speech spectrum would undoubtedly cause the performance to decrease. In some cases, the spectral subtraction might produce speech-like time-varying characteristics and further decrease intelligibility.

Because CI users are more susceptible to speech babble than steady noise, a set of time constant values may have to be optimized for speech babble, or another speech enhancement algorithm suitable for competing speech may need to be developed. CI users would then be able to switch from one algorithm to another, depending on the demands of the listening situation.

ACKNOWLEDGMENTS

We appreciate the efforts of all subjects. We also would like to thank John J. Galvin III for assistance in editing the manuscript. The research was supported by Grants No. R01-DC04993 from NIDCD.

Kiefer, J., Müller, J., Pfennigdorff, T., Schön, F., Helms, J., von Ilberg, C., Baumgartner, W., Gstöttner, W., Ehrenberger, K., Arnold, W., Stephan, K., Thumfart, W., and Baur, S. (**1996**). "Speech understanding in quiet and in noise with the CIS speech coding strategy (MED-EL Combi-40) compared to the multipeak and spectral peak strategies (Nucleus)," ORL **58**, 127–135.

Müller-Deile, J., Schmidt, B. J., and Rudert, H. (**1995**). "Effects of noise on speech discrimination in cochlear implant patients," Ann. Otol. Rhinol. Laryngol. Suppl. **166**, 303–306.

Fetterman, B. L., and Domico, E. H. (**2002**). "Speech recognition in background noise for cochlear implant patients," Otolaryngol.-Head Neck Surg. **126**, 257–263.

Hochberg, I., Boothroyd, A., Weiss, M., and Hellman, S. (**1992**). "Effects of noise and noise suppression on speech perception by cochlear implant users," Ear Hear. **13**, 263–271.

Boll, S. F. (**1979**). "Suppression of acoustic noise in speech using spectral subtraction," IEEE Trans. Acoust., Speech, Signal Process. **27**, 113–120.

Ephraim, Y., and Malah, D. (**1984**). "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," IEEE Trans. Acoust., Speech, Signal Process. **32**, 1109–1121.

Gustafsson, H., Nordholm, S. E., and Claesson, I. (**2001**). "Spectral subtraction using reduced delay convolution and adaptive averaging," IEEE Trans. Speech Audio Process. **9**, 799–807.

Lockwood, P., and Boudy, J. (**1992**). "Experiments with a nonlinear spectral subtractor (NSS), hidden Markov models and projection, for robust recognition in cars," Speech Commun. **11**, 215–228.

Berouti, M., Schwartz, R., and Makhoul, J. (**1979**). "Enhancement of speech corrupted by acoustic noise," in *Proc. IEEE ICASSP*, pp. 208–211.

Marzinzik, M., and Kollmeier, B. (**2002**). "Speech pause detection for noise spectrum estimation by tracking power envelope dynamics," IEEE Trans. Speech Audio Process. **10**, 109–117.

Oppenheim, A. V., and Schafer, R. W. (**1989**). *Discrete-Time Signal Processing* (Prentice–Hall, Englewood Cliffs, NJ).

Nilsson, M., Soli, S. D., and Sullivan, J. A. (**1994**). "Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise," J. Acoust. Soc. Am. **95**, 1085–1099.

A beamforming video recorder for integrated observations of dolphin behavior and vocalizations (L)

Keenan R. Ball[a)]
Woods Hole Oceanographic Institution, Dept 4, MS 18, 86 Waters Street, Woods Hole, Massachusetts 02543

John R. Buck[b)]
Department of Electrical and Computer Engineering & School for Marine Science and Technology, University of Massachusetts Dartmouth, 285 Old Westport Road, North Dartmouth, Massachusetts 02747-2300

(Received 11 August 2003; revised 26 August 2004; accepted 7 October 2004)

In this Letter we describe a beamforming video recorder consisting of a video camera at the center of a 16 hydrophone array. A broadband frequency-domain beamforming algorithm is used to estimate the azimuth and elevation of each detected sound. These estimates are used to generate a visual cue indicating the location of the sound source within the video recording, which is synchronized to the acoustic data. The system provided accurate results in both lab calibrations and a field test. The system allows researchers to correlate the acoustic and physical behaviors of marine mammals during studies of social interactions. © *2005 Acoustical Society of America.*
[DOI: 10.1121/1.1831284]

PACS numbers: 43.80.Ev, 43.80.Ka, 43.80.Jz, 43.30.Sf [WA] Pages: 1005–1008

I. INTRODUCTION

Many cetacean species are acoustically active in social contexts. Additionally, these species spend a majority of their lives underwater where it is difficult for human researchers to observe them. These combined factors challenge researchers studying the physical and acoustic behavior of individual animals during social interactions. Bottlenose dolphins (*Tursiops truncatus*) and other smaller cetaceans increase these challenges by interacting in complex fission–fusion social structures, with animals in close physical proximity. Associating acoustic signals with individual dolphins is essential to studies of acoustic repertoires, juvenile acoustic development,[1] social alliances,[2] and the signature whistle hypothesis.[3] Ideally, observations should be made with minimal perturbations to the animals' natural behavior and environment. Moreover, the observations should be archival in the sense that subsequent investigators should be able to see and hear the animals' behaviors directly in their original form, and not have to rely on the original observers' detections and classifications of the behaviors. Data in this form allow both independent confirmation of behavioral hypotheses and reanalysis in light of subsequently proposed alternative hypotheses. Lastly, the system should be portable for situations where the animals do not remain in a single location.

Previous techniques used to associate sounds with individual marine mammals include emitted bubble streams (e.g., Ref. 4) isolating animals (e.g., Refs. 1, 3), tags (e.g., Refs. 5, 6), and hydrophone arrays (e.g., Refs. 7, 8). None of these approaches produced archival video footage, relying on human observers to link the animals to the sounds made. Three recent systems[9–11] integrate hydrophone arrays with video recordings to produce archival observations, but still differ in significant ways from the system described in this Letter. The system in Ref. 10 uses only two hydrophones spaced at five times the human intra-aural distance to provide coarse localization cues (left, right, or both) while reviewing the video recordings. The system provides no vertical resolution cues.

Thomas *et al.*[11] combined an elevated video camera with a distributed array of eight hydrophones around a lagoon perimeter to observe behavior. The location of each sound source was determined by cross-correlating the hydrophone signals to estimate the relative arrival times of the sounds at the array. The resulting location estimate was projected into the video image. Calibration tests determined an accuracy of roughly 2 m for the system. While the system produces large-scale archival overhead video records of the physical and acoustic behaviors of the animals with little or no perturbation of the observed animals' behavior, the limited accuracy makes it impossible to discern acoustic behavior among closely spaced dolphins, e.g., mother–calf pairs, and the need for the manual selection of sounds makes the video post-processing labor intensive.

Au and Herzing[9] developed a two-dimensional Y-shaped four hydrophone array including a video camera to study Atlantic spotted dolphin (*Stenella frontalis*) echolocation clicks. This system used differences in the clicks' arrival times to estimate the range to the echolocating dolphin, but did not estimate the bearing to the animal.[12] Because it was designed solely for click analysis, the system has memory limitations that prevent it from recording and analyzing entire whistles or tracking a whistling dolphin moving across multiple video frames.

In this Letter we describe a beamforming video recorder (BVR) designed to produce archival video recordings of complicated social interactions including whistles and

[a)]Formerly at the Department of Electrical and Computer Engineering, University of Massachusetts, Dartmouth, 285 Old Westport Road, North Dartmouth, Massachusetts 02747-2300; electronic mail: kball@whoi.edu
[b)]Electronic mail: johnbuck@ieee.org

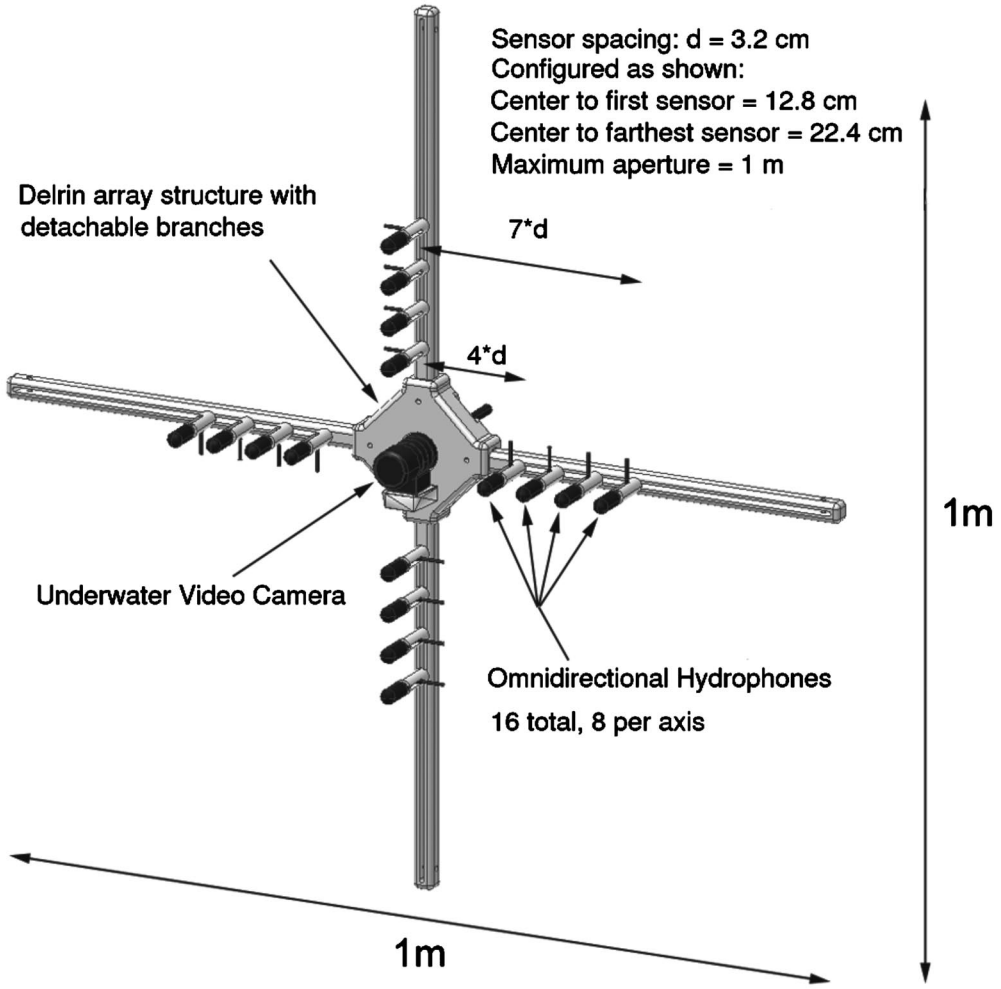


FIG. 1. Two-dimensional handheld audio/video array with detachable branches. The array consists of 16 hydrophones, 8 per axis, and an underwater video camera located at the center. The diver hand holds, mounting brackets, and tether are not shown.

echolocation clicks. The video recordings include visual cues indicating the location of each sound observed, tracking across video frames where necessary. The BVR provides short range underwater observations with a high resolution beamformer that can potentially distinguish among several vocally active dolphins in view of the camera.

II. SYSTEM DESCRIPTION

The BVR consists of a short-baseline Mills cross 16 hydrophone array incorporating a video camera at the array center and associated audio and video recording hardware. Placing the video camera at the array center gives a common reference frame for the bearing estimates and the video image, removing the need for field recalibration and registration as in Ref. 11. The system is largely comprised of off-the-shelf parts, minimizing the need for custom hardware. The acoustic signals are processed using broadband frequency–domain beamforming algorithms, allowing for automated processing and independent position estimates for each video frame. The visual cues incorporated into the video record can track a whistling animal as it moves across the image. The resulting video stream is an archival record linking the acoustic and physical behaviors of small cetaceans in dense social contexts.

A. Hydrophone array and camera

Figure 1 is a diagram of the BVR array and camera with the hydrophones configured in the reduced aperture as used for the tests described in Sec. III B. The array is constructed of Delrin with detachable branches. Each of the four arms is 45.7 cm long by 3.2 cm square, holding four hydrophones with integrated 10 dB preamplifiers (High Tech Inc., HTI-96-MIN). The hydrophones can be reconfigured to suit recording conditions using a slot in each arm. Calibrations confirmed that the phase variations among the hydrophones produce a negligible bias on the beamforming results. The array's center piece is 17.8 cm tall by 18.4 cm wide by 4.4 cm deep. A 5 cm diameter hole in this piece mounts a DeepSea 2050 (DeepSea Power and Light, San Diego, CA) underwater camera at the array center. The maximum total aperture of the array is roughly 1 m, making it convenient to deploy from a dock or moving boat in an inverted periscope configuration, or to be maneuvered by a swimmer. Due to the low visibility during the field test described in Sec. III B, the hydrophones were positioned at ± 13.3, ± 16.5, ± 19.7, and ± 22.9 cm on each axis relative to the origin at the array center, giving a total aperture of about 46 cm.

B. Recording system

The 16 hydrophones are connected through custom breakout boxes to a Tascam MX-2424 multitrack hard disk recorder. The MX-2424 synchronously records all 16 channels at 44.1 kHz with 24 bit resolution directly to its internal SCSI hard disk. An additional channel on the MX-2424 records a sync signal that is also recorded on the left channel of the video camera. This allows the video and acoustic array recordings to be synchronized with a precision of 1/44.1 kHz $=22.7\ \mu$s. The video signal from the DeepSea 2050 underwater camera is recorded by a Sony DCR-TRV-530 Digital 8 camcorder. To simplify processing, the video is converted from the NTSC 29.97 frames-per-second DF (Drop Frame) video standard to the 30 frames-per-second ND (Non-Drop Frame) standard. At the 30 Hz frame rate, there are exactly 1470 samples of 44.1 kHz acoustic data per video frame.

C. Beamforming algorithm

The recorded data is processed in Matlab (The Mathworks, Natick, MA) to synthesize video recordings with localization cues. Whistles and echolocation clicks were detected by comparing the average acoustic energy in each 1470 sample frame against an empirically determined threshold value. The azimuth and elevation of each detected sound are estimated using the broadband frequency–domain algorithm for sparse arrays in Ref. 13. Equivalent angular resolution could be obtained using an array of only two hydrophones, however, the multiple hydrophone array geometry employed reduces the number of grating lobes and attenuates the amplitudes of the sidelobes in the frequency–wave number response or beam pattern (Ref. 14, Sec. 2.2). Consequently, the BVR rejects noise from undesired bearings more robustly than could a two hydrophone per axis array. The system beamforms the array data at the peak frequency in each block, and also at harmonics of the peak frequency. Frequency–domain beamforming allows automated processing of the acoustic data, updating the position estimate 30 times a second, which is not possible for systems that use time–domain autocorrelations on an entire whistle such as Ref. 11. The estimated elevation and azimuth angles from the beamformer were converted to pixels in the video image using conversions of 11.083 pixels/degree in azimuth and 8.727 pixels/deg in elevation established during calibration tests in the SMAST Acousto-Optic tank. A localization cue of a small + sign was spliced into each frame of video based on these pixel coordinates. Elevations or azimuths just offscreen were indicated by highlighting the nearest border of the video image.

III. RESULTS

A. Calibration tests

The BVR was calibrated in both indoor tank tests and a free swimming outdoor test. The tank tests transmitted five cycles of a 5 kHz sinusoid with three harmonics (10, 15, and 20 kHz) from a source at a known fixed range and bearing. These tests established that for the array geometry in Sec. II A, any source at a range of 1 m or more was sufficiently in the farfield to produce an accurate bearing estimation. The tests varied the elevation and azimuth of the source over the camera's field of view, and the BVR nearly always placed the + cue on the source or within five pixels of the source. This is an error of less than 4 cm even at a 4 m range. It is also possible to use the BVR to triangulate the range to a source using the two outermost sensors on each arm to obtain four separate bearing estimates. When the hydrophones are set to provide the maximum aperture (about 1 m, twice that used in Sec. III B), this triangulation approach produced range estimates accurate to $\pm 10\%$ for ranges of 1–4 m. The free swimming test was conducted off the UMass Dartmouth SMAST pier in water roughly 3–4 m deep over a bottom with a mix of rocks and sand. The source played recorded dolphin whistles with three to four harmonics. The video cue tracked the source closely as the current and the swimmer's motion caused the source to move within the video frame.

B. Field test

The BVR field test took place at the Dolphin Connection at Hawk's Cay Resort in Duck Key, Florida. The Dolphin Connection was chosen as the testing site because of its natural lagoon setting, the number of well-trained dolphins, and the representative acoustic conditions including snapping shrimp and nearby boat traffic. The sand and coral bottom of the lagoon provide realistic boundary conditions for acoustic propagation. The seaward edges of the lagoon are formed by plastic fencing secured to wooden posts embedded in the bottom. The dolphins generally acclimatize well to the novel activities because the Dolphin Connection frequently hosts research projects.

The dolphin training staff suggested that a dummy array be used to desensitize the dolphins during the months preceding the experiments using the actual array. The dummy array, constructed out of Delrin to closely resemble the actual array, was installed in March 2002. The array was mounted to the dock with a clamping mount holding a 1.2 m PVC tube with a 0.61 m horizontal tube glued on the top end. A buckled strap closed the clamp mount while allowing the rapid release and removal of the array in the event of an adverse reaction. The array could be rotated in order to aim the camera at any activity in the lagoon.

The field test took place during the last week of May and the first week of June 2002. Most testing took place during 1 h sessions in the morning before the trainers began their daily morning cleaning and feeding routine. We obtained useful data on five of seven days of deployment. In total, over 145 minutes of data were processed, including over 57 distinct whistles. The narrowband signal-to-noise ratio (SNR) at the fundamental frequency was at least 80 dB in all of the data processed. For all 57 whistles, the cue was placed on or very close to (two to three pixels) the dolphin's head, indicating a maximum error of less than 0.5 deg of elevation or azimuth. Figure 2 shows a still frame from the video recorded on 28 May 2002 observing a adolescent male dolphin named Kai when he was the only dolphin in the lagoon. The contrast of this image was increased from the original color image for a black and white publication. More video data including the sample frame presented in Fig. 2 are available

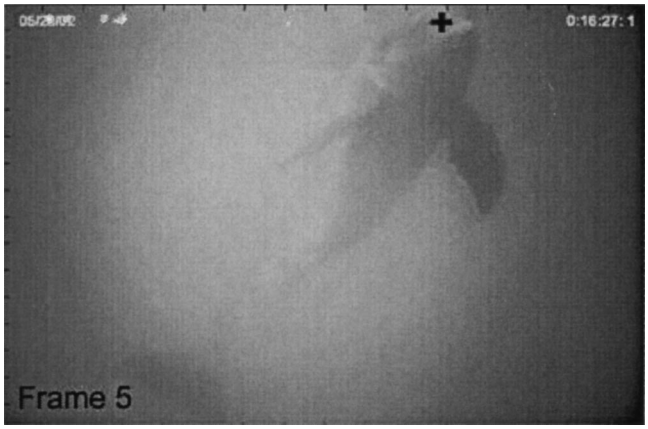


FIG. 2. Representative still frame from the BVR output, recorded on 28 May 2002 showing Kai whistling. The cue accurately followed Kai's head throughout a 3 s whistle including many echolocation clicks. The original color video frame has been digitally retouched to improve the contrast for a black and white publication.

at the website indicated in Ref. 15. In all trials, the BVR performed well through interference such as snapping shrimp, boat noise, and echolocation clicks. The high SNRs in the recordings meant that the limit on the BVR's performance was generally underwater visibility (4–5 m), and not the beamforming algorithm.

IV. DISCUSSION

The BVR performed very well monitoring dolphins in a natural lagoon. In almost every test, the cue was placed on the head of a dolphin in the video image. In most tests, the few erroneous cues, if any, were single frame transients and did not generate confusion about the identity of the whistling animal. The only whistle series that consistently confounded the BVR was when the whistling dolphin swam beneath the dock along the fence, causing a complicated multipath arrival structure. Although we never observed a whistle exchange between two dolphins, the BVR's consistent precision placing the cue on the dolphin's head gives us confidence that the system will perform well when such an event is observed.

Our efforts to desensitize the dolphins to the array's presence were only partially successful, as the animals often focused their behavior on the array, scientists, and trainers. The animals' focus on the array was useful for our engineering test, providing a substantial dataset in a short period, but would be a drawback in any behavioral experiment. In such a test, the animals must be more thoroughly desensitized to the array's presence, or perhaps the array should be physically isolated by fencing and deployed unattended to minimize the animal's interest.

In conclusion, video recordings incorporating visual cues indicating the origins of sounds are a powerful tool for studies linking physical and acoustic behavior in cetaceans. The BVR system presented in this paper is a portable system suitable for observing highly active social contexts and producing archival recordings of these observations. The BVR holds promise for studying open questions in bottlenose dolphin behavior.

ACKNOWLEDGMENTS

This paper is dedicated to the memory of Gina Wood, the head trainer at the Dolphin Connection. Gina generously and creatively contributed to the planning and execution of the field test. This work would not have been possible without the cooperation and contributions of the training staff at the Dolphin Connection. Doug and Cheryl Messinger generously allowed us access to the animals at their facility. Jason Davis, Lisa Davis, Amy Dobelle, Adrien Domske, and Sylvia Rickett patiently came in early and stayed late to help us with our field test. Conversations with Peter Tyack inspired the BVR. Larry Reinhart and Rob Fisher at UMass Dartmouth gave valuable advice and assistance in the design and calibration of the BVR. Will Moore machined the BVR. Two reviewers made numerous helpful suggestions to improve this Letter. This research was funded by NSF Ocean Sciences CAREER award 9733391. This Letter is UMass Dartmouth School for Marine Science and Technology Contribution No. 040301.

[1] L. S. Sayigh, P. L. Tyack, R. S. Wells, and M. D. Scott, "Signature whistles of free-ranging bottlenose dolphins *Tursiops truncatus*: stability and mother-offspring comparisons," Behav. Ecol. Sociobiol. **26**, 247–260 (1990).

[2] R. C. Connor, M. R. Heithaus, and L. M. Barre, "Complex social structure, alliance stability and mating access in a bottlenose dolphin 'super-alliance'," Proc. R. Soc. London **268**, 263–267 (2001).

[3] M. C. Caldwell and D. K. Caldwell, "Individualized whistle contours in bottle-nosed dolphins (*Tursiops truncatus*)," Nature (London) **207**, 434–435 (1965).

[4] M. E. Dahlheim and F. Awbrey, "A classification and comparison of vocalizations of captive killer whales (*Orcinus orca*)," J. Acoust. Soc. Am. **72**, 661–670 (1982).

[5] M. Johnson and P. L. Tyack, "A digital acoustic recording tag for measuring the response of wild marine mammals to sound," IEEE J. Ocean. Eng. **28**, 3–12 (2003).

[6] P. L. Tyack, "An optical telemetry device to identify which dolphin produces a sound," J. Acoust. Soc. Am. **78**, 1892–1895 (1985).

[7] W. A. Watkins and W. E. Schevill, "Sound source location by arrival times on a non-rigid 3-dimensional hydrophone array," Deep-Sea Res. **19**, 691–706 (1972).

[8] C. W. Clark, "A real-time direction finding device for determining the bearing to the underwater sounds of southern right whales, *Eubalaena australis*," J. Acoust. Soc. Am. **68**, 508–511 (1980).

[9] W. W. L. Au and D. L. Herzing, "Echolocation signals of wild atlantic spotted dolphin (*Stenella frontalis*)," J. Acoust. Soc. Am. **113**, 598–604 (2003).

[10] K. M. Dudzinski, C. W. Clark, and B. Wursig, "A mobile video/acoustic system for simultaneous underwater recording of dolphin interactions," Aq. Mam. **21**, 187–193 (1995).

[11] R. E. Thomas, K. M. Fristrup, and P. L. Tyack, "Linking the sounds ofdolphins to their locations and behavior using video and multichannel acoustic recordings," J. Acoust. Soc. Am. **112**, 1692–1701 (2002).

[12] W. W. L. Au (personal communication, 2003).

[13] A. Thode, T. Norris, and J. Barlow, "Frequency beamforming of dolphin whistles using a sparse three-element towed array," J. Acoust. Soc. Am. **107**, 3581–3584 (2000).

[14] H. L. Van Trees, "Optimum array processing," Part IV of *Detection, Estimation and Modulation Theory* (Wiley, New York, 2002).

[15] See EPAPS Document No. EPAPS-JASMAN-117-705501 for downloadable files containing the video files showing the beamforming video recorder output as it tracks whistles and clicks from a dolphin and places a + cue on the video image based on the beamformed array data. A direct link to this document may be found in the online article's HTML reference section. The document may also be reached via the EPAPS homepage (http://www.aip.org/pubservs/epaps.html) or from ftp.aip.org in the directory /epaps. See the EPAPS homepage for more information.

An efficient method for evaluating diffuse field joint acceptance functions for cylindrical and truncated conical geometries

Jean-Pierre Coyette,[a] Grégory Lielens, and Mickaël Robbé
Free Field Technologies S.A., Place de l'Université, 16 B1348 Louvain-la-Neuve, Belgium

Pascale Neple
Airbus-France, Acoustic and Environmental Department, Route de Bayonne, 316 F31060 Toulouse Cedex 03, France

(Received 18 May 2004; revised 30 November 2004; accepted 30 November 2004)

The evaluation of the response of elastic structures subjected to distributed random excitations is usually performed in the modal space. Random excitations (like acoustic diffuse fields) are usually modeled as weakly stationary random processes and are assumed to be homogeneous. Their characterization basically relies on the power spectral density (PSD) function of the pressure at a particular reference position and a suitable spatial correlation function. In the modal space, the distributed random excitation is characterized by a modal PSD matrix made from the joint acceptance functions related to the mode pairs. The joint acceptance function is a double surface integral involving the product of the considered mode shapes and the spatial correlation function. The paper shows how to evaluate efficiently this quadruple integral for cylindrical and truncated conical structures excited by an acoustic diffuse field. Basically, the procedure relies on the derivation of alternative expressions for the spatial correlation function. The related expressions prove to be more convenient for these geometries and are leading to a reduction of the double surface integral to a combination of simple integrals. A very substantial breakdown of the computational cost can be achieved using the resulting expressions. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1850368]

PACS numbers: 43.20.−f, 43.28.Lv, 43.40.Ey, 43.40.Hb, 43.20.Bi [NX] Pages: 1009–1019

I. INTRODUCTION

The study and the optimization of vibro-acoustic performances are important challenges for aeronautical structures (cabin and cockpit, for example). Usually these structures are modeled as canonical (cylindrical or truncated conical) shapes. The assessment of their noise transmission characteristics usually refers to acoustic diffuse field excitations. The target frequency range (typically from 40 Hz up to 12.5 kHz) prevents the use of numerical methods which are basically restricted to the ''low'' frequency range. In such circumstances, analytical solution procedures prove to be useful especially if they can rely on modal descriptions. The acoustic diffuse field, on the other hand, is usually modeled as a weakly stationary random process characterized by its power spectrum at a reference position and a spatial correlation function. In a modal context, the distributed nature of the excitation requires the evaluation of a joint acceptance function for each mode pair.[1,2] The joint acceptance function is a double surface (i.e., quadruple) integral involving the product of the mode shapes and the spatial correlation function. In the most general case of nonregular (i.e., arbitrary) surfaces, the evaluation of acceptance functions is usually performed by numerical integration and can represent a significant computational cost. In the particular case of cylindrical or truncated conical surfaces, we show how to reduce the quadruple integral to a simple integral or a product of simple integrals. In both cases, the reduction procedure relies on an appropriate re-derivation of the spatial correlation function of the diffuse field. In contrast with earlier studies[3] where the spatial correlation function was assumed to be separable along longitudinal and transversal directions, the present approach still refers to the exact correlation function of a perfectly diffuse field.

The outline of the paper is as follows. In Sec. II, the handling of distributed random excitations in a modal context is summarized. Sections III and IV derive the original forms of the spatial correlation function and the joint acceptance function for the cylindrical and truncated conical cases. Section V is related to the formulation of the spatial correlation function in two alternative forms that prove to be more convenient for evaluating the joint acceptance function. Section VI deals with the evaluation of resulting integrals while Sec. VII presents numerical results which further validate the reduction procedure.

II. MODAL APPROACH FOR HANDLING DISTRIBUTED RANDOM EXCITATIONS

A. Deterministic input/output relation

We consider the evaluation of the time-harmonic response (time dependence like $e^{i\omega t}$ where ω is the circular frequency) of a linear elastic structure subjected to a distributed load. The deterministic response $u(\zeta,\omega)$ at a point ζ and at frequency ω can be obtained as a linear combination of eigenmodes ψ_m:

[a]Electronic mail: jean-pierre.coyette@fft.be

$$u(\zeta,\omega)=\sum_{m=1}^{\infty}\psi_m(\zeta)y_m(\omega), \tag{1}$$

where y_m is the participation factor of mode m.

In practice, the series is truncated after M modes so that Eq. (1) can be rewritten in the matrix form:

$$u(\zeta,\omega)=\psi(\zeta)y(\omega), \tag{2}$$

where ψ is the matrix $(3\times M)$ of mode shapes while y is the vector $(M\times 1)$ of the modal participation factors.

The equation of motion in the modal space can be written as

$$Z(\omega)y(\omega)=x(\omega) \tag{3}$$

or

$$y(\omega)=H(\omega)x(\omega), \tag{4}$$

where $Z(\omega)$ and $H(\omega)$ are the modal dynamic stiffness matrix and the modal receptance matrix, respectively, while the modal load vector $x(\omega)$ can be related to the distributed pressure p along the boundary Γ through

$$x(\omega)=\int_{\Gamma}p(\zeta,\omega)\mathbf{n}(\zeta)\cdot\psi(\zeta)d\Gamma(\zeta), \tag{5}$$

where $\mathbf{n}$ is the local normal to the boundary surface.

The modal dynamic stiffness matrix is diagonal:

$$Z_{mn}=(\omega_m^2+2i\xi_m\omega\omega_m-\omega^2)\delta_{mn}, \tag{6}$$

where ω_m is the circular frequency of mode m and ξ_m is the related viscous damping ratio.

B. Random distributed excitation

If the distributed pressure loading is described as a weakly stationary random process,[1] its main characteristic is the cross-correlation function $R_{pp'}$ of pressures p and p' at two particular points ζ and ζ' along the boundary Γ:

$$R_{pp'}(\tau)=E[p(\zeta,t)p(\zeta',t+\tau)], \tag{7}$$

where E is the mathematical expectation operator.

The assumption of a ''weakly'' stationary process means that the statistical descriptors (up to the second order) of the considered process depend only on the time shift τ between the two observations and not on the absolute time values.

The cross power spectral density function $S_{pp'}(\omega)$ is the Fourier transform of $(1/2\pi)R_{pp'}(\tau)$. For an homogeneous process, this function can be presented as the product of a power spectrum S_{ref} at a reference point and a particular correlation function f_c:

$$S_{pp'}(\omega)=S_{\mathrm{ref}}(\omega)f_c(\zeta,\zeta',k). \tag{8}$$

The correlation function f_c involves the relative position of the two points and the wave number k related to the frequency ω and the sound speed c through

$$k=\frac{\omega}{c}. \tag{9}$$

Since the modal load vector is related to the distributed pressure through Eq. (5), one can easily formulate the matrix S_x of power spectral densities of modal loads as

$$S_x(\omega)=S_{ref}(\omega)J(k), \tag{10}$$

where J is the so-called joint acceptance matrix (M by M matrix) given by

$$J(k)=\int_{\Gamma}\int_{\Gamma}\psi^T(\zeta)\cdot\mathbf{n}^T(\zeta)f_c(\zeta,\zeta',k)\mathbf{n}(\zeta')\cdot\psi(\zeta')\times d\Gamma(\zeta)d\Gamma(\zeta'). \tag{11}$$

C. Evaluation of the random response

The random response is first evaluated in the modal space and then resolved in the physical space. The power spectral density (PSD) matrix S_y of the modal response is obtained as

$$S_y(\omega)=H^*(\omega)S_x(\omega)H(\omega), \tag{12}$$

where an asterisk $(*)$ denotes the complex conjugate operator.

Owing to the linear relation [Eq. (2)] between the physical response u and the modal response y, the PSD matrix S_u of displacements is computed from S_y using

$$S_u(\omega)=\psi^T(z)S_y(\omega)\psi(z). \tag{13}$$

III. SPATIAL CORRELATION FUNCTION OF AN ACOUSTIC FIELD

A. Acoustic diffuse field

An acoustic diffuse field is obtained by summing up the effect of an infinite number of plane waves originating from all spatial directions and having arbitrary (random) phases. In practice, diffuse field conditions are obtained by activating a set of uncorrelated acoustic sources in a reverberant chamber.[4] The multiple reflections along rigid wall boundaries allow for the generation of a diffuse field. This conceptually simple field proves to be useful for defining a particular class of acoustic excitations frequently encountered in practice for vibrational or acoustic transmission studies. It is usually assumed that the tested structures located in the reverberant chamber do not participate to the generation of the diffuse field. This assumption could not be valid in some particular cases (structure volume significant versus chamber volume, convex structure). Additionally the size of the structure relative to the wavelength could also have a strong influence on the validity of this assumption.

The spatial correlation function related to a diffuse field can be easily formulated by considering an infinite number of plane waves. This derivation can be found in many standard textbooks and leads to the following result:[5]

$$f_c(\zeta,\zeta',k)=\frac{\sin(kd)}{kd}, \tag{14}$$

where $d=|\zeta-\zeta'|$ is the distance between the two points.

B. Cylindrical and truncated conical geometries

If the boundary surface Γ is cylindrical or truncated conical, a particular coordinate system is usually selected.

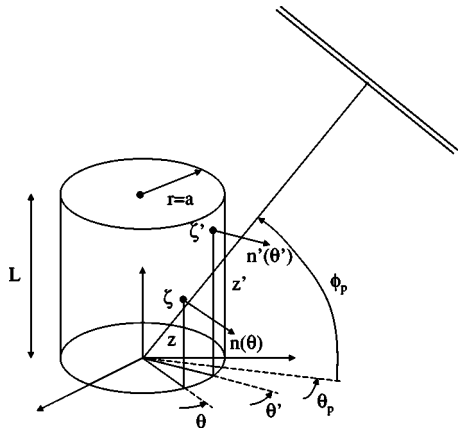


FIG. 1. Cylindrical surface and related coordinate system.

For cylindrical shells, cylindrical coordinates (r,θ,z) (Fig. 1) are a natural choice for handling eigenmodes. A cylindrical surface of radius a corresponds to

$$r=a. \tag{15}$$

The spatial correlation function related to two points $\zeta=(a,\theta,z)$ and $\zeta'=(a,\theta',z')$ along this cylindrical surface can be expressed as

$$f_c(\zeta,\zeta',k)\equiv f_c^{\mathrm{cyl}}(\theta,z;\theta',z';k) = \frac{\sin(k\sqrt{2a^2(1-\cos(\theta-\theta'))+(z-z')^2})}{k\sqrt{2a^2(1-\cos(\theta-\theta'))+(z-z')^2}}. \tag{16}$$

For truncated conical geometries, a spherical coordinate system (r,θ,ϕ) (Fig. 2) is usually selected. A truncated conical surface corresponds to

$$\phi=\alpha, \tag{17}$$

where α is the semi-vertex angle.

The spatial correlation function related to two points $\zeta=(r,\theta,\alpha)$ and $\zeta'=(r',\theta',\alpha)$ along such a surface takes the following form:

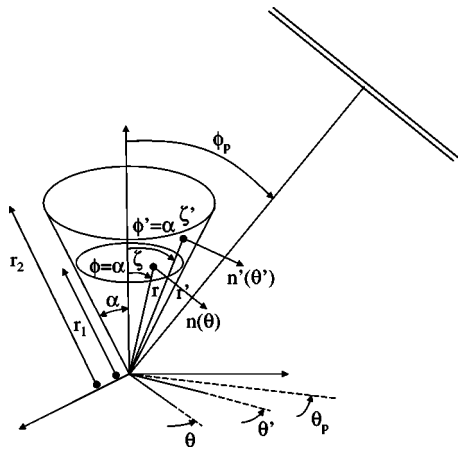


FIG. 2. Truncated conical surface and related coordinate system.

$$f_c(\zeta,\zeta',k)\equiv f_c^{tc}(r,\theta;r',\theta';k) = \frac{\sin(k\sqrt{2rr'\sin^2\alpha(1-\cos(\theta-\theta'))+(r-r')^2})}{k\sqrt{2rr'\sin^2\alpha(1-\cos(\theta-\theta'))+(r-r')^2}}. \tag{18}$$

IV. JOINT ACCEPTANCE FUNCTIONS FOR CYLINDRICAL AND TRUNCATED CONICAL SHELLS

The joint acceptance function is built from available mode shapes. For a cylindrical shell of length L with rigid end diaphragms, mode shapes can be expressed by referring to circumferential $(n\geqslant 0)$ and axial $(s\geqslant 1)$ modal orders:

$$\psi_{ns}(\theta,z)\cdot\mathbf{n}(\theta)=\cos(n\theta)\sin\left(\frac{s\pi z}{L}\right), \tag{19}$$

where $0\leqslant\theta\leqslant 2\pi$ and $0\leqslant z\leqslant L$.

The joint acceptance function for modes (n,s) and (m,t) can be expressed as

$$J^{\mathrm{cyl}}_{(n,s),(m,t)}(k)=\int_0^L\int_0^L\int_0^{2\pi}\int_0^{2\pi} f_c^{\mathrm{cyl}}(\theta,z;\theta',z';k) \times\cos(n\theta)\cos(m\theta')\sin\left(\frac{s\pi z}{L}\right) \times\sin\left(\frac{t\pi z'}{L}\right)a^2\,d\theta\,d\theta'\,dz\,dz', \tag{20}$$

where f_c^{cyl} is given by Eq. (16).

For a truncated conical shell, kinematical functions[6] (which do not possess the orthogonality properties of modes) form a convenient basis for setting up a variational approach. These kinematical functions are expressed by referring to circumferential $(n\geqslant 0)$ and axial $(s\geqslant 1)$ orders:

$$\psi_{ns}(r,\theta)\cdot\mathbf{n}(\theta)=r^2\cos(n\theta)\sin\left(\frac{s\pi(r-r_1)}{r_2-r_1}\right), \tag{21}$$

where $0\leqslant\theta\leqslant 2\pi$ and $r_1\leqslant r\leqslant r_2$ (Fig. 2).

In the sequel, these kinematical functions will be referenced as "modes." The joint acceptance function for "modes" (n,s) and (m,t) can be expressed as

$$J^{tc}_{(n,s),(m,t)}(k)=\int_{r_1}^{r_2}\int_{r_1}^{r_2}\int_0^{2\pi}\int_0^{2\pi} f_c^{tc}(r,\theta;r',\theta';k) \times\cos(n\theta)\cos(m\theta')r^3\sin\left(\frac{s\pi(r-r_1)}{r_2-r_1}\right)r'^3 \times\sin\left(\frac{t\pi(r'-r_1)}{r_2-r_1}\right)\sin^2\alpha\,d\theta\,d\theta'\,dr\,dr', \tag{22}$$

where f_c^{tc} is given by Eq. (18).

It can be demonstrated easily that the joint acceptance function is zero when $n\neq m$ for both cylindrical and truncated conical geometries.

Inspection of Eqs. (20) and (22) shows that the spatial correlation functions as described by Eqs. (16) and (18) are not appropriate for an efficient evaluation of the joint acceptance function since all integration variables are coupled. More suitable forms are derived in the next section.

V. ALTERNATIVE EVALUATION METHODS

A. Proposed strategy

The strategy proposed for achieving a reduction of joint acceptance functions to a simple form relies first on the derivation of alternative expressions for the spatial correlation function. Such expressions prove to be more convenient for axisymmetric geometries and, more specifically, for cylindrical and truncated conical cases. For this purpose, the spatial correlation function is re-derived assuming that the diffuse field results from an infinite number of plane waves originating from all spatial directions and having arbitrary phases. Each plane wave is decomposed in an appropriate way using cylindrical or spherical coordinates. The resulting expressions of the spatial correlation functions are then injected in the joint acceptance function in order to get a more suitable form.

B. Cylindrical case

The incident field due to a plane wave of unit amplitude with incidence (θ_p,ϕ_p) can be developed in cylindrical coordinates (Fig. 1) (r,θ,z):

$$p_{\theta_p,\phi_p}(r,\theta,z)=\sum_{q=0}^{\infty}\epsilon_q(-i)^q J_q(kr\cos\phi_p)\times\cos(q(\theta-\theta_p))e^{-ikz\sin\phi_p}, \tag{23}$$

where J_q is the Bessel function of order q and ϵ_q is the Neumann factor ($\epsilon_q=1$ for $q=0$, $\epsilon_q=2$ for $q\neq 0$). The basis of such expansion can be found in Refs. 7 and 8.

The spatial correlation function related to two points along the cylindrical surface of radius a can be obtained by summing the effects of plane waves with variable incidences ($0<\theta_p<2\pi, -\pi/2<\phi_p<\pi/2$):

$$f_c^{\mathrm{cyl}}(\theta,z;\theta',z';k)=\frac{1}{4\pi}\int_0^{2\pi}\int_{-\pi/2}^{\pi/2} p_{\theta_p,\phi_p}(a,\theta,z)\times p^*_{\theta_p,\phi_p}(a,\theta',z')\times\cos\phi_p\,d\phi_p\,d\theta_p. \tag{24}$$

Substitution of Eq. (23) into Eq. (24) leads to the following result:

$$f_c^{\mathrm{cyl}}(\theta,z;\theta',z';k)=\frac{1}{4\pi}\int_0^{2\pi}\int_{-\pi/2}^{\pi/2}\sum_{q=0}^{\infty}\sum_{q'=0}^{\infty}\epsilon_q\epsilon_{q'}(-i)^q(i)^{q'}\times J_q(ka\cos\phi_p)J_{q'}(ka\cos\phi_p)\times\cos(q(\theta-\theta_p))\cos(q'(\theta'-\theta_p))\times e^{ik(z'-z)\sin\phi_p}\cos\phi_p\,d\phi_p\,d\theta_p. \tag{25}$$

The joint acceptance function [Eq. (20)] related to modes (n,s) and (n,t) is therefore given by

$$J^{\mathrm{cyl}}_{(n,s),(n,t)}(k)\equiv\mathcal{I}_C=a^2\int_0^L\int_0^L\int_0^{2\pi}\int_0^{2\pi} f_c^{\mathrm{cyl}}(\theta,z;\theta',z';k)\times\cos(n\theta)\cos(n\theta')\times\sin\left(\frac{s\pi z}{L}\right)\sin\left(\frac{t\pi z'}{L}\right)d\theta\,d\theta'\,dz\,dz'. \tag{26}$$

Substitution of Eq. (25) into Eq. (26) and further use of orthogonality properties of trigonometric functions give

$$\mathcal{I}_C=\frac{2a^2\pi^2}{\epsilon_n}\int_{-\pi/2}^{\pi/2}J_n^2(ka\cos\phi_p)\times\cos\phi_p\left(\int_0^L e^{-ikz\sin\phi_p}\sin\left(\frac{s\pi z}{L}\right)dz\right)\times\left(\int_0^L e^{ikz'\sin\phi_p}\sin\left(\frac{t\pi z'}{L}\right)dz'\right)d\phi_p \tag{27}$$

or

$$\mathcal{I}_C=\frac{2a^2\pi^2}{\epsilon_n}\int_{-\pi/2}^{\pi/2}J_n^2(ka\cos\phi_p)\mathcal{H}_s^{(-)}(\phi_p)\times\mathcal{H}_t^{(+)}(\phi_p)\cos\phi_p d\phi_p \tag{28}$$

with

$$\mathcal{H}_s^{(-)}(\phi)=\int_0^L e^{-ikz\sin\phi}\sin\left(\frac{s\pi z}{L}\right)dz, \tag{29}$$

and

$$\mathcal{H}_t^{(+)}(\phi)=\int_0^L e^{ikz\sin\phi}\sin\left(\frac{t\pi z}{L}\right)dz. \tag{30}$$

It will be shown in the next section how to evaluate analytically these last two integrals. Consequently the original quadruple integral has been reduced to the simple integral given by Eq. (28).

C. Truncated conical case

The spatial correlation function for a truncated conical geometry can be reformulated equivalently using a decomposition of plane waves into spherical harmonics. These coordinates are effectively more appropriate for handling the double surface integral along a truncated conical surface.

The incident field p_{θ_p,ϕ_p} related to a plane wave of unit amplitude with incidence (θ_p,ϕ_p) can be decomposed into spherical harmonics:[7,8]

$$p_{\theta_p,\phi_p}(r,\theta,\phi)=4\pi\sum_{q=0}^{\infty}\sum_{l=-q}^{q}(i)^q j_q(kr)\times Y^*_{ql}(\theta,\phi)Y_{ql}(\theta_p,\phi_p), \tag{31}$$

where j_q is the spherical Bessel function of order q and Y_{ql} is the spherical harmonics given by

$$Y_{ql}(\theta,\phi)=C_{ql}P_{q|l|}(\cos\phi)e^{il\theta}. \tag{32}$$

In this expression, the coefficient C_{ql} is given by

$$C_{ql}=\sqrt{\frac{(2q+1)(q-|l|)!}{4\pi(q+|l|)!}} \tag{33}$$

while P_{ql} is the Legendre function.

It should be emphasized that points along the truncated conical surface are characterized by $\phi=\alpha$.

The spatial correlation function for the diffuse field can again be obtained by summing up the effects of plane waves originating from all spatial directions:

$$f_c^{tc}(r,\theta;r',\theta';k)=\frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi} p_{\theta_p,\phi_p}(r,\theta,\alpha) \times p^*_{\theta_p,\phi_p}(r',\theta',\alpha)\sin\phi_p d\phi_p d\theta_p. \tag{34}$$

Substitution of Eq. (31) (with $\phi=\alpha$) into Eq. (34) leads after use of orthogonality properties of spherical harmonics to the following result:

$$f_c^{tc}(r,\theta;r',\theta';k)=4\pi\sum_{q=0}^{\infty}\sum_{l=-q}^{q} i^q(i^q)^* j_q(kr) \times j_q^*(kr')Y_{ql}^*(\theta,\alpha)Y_{ql}(\theta',\alpha). \tag{35}$$

The joint acceptance function related to modes (n,s) and (n,t) is therefore given by

$$J^{tc}_{(n,s),(n,t)}(k)\equiv\mathcal{I}_T=4\pi\sum_{q=0}^{\infty}\sum_{l=-q}^{q}\int_{r_1}^{r_2}\int_{r_1}^{r_2}\int_0^{2\pi}\int_0^{2\pi} i^q(i^q)^* \times j_q(kr)j_q^*(kr')Y_{ql}^*(\theta,\alpha) \times Y_{ql}(\theta',\alpha)\cos(n\theta)\cos(n\theta') r^3 \times\sin\left(\frac{s\pi(r-r_1)}{r_2-r_1}\right)r'^3\sin\left(\frac{t\pi(r'-r_1)}{r_2-r_1}\right) \times\sin^2\alpha\, d\theta\, d\theta'\, dr\, dr'. \tag{36}$$

Substitution of Eq. (32) into Eq. (36) leads to the following result:

$$\mathcal{I}_T=4\pi\sin^2\alpha\sum_{q=0}^{\infty}\sum_{l=-q}^{q} i^q(i^q)^* C_{ql}^2 P_{q|l|}^2(\cos\alpha) \times\int_{r_1}^{r_2} j_q(kr)\sin\left(\frac{s\pi(r-r_1)}{r_2-r_1}\right)r^3 dr \times\int_{r_1}^{r_2} j_q^*(kr')\sin\left(\frac{t\pi(r'-r_1)}{r_2-r_1}\right)r'^3 dr' \times\int_0^{2\pi} e^{-il\theta}\cos(n\theta)d\theta\int_0^{2\pi} e^{-il\theta'}\cos(n\theta')d\theta'. \tag{37}$$

Integrals involving θ and θ' can be evaluated as follows:

$$\int_0^{2\pi} e^{-il\theta}\cos(n\theta)d\theta=\frac{2\pi\delta_{|l|n}}{\epsilon_{|l|}}, \tag{38}$$

where the Neumann factor $\epsilon_l=1$ for $l=0$ and $\epsilon_l=2$ for $l\neq 0$.

Substitution of Eq. (38) into Eq. (37) leads [after use of $i^q(i^q)^*=1$] to the final result:

$$\mathcal{I}_T=\frac{16\pi^3\sin^2\alpha}{\epsilon_n^2}\sum_{q=n}^{+\infty} C_{qn}^2 P_{qn}^2(\cos\alpha)\mathcal{G}(q,s)\mathcal{G}^*(q,t) \tag{39}$$

with

$$\mathcal{G}(q,s)=\int_{r_1}^{r_2} j_q(kr)\sin\left(\frac{s\pi(r-r_1)}{r_2-r_1}\right)r^3 dr. \tag{40}$$

The quadruple integral is thus reduced to an equivalent form [Eq. (39)] involving the product of two simple integrals.

VI. EVALUATION OF RESULTING INTEGRALS

A. Cylindrical case

The expression of the joint acceptance function given by Eq. (28) already has attractive features versus the original expression given by Eq. (20). We show now that the product of integrals $\mathcal{H}_s^{(-)}(\phi_p)\mathcal{H}_t^{(+)}(\phi_p)$ can be computed analytically. The evaluation of $\mathcal{I}_C$ is therefore reduced to the evaluation of a simple integral. In order to clarify the paper, proofs are given in the Appendix.

Lemma 1: If we define $\beta=k\sin(\phi_p)$ *and* $\gamma=s\pi/L$, *one can show that the following results hold*:

$$\mathcal{H}_s^{(+)}(\phi_p)=\begin{cases}\gamma\dfrac{(-1)^s e^{i\beta L}-1}{\beta^2-\gamma^2} & if\ \ \beta^2\neq\gamma^2\\ \dfrac{iL\beta}{2\gamma} & if\ \ \beta^2=\gamma^2,\end{cases} \tag{41}$$

and

$$\mathcal{H}_s^{(-)}(\phi_p)=\begin{cases}\gamma\dfrac{(-1)^s e^{-i\beta L}-1}{\beta^2-\gamma^2} & if\ \ \beta^2\neq\gamma^2\\ -\dfrac{iL\beta}{2\gamma} & if\ \ \beta^2=\gamma^2.\end{cases} \tag{42}$$

From Lemma 1, $\mathcal{I}_C$ is reduced to a simple integral. However the following proposition gives simpler formulas to compute the product $\mathcal{H}_s^{(-)}(\phi_p)\mathcal{H}_t^{(+)}(\phi_p)$.

Proposition 1: We define $\gamma_s=s\pi/L$ *and* $\gamma_t=t\pi/L$. *The function* $\mathcal{H}_{st}$ *defined by*

$$\mathcal{H}_{st}:\left[-\frac{\pi}{2},\frac{\pi}{2}\right]\to\mathbb{C},\qquad \phi_p\to\mathcal{H}_s^{(-)}(\phi_p)\mathcal{H}_t^{(+)}(\phi_p),$$

verifies:

(1) *If* $\beta^2\neq\gamma_s^2$ *and* $\beta^2\neq\gamma_t^2$,

$$\mathcal{H}_{st}(\phi_p)=\Delta_{st}(\phi_p)\times\begin{cases}(1-(-1)^t\cos(\beta L)) & if\ \ s+t\ \ is\ \ even\\ (-1)^s i\sin(\beta L) & if\ \ s+t\ \ is\ \ odd.\end{cases} \tag{43}$$

where $\Delta_{st}(\phi_p)=2\gamma_s\gamma_t/[(\beta^2-\gamma_s^2)(\beta^2-\gamma_t^2)]$

(2) *If* $\beta^2=\gamma_s^2$ *or* $\beta^2=\gamma_t^2$,

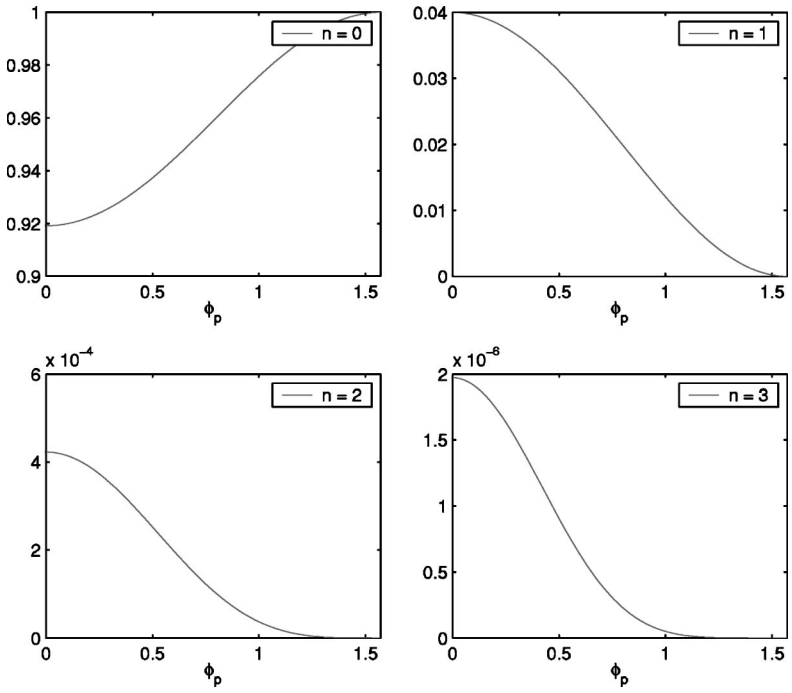


FIG. 3. Function $J_n^2(ka\cos\phi_p)$ with $ka=0.408\,41$ for $\phi_p\in[0,\pi/2]$ and $n=0$, 1, 2, 3.

$$\mathcal{H}_{st}(\phi_p)=\begin{cases}\dfrac{L^2}{4} & if\ \ s=t\\ 0 & if\ \ s+t\ \ is\ \ even\ \ and\ \ s\neq t\\ \dfrac{iL}{\beta}\dfrac{ts}{s^2-t^2} & if\ \ s+t\ \ is\ \ odd.\end{cases} \tag{44}$$

Equation (44) is useless for an exact evaluation of $\mathcal{I}_C$ since it concerns a discrete number of values. However a numerical integration method can encounter problems for the evaluation of Eq. (43) when β is close to $|\gamma_s|$ or $|\gamma_t|$. More details are given in Sec. VII.

If one introduces the above results into Eq. (28), one obtains the final result:

$$\text{If } s+t \text{ is odd } \mathcal{I}_C=0, \tag{45}$$

$$\text{If } s+t \text{ is even } \mathcal{I}_C=\frac{4\pi^2a^2}{\epsilon_n}\times\int_0^{\pi/2}J_n^2(ka\cos(\phi_p))\times\mathcal{H}_{st}(\phi_p)\cos(\phi_p)d\phi_p, \tag{46}$$

where $\mathcal{H}_{st}$ is evaluated using Eq. (43) or Eq. (44).

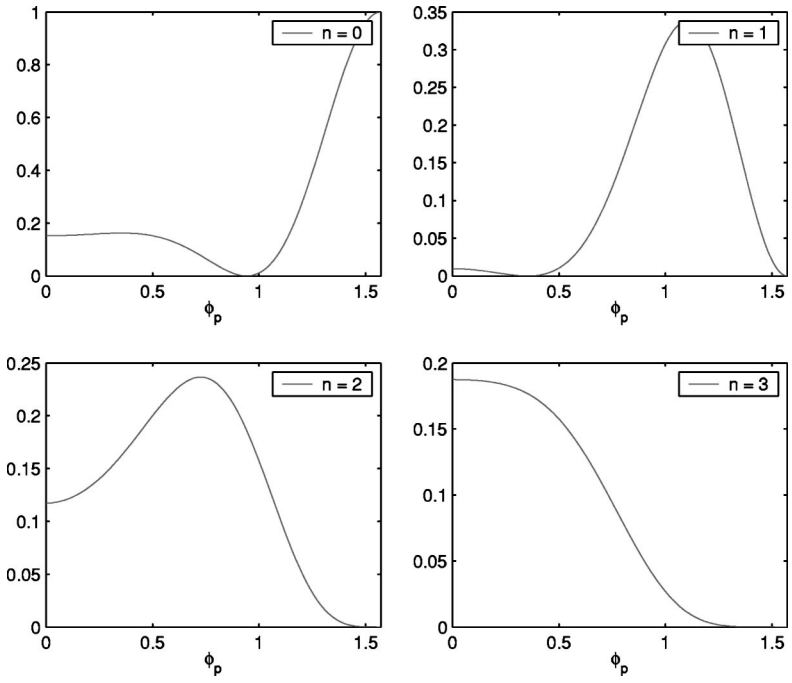


FIG. 4. Function $J_n^2(ka\cos\phi_p)$ with $ka=4.0841$ for $\phi_p\in[0,\pi/2]$ and $n=0$, 1, 2, 3.

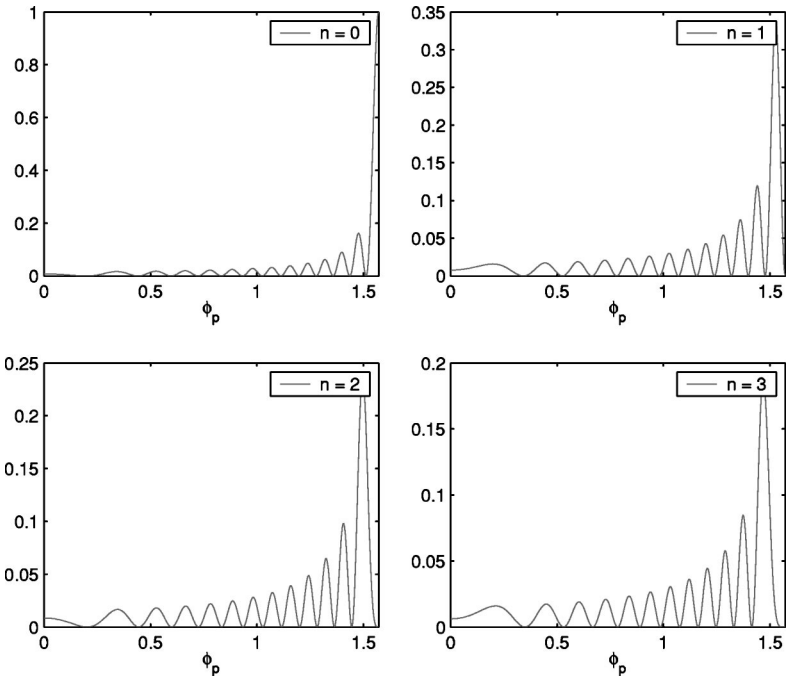


FIG. 5. Function $J_n^2(ka\cos\phi_p)$ with $ka=40.8408$ for $\phi_p\in[0,\pi/2]$ and $n=0$, 1, 2, 3.

Additionally, one can deduce the following result from Proposition 1:

$$J_{(n,s),(n,t)}(k)=J_{(n,t),(n,s)}(k). \tag{47}$$

B. Truncated conical case

The integrals $\mathcal{G}(q,s)$ can be computed using recurrence formulas between spherical Bessel functions of different orders. Only the first term of the series must be computed numerically. This evaluation requires the computation of a simple integral.

Theoretically this way looks attractive but its numerical application fails because recurrence relations are numerically ill-posed. Rounding-errors grow quickly so that final results are inaccurate.

Another potential problem is the convergence of the infinite series in Eq. (39). The fact that this series refers to Bessel and Legendre functions does not ease its mathematical study. However in all numerical experiments performed, the modulus of the principal term of the series decreases very quickly.

VII. NUMERICAL EXPERIMENTS

A. Cylindrical case

1. Preliminary remarks

Equation (45) and Eq. (46) support the evaluation of $\mathcal{I}_C$. The successful evaluation of this integral requires the preliminary observation of the various terms involved and the identification of potential numerical problems. In this way, an appropriate integration strategy can be selected.

The simplest function to integrate in $\mathcal{I}_C$ is clearly $\cos(\phi_p)$. Note that the cosine function is decreasing on $[0,\pi/2]$ and also appears in the argument of the Bessel functions.

These Bessel functions can be computed using AMOS FORTRAN library. The behavior of Bessel functions is obviously depending on the value of ka (see Figs. 3–5). *A priori*, no specific rule can be applied in order to locate optimally integration points.

The evaluation of function $\mathcal{H}_{st}$ requires some caution. Its expression given in Proposition 1 can be divided in two parts, the first one is given by $\Delta_{st}(\phi_p)$ while the second one contains a trigonometric function.

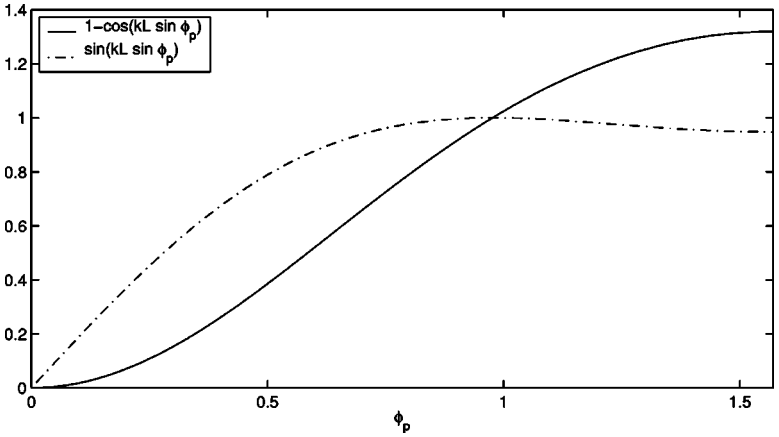


FIG. 6. Functions $1-\cos(kL\sin\phi_p)$ and $\sin(kL\sin\phi_p)$ for $kL=1.8960$.

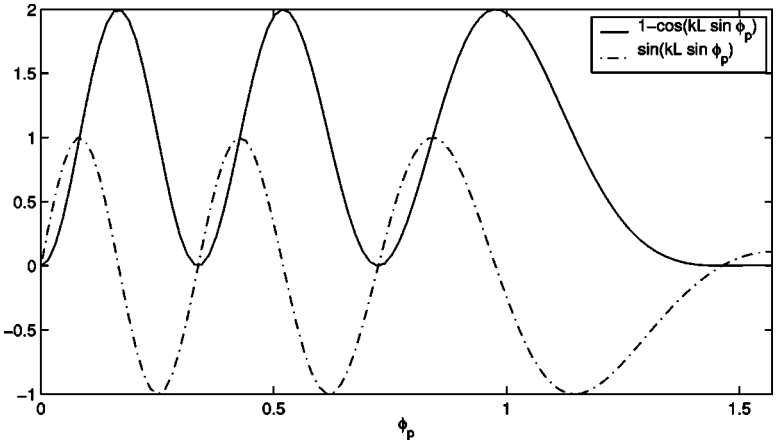


FIG. 7. Functions $1-\cos(kL \sin \phi_p)$ and $\sin(kL \sin \phi_p)$ for $kL = 18.9605$.

The computation of the second part is quite easy. Roughly speaking, it can be considered, for numerical integration, as a periodic function (as shown in Figs. 6 and 7).

Numerical integration of $\Delta_{st}(\phi_p)$ requires more care. The denominator of $\Delta_{st}(\phi_p)$ is close to zero when β is close to $|\gamma_s|$ or $|\gamma_t|$ but $\mathcal{H}_{st}$ is clearly a continuous function and therefore theoretically $\mathcal{H}_{st}(\phi_p)$ is well defined for any ϕ_p value.

However numerical evaluation of $\Delta_{st}(\phi_p)$ can contain large numerical errors around $|\gamma_s|$ or $|\gamma_t|$. A simple solution is to take, when β is close $|\gamma_s|$ or $|\gamma_t|$, the value of $\mathcal{H}_{st}$ given by Eq. (44) which is valid when β is exactly equal to $|\gamma_s|$ or $|\gamma_t|$.

This strategy is illustrated in Fig. 8. It introduces only a small perturbation in the evaluation of $\mathcal{I}_C$.

2. Numerical results

Numerical results have been produced for the following cylindrical shell configuration:[9] $a = 0.221$ m, $L = 1.026$ m. The surrounding fluid is the air with a speed of sound $c = 340$ m/s. The joint acceptance function is evaluated at a frequency of 100 Hz. The related wave number value is $k = 1.848$.

The numerical integration is performed using the trapezoidal method with m points. The evaluation of the quadruple integral therefore requires m^4 evaluations of functions while only m evaluations are requested for the simple integral.

Table I gives estimations of $\mathcal{I}_C$ obtained for different values of n, s, and t. Table II shows the differences between evaluations of Table I and the estimation of $\mathcal{I}_C$ obtained with $m = 400$. These differences have been produced only for the two cases leading to a significant value of the resulting integral.

Examination of these results shows that the simple integral converges quite quickly. The quadruple integral is not easy to evaluate accurately. Computational times (processor AMD Athlon XP1800+) required for simple and quadruple integrals are summarized in Table III.

B. Truncated conical case

1. Preliminary remarks

The numerical evaluation of $\mathcal{G}(q,s)$ and $\mathcal{G}^*(q,t)$ is quite easy since it requires the evaluation of well-known spherical Bessel functions. The integration method must be chosen with respect to kr and the order q. The other functions are

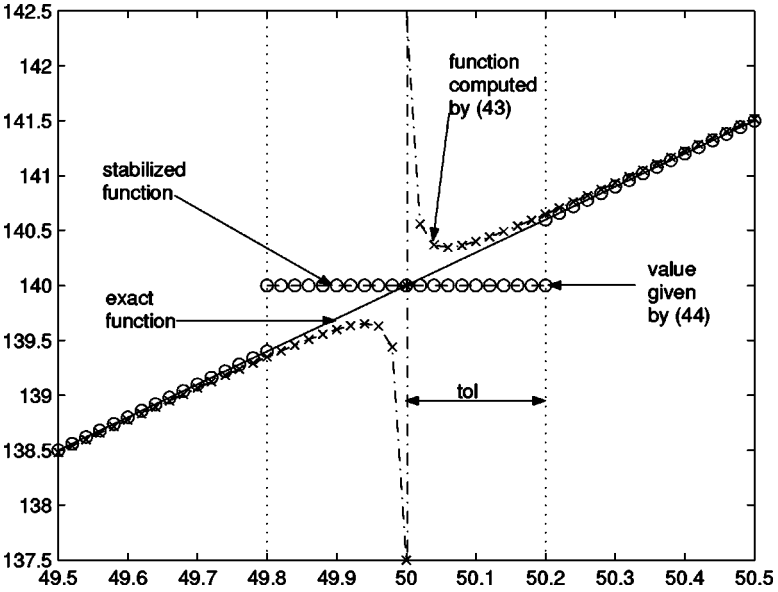


FIG. 8. Behavior of function $\mathcal{H}_{st}$ in the vicinity of the singularity.

TABLE I. Estimations of $\mathcal{I}_C$ based on quadruple and simple integrals for different values of n, s, and t using the trapezoidal method with m points.

$n=0$, $s=1$, $t=1$	Quadruple integral	Simple integral
$m=5$	7.5655e−001	7.3750e−001
$m=10$	7.3994e−001	7.3506e−001
$m=20$	7.3603e−001	7.3506e−001
$m=50$	7.3506e−001	7.3506e−001
$n=7$, $s=1$, $t=1$	Quadruple integral	Simple integral
$m=5$	9.4165e−005	1.1077e−018
$m=10$	3.6772e−007	1.1077e−018
$m=20$	1.6616e−018	1.1077e−018
$m=50$	9.6119e−019	1.1077e−018
$n=0$, $s=1$, $t=7$	Quadruple integral	Simple integral
$m=5$	2.5803e−001	9.2993e−002
$m=10$	1.1483e−001	9.2700e−002
$m=20$	9.7633e−002	9.2700e−002
$m=50$	9.3482e−002	9.2700e−002
$n=0$, $s=1$, $t=2$	Quadruple integral	Simple integral
$m=5$	−5.6753e−017	0
$m=10$	2.7815e−018	0
$m=20$	−5.2845e−018	0
$m=50$	2.7263e−018	0

smooth but one should consider the fact that Bessel functions can have small values when r^3 is large. So when r_1 is very small, then the r^3 term can have a large contribution to $\mathcal{G}(q,s)$ and $\mathcal{G}^*(q,t)$.

The main problem is to know when the series can be truncated. We use a very simple criterion. We decide to truncate the series when

$$C_{pn}P_{pn}(\cos(\alpha))\mathcal{G}(p,s)\mathcal{G}^*(p,t) \leqslant \mathrm{tol}\sum_{q=n}^{n_u} C_{qn}P_{qn}(\cos(\alpha))\mathcal{G}(q,s)\mathcal{G}^*(q,t),$$

where tol is a threshold parameter, generally 10^{-12}. Despite this very small value, only a few terms are computed. In our numerical tests, the series always converged quickly.

2. Numerical experiments

Numerical results have been produced for the following truncated conical shell configuration:[10] $r_1=0.689\,\mathrm{m}$, $r_2=1.705\,\mathrm{m}$, and $\alpha=16°$. The surrounding fluid is the air with a speed of sound $c=340\,\mathrm{m/s}$. The joint acceptance function is evaluated at a frequency of 100 Hz. The related wave number value is $k=1.848$.

TABLE II. Absolute differences of various estimations of $\mathcal{I}_C$ with respect to an estimation computed with $m=400$.

$n=0$, $s=1$, $t=1$	Quadruple integrals	Simple integral
$m=5$	1.4486e−003	1.8950e−004
$m=10$	3.5605e−004	4.7303e−005
$m=20$	8.8598e−005	1.1800e−005
$m=50$	1.4135e−005	1.8633e−006
$n=0$, $s=1$, $t=7$	Quadruple integral	Simple integral
$m=5$	8.7084e−002	1.3602e−004
$m=10$	1.1644e−002	3.4218e−005
$m=20$	2.5925e−003	8.5521e−006
$m=50$	4.0230e−004	1.3509e−006

TABLE III. CPU times for the evaluation of $\mathcal{I}_C$ (simple and quadruple integrals) using the trapezoidal method with m points.

Integral	$m=5$	$m=10$	$m=20$	$m=50$
Simple	4.3 ms	5.8 ms	10.8 ms	25.3 ms
Quadruple	880 ms	12.1 s	3 mn 1.07 s	1 h 54 mn 6.63 s

The numerical integration is again performed using the trapezoidal method with m points. The evaluation of the quadruple integral therefore requires m^4 evaluations of functions while only m evaluations are requested for the simple integral.

Table IV gives estimations of $\mathcal{I}_T$ obtained for different values of n, s and t. The number of terms N_q retained in the truncated series ($N_q=n_u-n+1$) is also indicated. Table V shows the differences between evaluations of Table IV and the estimation of $\mathcal{I}_T$ obtained with $m=400$.

Examination of these results shows that the simple integral converges quite quickly. The quadruple integral is not easy to evaluate accurately. Computational times (processor AMD Athlon XP1800+) required for simple and quadruple integrals are summarized in Table VI.

VIII. CONCLUSIONS

The evaluation of the joint acceptance functions related to a diffused field excitation along cylindrical and truncated conical shells has been addressed starting from the derivation of alternative forms of the spatial correlation function. In this way, the quadruple integral can be reduced to a combination of simple integrals. The proposed reduction scheme has been

TABLE IV. Estimations of $\mathcal{I}_T$ based on quadruple and simple integrals for different values of n, s, and t using the trapezoidal method with m points.

$n=0$, $s=1$, $t=1$	Quadruple integral	Simple integral	N_q
$m=5$	0.3830e+01	0.3830e+01	11
$m=10$	0.3714e+01	0.3714e+01	11
$m=20$	0.3686e+01	0.3686e+01	11
$m=50$	0.3678e+01	0.3678e+01	11
$n=7$, $s=1$, $t=1$	Quadruple integral	Simple integral	N_q
$m=5$	0.3869e−02	0.1524e−13	10
$m=10$	0.4316e−04	0.1361e−13	10
$m=20$	0.1318e−13	0.1318e−13	10
$m=50$	0.1306e−13	0.1306e−13	10
$n=0$, $s=1$, $t=7$	Quadruple integral	Simple integral	N_q
$m=5$	0.1810e+01	0.1810e+01	11
$m=10$	0.8082e+00	0.8082e+00	11
$m=20$	0.6861e+00	0.6861e+00	11
$m=50$	0.6566e+00	0.6566e+00	11
$n=1$, $s=1$, $t=2$	Quadruple integral	Simple integral	N_q
$m=5$	−0.2318e+01	−0.2318e+01	10
$m=10$	−0.2171e+01	−0.2171e+01	10
$m=20$	−0.2136e+01	−0.2136e+01	10
$m=50$	−0.2127e+01	−0.2127e+01	10

TABLE V. Absolute differences(*) or relative differences(○) of various estimations of $\mathcal{I}_T$ with respect to an estimation computed with $m=400$.

$n=0,\ s=1,\ t=1$	Quadruple integrals(○)	Simple integral(○)
$m=5$	0.4174e−01	0.4174e−01
$m=10$	0.1025e−01	0.1025e−01
$m=20$	0.2546e−02	0.2546e−02
$m=50$	0.4016e−03	0.4016e−03
$n=7,\ s=1,\ t=1$	Quadruple integral(*)	Simple integral(*)
$m=5$	0.3869e−02	0.2198e−14
$m=10$	0.4316e−04	0.5765e−15
$m=20$	0.1456e−15	0.1453e−15
$m=50$	0.2364e−16	0.2301e−16
$n=0,\ s=1,\ t=7$	Quadruple integral(○)	Simple integral(○)
$m=5$	0.8572e+01	0.1779e+01
$m=10$	0.2411e+00	0.2411e+00
$m=20$	0.5360e−01	0.5360e−01
$m=50$	0.8209e−02	0.8209e−02
$n=1,\ s=1,\ t=2$	Quadruple integral(*)	Simple integral(*)
$m=5$	0.6889e−02	0.6889e−02
$m=10$	0.1644e−02	0.1644e−02
$m=20$	0.4055e−03	0.4055e−03
$m=50$	0.6382e−04	0.6382e−04

validated by numerical experiments. A substantial reduction of the computational cost can be observed. The developed procedure can be applied to other classes of axisymmetric structures where the mode shapes can be factorized as a product of a function of the axial coordinate z and a trigonometric function of the circumferential coordinate θ. Such generalizations are trivial and should lead again to significant computational savings.

APPENDIX: PROOFS OF SEC. VI

Proof 1 (Lemma 1): We can formulate $\mathcal{H}_s^{(+)}(\phi_p)$ *as*

$$\mathcal{H}_s^{(+)}(\phi_p)=\int_0^L e^{i\beta y}\sin(\gamma y)dy=\frac{1}{2i}\int_0^L e^{i(\beta+\gamma)y}dy - \frac{1}{2i}\int_0^L e^{i(\beta-\gamma)y}dy$$

where

$$\int_0^L e^{\alpha y}dy=\begin{cases}\dfrac{1}{\alpha}(e^{\alpha L}-1) & if\ \ \alpha\neq 0\\ L & otherwise,\end{cases}$$

therefore:

TABLE VI. CPU times for the evaluation of $\mathcal{I}_T$ (quadruple and simple integrals) using the trapezoidal method with m points.

Integral	$m=5$	$m=10$	$m=20$	$m=50$
Simple	0.24 s	0.36 s	0.67 s	1.6 s
Quadruple	0.94 s	12.52 s	3 mn 9.43 s	1 h 58 mn 9.93 s

$$if\ \ \beta^2\neq\gamma^2,\quad \mathcal{H}_s^{(+)}(\phi_p)=\frac{e^{i(\beta-\gamma)L}-1}{2(\beta-\gamma)}-\frac{e^{i(\beta+\gamma)L}-1}{2(\beta+\gamma)}=\gamma\frac{(-1)^s e^{i\beta L}-1}{\beta^2-\gamma^2},$$

$$if\ \ \beta=\gamma,\quad \mathcal{H}_s^{(+)}(\phi_p)=\frac{1}{2i}\left(\frac{e^{2i\gamma L}-1}{2i\gamma}-L\right)=\frac{iL}{2},$$

$$if\ \ \beta=-\gamma,\quad \mathcal{H}_s^{(+)}(\phi_p)=\frac{1}{2i}\left(L+\frac{e^{-2i\gamma L}-1}{2i\gamma}\right)=-\frac{iL}{2},$$

using $\gamma L=s\pi$.

Proof 2 (Proposition 1): First, assume that $\beta^2\neq\gamma_s^2$ *and* $\beta^2\neq\gamma_t^2$.

From Lemma 1, we have

$$\mathcal{H}_{st}(\phi_p)=\frac{\gamma_s\gamma_t}{(\beta^2-\gamma_s^2)(\beta^2-\gamma_t^2)}(1+(-1)^{s+t}-(-1)^t e^{i\beta L}-(-1)^s e^{-i\beta L}).$$

If $s+t$ *is even,* $(-1)^t=(-1)^s$ *so*

$$\mathcal{H}_{st}(\phi_p)=\frac{2\gamma_s\gamma_t}{(\beta^2-\gamma_s^2)(\beta^2-\gamma_t^2)}(1-(-1)^t\cos(\beta L)).$$

If $s+t$ *is odd,* $(-1)^t=-(-1)^s$ *and*

$$\mathcal{H}_{st}(\phi_p)=(-1)^s\frac{2i\gamma_s\gamma_t}{(\beta^2-\gamma_s^2)(\beta^2-\gamma_t^2)}\sin(\beta L).$$

Now, assume that $\gamma_s=\xi\beta$ *with* $\xi=\pm 1$ *and* $\beta^2\neq\gamma_t^2$.

We have $\gamma_t=\alpha\xi\beta$ *with* $\alpha=t/s\neq 1$ *and* $\beta L=\xi s\pi$. *Lemma 1 implies that*

$$\mathcal{H}_{st}(\phi_p)=-\frac{iL}{2\beta}\frac{\alpha}{1-\alpha^2}((-1)^t e^{\xi is\pi}-1)$$

$$=-\frac{iL}{2\beta}\frac{\alpha}{1-\alpha^2}((-1)^{t+s}-1).$$

Then we have

$$\mathcal{H}_{st}(\phi_p)=\begin{cases}\dfrac{iL}{\beta}\dfrac{ts}{s^2-t^2} & if\ \ s+t\ \ is\ \ odd\\ 0 & if\ \ s+t\ \ is\ \ even.\end{cases}$$

Assume that $\gamma_t=\xi\beta$ *with* $\xi=\pm 1$ *and* $\beta^2\neq\gamma_s^2$.

As in the previous case, we can show that

$$\mathcal{H}_{st}(\phi_p)=\begin{cases}\dfrac{iL}{\beta}\dfrac{ts}{s^2-t^2} & if\ \ s+t\ \ is\ \ odd\\ 0 & if\ \ s+t\ \ is\ \ even.\end{cases}$$

To finish, assume that $\gamma_s=\xi_s\beta$ *and* $\gamma_t=\xi_t\beta$ *with* $\xi_s=\pm 1$ *and* $\xi_t=\pm 1$.

Clearly, $s=t$ *and* $\xi_s=\xi_t\equiv\xi$. *So we have*

$$\mathcal{H}_{st}(\phi_p)=\left(-\xi\frac{iL}{2}\right)\left(\xi\frac{iL}{2}\right)=\frac{L^2}{4}.$$

Proof 3 [Proof of Eqs. (45) and (46)]: From Proposition 1, we deduce that

$\mathcal{H}_{st}$ *is even if* $s+t$ *is even,* $\mathcal{H}_{st}$ *is odd if* $s+t$ *is odd, because* β *is an odd function in* ϕ_p.

The functions $J_n^2(ka\cos(\phi_p))$ *and* $\cos(\phi_p)$ *are even in* ϕ_p *and the integration domain is symmetric hence we use basic properties of odd and even functions to close the proof.*

[1] A. Preumont, *Random Vibrations and Spectral Analysis* (Kluwer Academic, Dordrecht, 1994).
[2] M. P. Norton, *Fundamentals of Noise and Vibration Analysis for Engineers* (Cambridge University Press, New York, 1991).
[3] C. M. Willis, W. H. Mayes, L. D. Pope, and D. C. Rennison, ''Development and validation of preliminary analytical models for aircraft interior noise prediction,'' J. Sound Vib. **82**, 541–575 (1982).
[4] F. Jacobsen and T. Roisin, ''The coherence of reverberant sound fields,'' J. Acoust. Soc. Am. **108**, 204–210 (2000).
[5] F. Jacobsen, ''The diffuse sound field,'' Technical report, The Acoustic Laboratory, Technical University of Denmark, Report 27, 1979.
[6] C. Mecitoglu, ''Vibration characteristics of a stiffened conical shell,'' J. Sound Vib. **197**, 191–206 (1996).
[7] E. Skudrzyk, *The Foundations of Acoustics* (Springer, Wien, 1971).
[8] F. P. Mechel, *Formula of Acoustics* (Springer, Berlin, 2002).
[9] C. Lesueur, *Rayonnement Acoustique des Structures—Vibro-acoustique, Interactions Fluide-Structure* (Acoustic radiation from structures—Vibro-acoustics, fluid-structure interaction) (Eyrolles, Paris, 1988).
[10] C. Lesueur, P. Neple, and A. El Hafidi, ''On sound transmission through a truncated isotropic conical shell: Experimental and preliminary numerical results,'' in Internoise 2001, The Hagues, pp. 907–910.

Radiative transfer theory for high-frequency power flows in fluid-saturated, poro-visco-elastic media

Eric Savin
Structural Dynamics and Coupled Systems Department, ONERA 29, avenue de la Division Leclerc, 92322 Châtillon cedex, France

(Received 28 July 2004; revised 1 December 2004; accepted 15 December 2004).

Some recent mathematical results are used to study the propagation features of the high-frequency vibrational energy density in three-dimensional fluid-saturated, isotropic poro-visco-elastic media. The theory of high-frequency asymptotics of the solutions of hyperbolic partial differential equations shows that their energy satisfies Liouville-type transport or radiative transfer equations for randomly heterogeneous materials. For long propagation times these equations can be approached by diffusion equations. The corresponding diffusion parameters—mean-free paths and diffusion constants—associated with Biot's linear model for such media, are derived. The analysis accounts for the thermal and viscous memory effects of the fluid phase, and the viscous memory effect of the solid phase through time convolution operators. In this respect it also extends the existing mathematical results. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1856271]

PACS numbers: 43.20.Gp, 43.20.Bi, 43.40.At [RLW] Pages: 1020–1031

I. INTRODUCTION

The theory of radiative transfer describes the high-frequency energy evolution in a linear random medium. It originates from considerations on the propagation of light through turbulent atmospheres.[1] In the frame of the classical multiple scattering theory of waves,[2] a radiative transfer equation for the intensity can be obtained from a Bethe–Salpeter (BS) equation under certain hypotheses, in particular when the wavelength is small compared to the macroscopic features of the medium, and perturbations of the latter are weak. In addition, the so-called ladder approximation is often invoked for the BS equation to make it suitable for analytical developments. However, it is only recently that a systematic derivation of transport equations for the energy in random electromagnetic, acoustic, or elastic media has been proposed.[3–6] The analysis is based (i) on the use of a Wigner transform, the high-frequency, non-negative limit of which is the angularly resolved energy density in time and space; and (ii) on the introduction of an explicit scaling of wave propagation patterns. The mathematical framework, called microlocal analysis, is very general and applies to any kind of first-order hyperbolic partial differential equations. The purpose of this paper is to present how the theory applies to vibrations of fluid-saturated porous media as described by Biot's equations. The main objective is to derive the acoustic diffusion parameters for such materials, because they could be linked in some ways to their acoustic insulation and dissipation properties. Thermal dissipation and viscosity of the fluid as well as viscosity of the solid skeleton are taken into account in the analysis. This requires nontrivial extensions of the existing theory to the case of time convolution operators, as done in Ref. 7 for visco-elastic media. These results are adapted to poro-visco-elastic media in the present paper.

The scattering of a fast compression wave by a spherical inclusion in a fluid-saturated porous medium was considered in Ref. 8. These results are limited mostly to the lower frequency range and the case of a small Biot's modulus. In Ref. 9 this problem was re-examined for higher frequencies in the frame of a Born approximation. Waves scattered by an obstacle embedded in a porous medium were calculated by representational integral equations in Ref. 10, using *ad hoc* Green's kernels derived in the course of the paper. In Ref. 11 the author analyzed multiply scattered waves by perturbative expansions, but his work is limited to delta-correlated scatterers and by approximations inherent to pertubation techniques. None of these studies has considered the high-frequency transport and diffusion limits for arbitrarily random media as done here. Furthermore, this analysis is not reduced to a particular shape or distribution of scatterers, such as spherical or delta-correlated heterogeneities. The sole restriction is that random perturbations of the medium parameters shall be weak and isotropic, as effectively defined by the theory presented below. Also, one considers that they vary—either smoothly or not—at a length scale comparable to the wavelength, which is the *a priori* worst configuration since one can expect strong interactions between waves and heterogeneities of similar sizes.

In this paper, we first recall the basic mathematical tools we have used to describe the high-frequency energy density and its evolution with time for rapidly oscillating initial conditions. Then, this setting—extended to time convolution operators—is applied to Biot's equations describing the motion of a fluid-saturated, isotropic poro-visco-elastic medium. The effects of thermal dissipation at the interface between the fluid and solid phases and viscosity of the fluid are also considered. One concurrently obtains the exact analytical expressions of the compressional wave speeds in this regime. The influence of random perturbations of the mechanical parameters of the solid skeleton is analyzed in Sec. IV and the diffusion regime is depicted in Sec. V, where analytical expressions of the diffusion parameters, the transport mean free paths, and the diffusion constant are given. A numerical example is examined in Sec. VI using experimental data taken from the literature. Some conclusions are drawn in Sec. VII.

One notes that the same analysis of high-frequency energy propagation features has been proposed in Ref. 12, but calculations were not performed thoroughly and memory effects (viscosity and thermal coupling) were not taken into account.

II. THEORETICAL BACKGROUND: WIGNER MEASURES AS ENERGY DENSITY ESTIMATES

We summarize in this section the basic developments of Refs. 3–6 generalized to incorporate the time convolution operators in Ref. 7. The work reported in Ref. 7 deals with the extension of the theory to the case of a visco-elastic material with memory effects. The work reported here specializes it to the case of a poro-elastic medium that considers the thermal and viscous effects of the fluid phase. The interested reader may consult these references for a detailed presentation and the proofs associated with the results given below. The materials presented in Sec. II A and Sec. II B are basically adapted from Ref. 13.

Consider a first-order hyperbolic system of the form

$$\mathbf{A}(\mathbf{x},0)\partial_t\mathbf{v}+\sum_{j=1}^{d}\mathbf{D}_j\partial_{x_j}(\mathbf{v}+\mathbf{B}(\mathbf{x},\cdot)*_t\mathbf{v})+\mathbf{\Omega}(\mathbf{x},\cdot)*_t\mathbf{v}=\mathbf{0}, \tag{1}$$

where $\mathbf{v}$ is the vector of state variables in $\mathbb{C}^n$, and $\mathbf{x}\in\mathbb{R}^d$ runs through the physical space; $d=1$, 2, or 3 is its dimension, but for this application one lets $d=3$ all along the paper. $\mathbf{A}(\mathbf{x},t)$ is a symmetric, positive-definite $n\times n$ real matrix; $\mathbf{D}_j$ are symmetric $n\times n$ real matrices independent of position $\mathbf{x}$ and time t; $\mathbf{B}(\mathbf{x},t)$ is a symmetric $n\times n$ real matrix; $\mathbf{\Omega}(\mathbf{x},t)$ is an $n\times n$ real matrix; and $*_t$ is the time convolution product $f*_tg(t)=\int_0^t f(\tau)g(t-\tau)\mathrm{d}\tau$, $t\geqslant 0$. The solution of Eq. (1) is denoted by $\mathbf{v}^\epsilon$ for initial conditions oscillating rapidly (as detailed below; see Sec. II A) at an arbitrarily small scale ϵ, considering slowly varying mechanical coefficients such that neither $\mathbf{A}$ nor $\mathbf{B}$ depend on ϵ. One also assumes that the sequence $(\mathbf{v}^\epsilon)$ is uniformly bounded in $(L^2(\mathbb{R}^3))^n$, where $L^2(\mathbb{R}^3)$ is the set of square integrable functions with respect to the space variable. Equation (1) is very general for wave equations arising from physical problems. It can embed Maxwell's equations for electromagnetic waves, the wave equation for acoustics, or Navier's equation of three-dimensional elastodynamics. For example, for an isotropic and synchronous visco-elastic medium characterized by its time-dependent Lamé's modulii $\lambda(\mathbf{x},t)$ and $\mu(\mathbf{x},t)$, and density $\varrho(\mathbf{x})$, the Navier equation of motion for the displacement vector $\mathbf{u}\in\mathbb{R}^3$ is

$$\varrho\,\partial_t^2\mathbf{u}=\mathrm{div}\,\boldsymbol{\sigma}(\mathbf{u}),$$

where the stress tensor $\boldsymbol{\sigma}$ is linked to the strain tensor $\boldsymbol{\epsilon}$ by Hooke's law

$$\boldsymbol{\sigma}=\lambda(\mathbf{x},0)\mathrm{div}\,\mathbf{u}\mathbf{I}_3+2\mu(\mathbf{x},0)\boldsymbol{\epsilon}+\partial_t\lambda(\mathbf{x},\cdot)*_t\mathrm{div}\,\mathbf{u}\mathbf{I}_3+2\partial_t\mu(\mathbf{x},\cdot)*_t\boldsymbol{\epsilon}.$$

Throughout the paper $\mathbf{I}_j$ is the $j\times j$ identity matrix. These equations have the form of Eq. (1), setting for example $\mathbf{v}=(\partial_t u_1,\partial_t u_2,\partial_t u_3,\epsilon_{11},\epsilon_{22},\epsilon_{33},\epsilon_{23},\epsilon_{13},\epsilon_{12},p)$ with $\epsilon_{ij}=\mu(\mathbf{x},0)(\partial_{\mathbf{x}_j}u_i+\partial_{\mathbf{x}_i}u_j)$ and $p=\lambda(\mathbf{x},0)\mathrm{div}\,\mathbf{u}$, together with $\mathbf{A}(\mathbf{x},t)=\mathrm{diag}(\varrho(\mathbf{x})\mathbf{I}_3,\ 1/2\mu(\mathbf{x},t)\mathbf{I}_3,\ 1/\mu(\mathbf{x},t)\mathbf{I}_3,\ 1/\lambda(\mathbf{x},t))$ and $\mathbf{B}(\mathbf{x},t)=\partial_t\mathbf{A}^{-1}(\mathbf{x},t)\mathbf{A}(\mathbf{x},0)$. Other definitions of $\mathbf{v}$ and $\mathbf{A}$ which lead to the same general form are obviously possible, as in Ref. 14, where the Navier equation of elastodynamics in $\mathbb{R}^d$ is written in the form of a symmetric hyperbolic system (1) with $n=d+1$.

A. Energy and power flow estimates

One injects high-frequency energy into the system described by Eq. (1) through, say, initial conditions of the form

$$\mathbf{v}^\epsilon(\mathbf{x},0)=\mathbf{v}_0^\epsilon(\mathbf{x}), \tag{2}$$

where the data $\mathbf{v}_0^\epsilon$ are bounded in $(L^2(\mathbb{R}^3))^n$ as ϵ goes to 0 and oscillating at most at the frequency $1/\epsilon$. For example, one can take $\mathbf{v}_0^\epsilon(\mathbf{x})=\mathbf{v}_0(\mathbf{x})e^{i\boldsymbol{\xi}\cdot\mathbf{x}/\epsilon}$, where $\mathbf{v}_0$ as well as its derivatives are in $(L^2(\mathbb{R}^3))^n$, or $\mathbf{v}_0^\epsilon(\mathbf{x})=\epsilon^{-3/2}\mathbf{v}_0(\mathbf{x}/\epsilon)$, where $\mathbf{v}_0$ is such that $\|\nabla\mathbf{v}_0\|_{L^2}<+\infty$, but more general conditions to be fulfilled by the initial data for the following to apply are given in Refs. 3 and 4. The energy density associated with the solution $\mathbf{v}^\epsilon$ of Eqs. (1)–(2) is defined by

$$\mathcal{E}^\epsilon(\mathbf{x},t)=\tfrac{1}{2}(\mathbf{v}^\epsilon,\mathbf{v}^\epsilon)_{\mathbf{A}(\mathbf{x},0)}, \tag{3}$$

while the densities of power flow along each direction $j=1$, 2, or 3 are

$$\Pi_j^\epsilon(\mathbf{x},t)=\tfrac{1}{2}(\mathbf{v}^\epsilon,\mathbf{v}^\epsilon)_{\mathbf{D}_j}, \tag{4}$$

with the notation $(\mathbf{u},\mathbf{v})_{\mathbf{M}}:=(\mathbf{M}\mathbf{u},\overline{\mathbf{v}})$, where $\overline{\mathbf{v}}$ is the complex conjugate of $\mathbf{v}$. A potential good candidate to study the high-frequency energetic features of the system (1)–(2) is the Wigner transform, because it has a direct link with the above quantities in the limit $\epsilon\to 0$, as well as some other nice evolution properties given below (see Sec. II B). Let us define for two functions $\mathbf{u}$ and $\mathbf{v}$ lying in a bounded subset of $(L^2(\mathbb{R}^3))^n$

$$\mathbf{W}^\epsilon(\mathbf{u},\mathbf{v})(\mathbf{x},\mathbf{k},t)=\frac{1}{(2\pi)^3}\int_{\mathbb{R}^3}e^{i\mathbf{k}\cdot\mathbf{y}}\mathbf{u}\left(\mathbf{x}-\frac{1}{2}\epsilon\mathbf{y},t\right)\otimes\overline{\mathbf{v}\left(\mathbf{x}+\frac{1}{2}\epsilon\mathbf{y},t\right)}\mathrm{d}\mathbf{y}. \tag{5}$$

Then, the $n\times n$ matrix-valued Wigner transform $\mathbf{W}^\epsilon[\mathbf{v}^\epsilon]:=\mathbf{W}^\epsilon(\mathbf{v}^\epsilon,\mathbf{v}^\epsilon)$ of $\mathbf{v}^\epsilon$ has (up to an extracted subsequence) a Hermitian limit $\mathbf{W}^\epsilon[\mathbf{v}^\epsilon]\rightharpoonup\mathbf{W}$ as $\epsilon\to 0$ in the sense of tempered distributions which is also a non-negative measure,[3,4] the so-called Wigner measure of $\mathbf{v}^\epsilon$.

Assuming that $\mathbf{v}^\epsilon$ of Eqs. (1)–(2) has the same oscillatory properties as the initial conditions, it can be shown, owing to Proposition 1.7 of Ref. 4 or Proposition 2.3.3 of Ref. 7, that the energy $\mathcal{E}(\mathbf{x},t)$ and power flow $\Pi_j(\mathbf{x},t)$ densities in the limit $\epsilon\to 0$ are estimated from the Wigner measure by

$$\lim_{\epsilon\to 0}\mathcal{E}^\epsilon(\mathbf{x},t):=\mathcal{E}(\mathbf{x},t)=\frac{1}{2}\int_{\mathbb{R}^3}\mathbf{A}(\mathbf{x},0):\mathbf{W}(\mathbf{x},\mathbf{k},t)\mathrm{d}\mathbf{k}, \tag{6}$$

and

$$\lim_{\epsilon\to 0}\Pi_j^\epsilon(\mathbf{x},t):=\Pi_j(\mathbf{x},t)=\frac{1}{2}\int_{\mathbb{R}^3}\mathbf{D}_j:\mathbf{W}(\mathbf{x},\mathbf{k},t)\mathrm{d}\mathbf{k}, \tag{7}$$

for $j=1$, 2, 3, respectively, where $\mathbf{A}:\mathbf{B}=\mathrm{Tr}(\mathbf{A}\mathbf{B}^{\mathrm{T}})$ is the usual tensor scalar product and Tr is the trace of a matrix.

B. Transport properties of Wigner measures

One now turns to the evolution properties of the Wigner measure associated with the sequence $(\mathbf{u}^{\epsilon})$ of solutions of the system (1)–(2). Define $[\mathbf{A},\mathbf{B}]=\mathbf{A}\mathbf{B}+\mathbf{B}^{*}\mathbf{A}^{*}$, where $\mathbf{A}^{*}=\overline{\mathbf{A}}^{\mathrm{T}}$ stands for the adjoint of a matrix. The Wigner measure satisfies the dispersion equation

$$[\mathrm{i}\mathbf{L},\mathbf{W}]=\mathbf{0}, \tag{8}$$

where $\mathbf{L}$ is the dispersion matrix of the above system defined by

$$\mathbf{L}(\mathbf{x},\mathbf{k})=\mathbf{A}^{-1}(\mathbf{x},0)\sum_{j=1}^{3} k_j\mathbf{D}_j\,. \tag{9}$$

It is self-adjoint with respect to the scalar product $(\mathbf{u},\mathbf{v})_{\mathbf{A}(\mathbf{x},0)}$; thus, its n right eigenvectors form an orthonormal family with respect to this scalar product. The latter and the associated real eigenvalues are denoted by $\mathbf{b}_{\alpha_r}(\mathbf{x},\mathbf{k})$ and $\omega_{\alpha_r}(\mathbf{x},\mathbf{k})$, respectively, for $1\leqslant\alpha\leqslant m\leqslant n$ and $1\leqslant r\leqslant R_\alpha$, where R_α is the order of multiplicity of eigenvalue ω_α for a mode α among the m existing ones, such that $\Sigma_{\alpha=1}^{m}R_\alpha=n$. R_α is assumed to be independent of $(\mathbf{x},\mathbf{k})$. From the dispersion equation the Wigner measure may be written

$$\mathbf{W}(\mathbf{x},\mathbf{k},t)=\sum_{\alpha=1}^{m}\sum_{r,s=1}^{R_\alpha}\mathrm{w}_{\alpha}^{rs}(\mathbf{x},\mathbf{k},t)\mathbf{b}_{\alpha_r}(\mathbf{x},\mathbf{k})\otimes\mathbf{b}_{\alpha_s}(\mathbf{x},\mathbf{k}), \tag{10}$$

where the w_{α}^{rs}'s are scalar coefficients, hereafter refered to as specific intensities. If one defines coherence matrices $\mathbf{W}_\alpha$, $1\leqslant\alpha\leqslant m$, by

$$W_{\alpha,rs}(\mathbf{x},\mathbf{k},t)=\mathrm{w}_{\alpha}^{rs}(\mathbf{x},\mathbf{k},t),\quad 1\leqslant r,s\leqslant R_\alpha, \tag{11}$$

then they will satisfy Liouville-type transport equations

$$\partial_t\mathbf{W}_\alpha+\{\omega_\alpha,\mathbf{W}_\alpha\}+[\mathbf{W}_\alpha,\mathbf{N}_\alpha]=\mathbf{0}. \tag{12}$$

$\{\ \}$ represents the usual Poisson's bracket defined by $\{f,g\}=\boldsymbol{\nabla}_{\mathbf{k}}f\cdot\boldsymbol{\nabla}_{\mathbf{x}}g-\boldsymbol{\nabla}_{\mathbf{x}}f\cdot\boldsymbol{\nabla}_{\mathbf{k}}g$ and the $\mathbf{N}_\alpha$'s are skew-symmetric matrices given by

$$N_{\alpha,rs}=\sum_{j=1}^{3}(\partial_{\mathbf{x}_j}\mathbf{b}_{\alpha_r},\mathbf{b}_{\alpha_s})_{\mathbf{D}_j}-(\boldsymbol{\nabla}_{\mathbf{x}}\omega_\alpha\cdot\boldsymbol{\nabla}_{\mathbf{k}}\mathbf{b}_{\alpha_r},\mathbf{b}_{\alpha_s})_{\mathbf{A}(\mathbf{x},0)}$$

$$-\frac{1}{2}\boldsymbol{\nabla}_{\mathbf{k}}\cdot\boldsymbol{\nabla}_{\mathbf{x}}\omega_\alpha\delta_{rs}\,, \tag{13}$$

for $1\leqslant r,s\leqslant R_\alpha$. These rotation matrices arise from the curvature of rays in heterogeneous media, leading to rotation of the polarization vector around a ray as waves propagate.[5] It is understood that if $R_\alpha=1$ the last term in Eq. (12) always vanishes and the latter writes

$$\partial_t\mathrm{w}_\alpha+\{\omega_\alpha,\mathrm{w}_\alpha\}=0, \tag{14}$$

dropping subscripts r and s. These results show that the transport equations satisfied by specific intensities are basically unchanged from the case without time-convolution operators.[7]

III. TRANSPORT EQUATIONS FOR ISOTROPIC POROUS MEDIA

Elastic waves in a fluid-saturated porous solid are described by the model of Ref. 15, taking into account the approach of Ref. 16 as well. The macroscopically homogeneous and isotropic porous material is constituted by a solid phase and a fluid phase which saturates the open porous space. The solid phase includes a solid part constituted by a synchronous (Poisson's coefficient ν is independent of time) visco-elastic material and the closed porosity. Both phases have average displacements denoted by $\mathbf{u}_\mathrm{s}$ and $\mathbf{u}_\mathrm{f}$ for the solid and fluid ones, respectively. The averaging process corresponds to the homogenization at a macroscopic scale of the quantities observed at the microscopic scale of the medium where they are not necessarily continuous and materials are heterogeneous. The ratio of the volume occupied by the fluid phase with respect to the overall volume of the porous medium is given by its porosity Φ such that $0\leqslant\Phi<1$. The density of the porous material is $\varrho=(1-\Phi)\varrho_\mathrm{s}+\Phi\varrho_\mathrm{f}$, where ϱ_s and ϱ_f are the densities of the materials constituting the solid and fluid phases, respectively. One assumes that Φ, ϱ_s, and ϱ_f are constant over the space. The average fluid motion with respect to the solid phase is given by the filtration velocity $\dot{\mathbf{w}}=\Phi(\dot{\mathbf{u}}_\mathrm{f}-\dot{\mathbf{u}}_\mathrm{s})$, where $\dot{\mathbf{u}}$ is the time derivative of $\mathbf{u}$ under the assumption of small transformations of the medium. The latter occupies an open domain Ω of $\mathbb{R}^3$.

A. Basic equations

The constitutive equations are written[16] for the porous material

$$\boldsymbol{\sigma}^{\mathrm{tot}}=(\lambda_0+\partial_t\lambda*_t)\boldsymbol{\nabla}\cdot\mathbf{u}_\mathrm{s}\mathbf{I}_3+2(\mu_0+\partial_t\mu*_t)\boldsymbol{\nabla}\otimes_s\mathbf{u}_\mathrm{s}-p\boldsymbol{\mathcal{B}}, \tag{15}$$

and for the fluid phase

$$p+\beta*_t p=-M(\boldsymbol{\mathcal{B}}:\boldsymbol{\nabla}\otimes_s\mathbf{u}_\mathrm{s}+\boldsymbol{\nabla}\cdot\mathbf{w}). \tag{16}$$

The pore fluid pressure is p, $\boldsymbol{\sigma}^{\mathrm{tot}}$ is the total macroscopic stress tensor of the porous medium, λ and μ are the time-dependent Lamé's modulii of the solid phase with $\lambda_0=\lambda(0)$ and $\mu_0=\mu(0)$, β is a time-dependent compressibility coefficient of the fluid phase which accounts for thermal effects, M is Biot's modulus, and B is the coupling coefficient between phases such that the coupling tensor is $\boldsymbol{\mathcal{B}}=B\mathbf{I}_3$ in the isotropic case. By definition[16]

$$\frac{1}{M}=\frac{\Phi}{K_\mathrm{f}}+\frac{B-\Phi}{K_\mathrm{s}},\quad B=1-\frac{K_\mathrm{b}}{K_\mathrm{s}},$$

where K_f is the fluid bulk modulus, $K_\mathrm{b}=\lambda_0+(2/3)\mu_0$ is the macroscopic bulk modulus of the dry solid phase, whereas K_s is the microscopic bulk modulus of the solid grains constituting the skeleton of the porous material such that $K_\mathrm{s}\geqslant K_\mathrm{b}$. Last, $\otimes_s$ is the symmetrized tensor product. Let

$$q=\lambda_0\boldsymbol{\nabla}\cdot\mathbf{u}_\mathrm{s}-Bp, \tag{17a}$$

$$\boldsymbol{\epsilon}=2\mu_0\boldsymbol{\nabla}\otimes_s\mathbf{u}_\mathrm{s}, \tag{17b}$$

then

$$\nabla\cdot\dot{\mathbf{u}}_s=\frac{1}{\lambda_0}\dot{q}+\frac{B}{\lambda_0}\dot{p}, \tag{18a}$$

$$\nabla\cdot\dot{\mathbf{w}}=-\frac{B}{\lambda_0}\dot{q}-\left(\frac{1}{M}+\frac{B^2}{\lambda_0}\right)\dot{p}-\frac{\dot{\beta}}{M}*_t p. \tag{18b}$$

The dynamic equation of motion of the bulk mixture is written

$$\nabla\left(\boldsymbol{\epsilon}+\frac{\partial_t\mu}{\mu_0}*_t\boldsymbol{\epsilon}\right)+\nabla\left(q+\frac{\partial_t\lambda}{\lambda_0}*_t(q+Bp)\right)=\varrho\ddot{\mathbf{u}}_s+\varrho_f\ddot{\mathbf{w}}. \tag{19}$$

The linearized Darcy's law under the assumption of small transformations of the homogeneous medium is written

$$-\eta_f\dot{\varphi}\boldsymbol{\mathcal{K}}^{-1}*_t\dot{\mathbf{w}}=\nabla p+\varrho_f(\alpha\ddot{\mathbf{u}}_f+(1-\alpha)\ddot{\mathbf{u}}_s), \tag{20}$$

and describes the motion of the pore fluid relative to the porous skeleton. In the above equation, η_f is the pore fluid viscosity, φ is a time-dependent viscosity correction factor, and $\boldsymbol{\mathcal{K}}$ is the intrinsic permeability tensor which depends on the geometry of the pore network solely; it is spherical in the isotropic case $\boldsymbol{\mathcal{K}}=\mathcal{K}\mathbf{I}_3$, where $\mathcal{K}$ is the coefficient of permeability. At last, α is the tortuosity defined as the ratio of the microscopic mean-square velocity of fluid particles with respect to the skeleton to their macroscopic quadratic velocity $|\dot{\mathbf{u}}_f-\dot{\mathbf{u}}_s|^2$. It is a purely geometrical quantity independent of fluid or solid densities such that $\alpha\geqslant 1$. Regarding the introduction of convolution products in Eqs. (15), we refer the reader to Fung's book[17] for a detailed presentation of the behavior of viscous materials including memory effects. Regarding the introduction of convolution products in Eqs. (16) and (20), we refer to the work of Fellah and Depollier,[18] extending their results to the case of a nonrigid solid frame. In this study, φ and β are proportional to the function $t\mapsto 1/\sqrt{t}$ for $t\downarrow 0$ (singular memory kernels). Note that Darcy's law may also be written $-\mathbf{Y}\dot{\mathbf{w}}=\nabla p+\varrho_f\ddot{\mathbf{u}}_s$, where $\mathbf{Y}:=\eta_f\dot{\varphi}\boldsymbol{\mathcal{K}}^{-1}*_t+(\alpha\varrho_f/\Phi)\partial_t$ is the viscodynamic operator introduced in Ref. 19.

Now, Eqs. (17) through (20) can be recast in the form of a first-order hyperbolic system of Eq. (1) with $n=14$ and $d=3$, introducing the state variable $\mathbf{v}=(\dot{\mathbf{u}}_s,\dot{\mathbf{w}},\boldsymbol{\epsilon},q,-p)$ of $\mathbb{C}^{14}$, where $\boldsymbol{\epsilon}$ is rearranged in a vector form as $\boldsymbol{\epsilon}=(\epsilon_{11},\epsilon_{22},\epsilon_{33},\epsilon_{23},\epsilon_{13},\epsilon_{12})$. This leads to the equation

$$\mathbf{A}_0\partial_t\mathbf{v}+\sum_{j=1}^{3}\mathbf{D}_j\partial_{x_j}(\mathbf{v}+\mathbf{B}*_t\mathbf{v})+\boldsymbol{\Omega}*_t\mathbf{v}=\mathbf{0}, \tag{21}$$

where $\mathbf{A}_0=\mathbf{A}(0)$, with

$$\mathbf{A}(t)=\begin{bmatrix}
\varrho\mathbf{I}_3 & \varrho_f\mathbf{I}_3 & 0 & 0 & 0 & 0\\
\varrho_f\mathbf{I}_3 & \dfrac{\alpha\varrho_f}{\Phi}\mathbf{I}_3 & 0 & 0 & 0 & 0\\
0 & 0 & \dfrac{1}{2\mu(t)}\mathbf{I}_3 & 0 & 0 & 0\\
0 & 0 & 0 & \dfrac{1}{\mu(t)}\mathbf{I}_3 & 0 & 0\\
0 & 0 & 0 & 0 & \dfrac{1}{\lambda(t)} & -\dfrac{B}{\lambda(t)}\\
0 & 0 & 0 & 0 & -\dfrac{B}{\lambda(t)} & \dfrac{1}{M}+\dfrac{B^2}{\lambda(t)}
\end{bmatrix},$$

$\mathbf{B}=\partial_t(\mathbf{A}^{-1})\mathbf{A}_0$ and a diagonal matrix $\boldsymbol{\Omega}$ given by $\boldsymbol{\Omega}=\mathrm{diag}(0,\eta_f\dot{\varphi}\boldsymbol{\mathcal{K}}^{-1},0,0,0,\dot{\beta}/M)$. Matrices $\mathbf{D}_1$, $\mathbf{D}_2$, and $\mathbf{D}_3$ are symmetric and have the form

$$-\mathbf{D}_j=\begin{bmatrix}
0 & 0 & \mathbf{K}_j & \mathbf{M}_j & \mathbf{k}_j & 0\\
 & 0 & 0 & 0 & 0 & \mathbf{k}_j\\
 & & 0 & 0 & 0 & 0\\
 & & & 0 & 0 & 0\\
 & & & & 0 & 0\\
 & & & & & 0
\end{bmatrix},$$

where $\mathbf{K}_j$ and $\mathbf{M}_j$ are the 3×3 matrices given by their elements $K_j^{rs}=\delta_{jr}\delta_{js}$ and $M_j^{rs}=(1-\delta_{jr})(1-\delta_{js})(1-\delta_{rs})$, and $\mathbf{k}_j$ is the vector given by its components $k_j^r=\delta_{jr}$, with $r, s\in\{1,2,3\}$ and δ_{rs} being the Kronecker symbol. The results of the previous section are applied to this system under the assumption that the wavelength, scaled by ϵ, remains much larger than the pore and solid grain sizes in order to ensure the linearity of the proposed model. One also considers "ϵ-oscillating" initial data $\mathbf{v}(\mathbf{x},0)=\mathbf{v}_0^{\epsilon}(\mathbf{x})$ with $\mathbf{v}=\mathbf{0}$ for $t<0$.

B. High-frequency energy transport properties

The dispersion matrix for a porous medium thus writes

$$-\mathbf{L}(\mathbf{k})=\begin{bmatrix} 0 & 0 & \frac{1}{\varrho_{\mathrm{sf}}}\mathbf{K}(\mathbf{k}) & \frac{1}{\varrho_{\mathrm{sf}}}\mathbf{M}(\mathbf{k}) & \frac{1}{\varrho_{\mathrm{sf}}}\mathbf{k} & -\frac{\Phi}{\alpha\varrho_{\mathrm{sf}}}\mathbf{k} \\ 0 & 0 & -\frac{\Phi}{\alpha\varrho_{\mathrm{sf}}}\mathbf{K}(\mathbf{k}) & -\frac{\Phi}{\alpha\varrho_{\mathrm{sf}}}\mathbf{M}(\mathbf{k}) & -\frac{\Phi}{\alpha\varrho_{\mathrm{sf}}}\mathbf{k} & \frac{\Phi\varrho}{\alpha\varrho_{\mathrm{sf}}\varrho_{\mathrm{f}}}\mathbf{k} \\ 2\mu_0\mathbf{K}(\mathbf{k}) & 0 & 0 & 0 & 0 & 0 \\ \mu_0\mathbf{M}(\mathbf{k}) & 0 & 0 & 0 & 0 & 0 \\ (\lambda_0+MB^2)\mathbf{k}^{\mathrm{T}} & MB\mathbf{k}^{\mathrm{T}} & 0 & 0 & 0 & 0 \\ MB\mathbf{k}^{\mathrm{T}} & M\mathbf{k}^{\mathrm{T}} & 0 & 0 & 0 & 0 \end{bmatrix}, \tag{22}$$

where $\alpha\varrho_{\mathrm{sf}}=\alpha(1-\Phi)\varrho_{\mathrm{s}}+(\alpha-1)\Phi\varrho_{\mathrm{f}}$, $\mathbf{k}=(k_1,k_2,k_3)$, and

$$\mathbf{K}(\mathbf{k})=\sum_{j=1}^{3}k_j\mathbf{K}_j=\mathrm{diag}(k_1,k_2,k_3),$$

$$\mathbf{M}(\mathbf{k})=\sum_{j=1}^{3}k_j\mathbf{M}_j=\begin{bmatrix} 0 & k_3 & k_2 \\ k_3 & 0 & k_1 \\ k_2 & k_1 & 0 \end{bmatrix}.$$

Let $|\mathbf{k}|=\sqrt{k_1^2+k_2^2+k_3^2}$ and $\hat{\mathbf{k}}=\mathbf{k}/|\mathbf{k}|$. One introduces vectors $\hat{\mathbf{z}}_1(\mathbf{k})$ and $\hat{\mathbf{z}}_2(\mathbf{k})$ such that $(\hat{\mathbf{k}},\hat{\mathbf{z}}_1,\hat{\mathbf{z}}_2)$ form an orthonormal triplet, for example $\hat{\mathbf{k}}=(\sin\theta\cos\phi,\sin\theta\sin\phi,\cos\theta)$, $\hat{\mathbf{z}}_1(\mathbf{k})=(\cos\theta\cos\phi,\cos\theta\sin\phi,-\sin\theta)$ and $\hat{\mathbf{z}}_2(\mathbf{k})=(-\sin\phi,\cos\phi,0)$ in spherical coordinates. Then, the eigenvalues of the dispersion matrix in the set $E=\mathbb{R}^3\times\mathbb{R}^3\backslash\{(\mathbf{x},\mathbf{k});\mathbf{k}=\mathbf{0}\}$ are

$$\begin{aligned} &\omega_0=0 \quad \text{with multiplicity 6,} \\ &\omega_{\mathrm{S}}^{\pm}=\pm c_{\mathrm{S}}|\mathbf{k}| \quad \text{each with multiplicity 2,} \\ &\omega_{\mathrm{P1}}^{\pm}=\pm c_{\mathrm{P1}}|\mathbf{k}| \quad \text{each with multiplicity 1,} \\ &\omega_{\mathrm{P2}}^{\pm}=\pm c_{\mathrm{P2}}|\mathbf{k}| \quad \text{each with multiplicity 1,} \end{aligned} \tag{23}$$

where c_{P1}^2 and c_{P2}^2 are the solutions of

$$X^2-(c_{\mathrm{P}}^2+c_1^2)X+c_{\mathrm{P}}^2c_2^2=0, \tag{24}$$

with parameters $c_{\mathrm{S}}^2=\mu_0/\varrho_{\mathrm{sf}}$, $c_{\mathrm{P}}^2=(\lambda_0+2\mu_0)/\varrho_{\mathrm{sf}}$, and

$$c_1^2=\frac{M}{\varrho_{\mathrm{sf}}}\left[B^2+\frac{\Phi}{\alpha}\left(\frac{\varrho-2B\varrho_{\mathrm{f}}}{\varrho_{\mathrm{f}}}\right)\right],$$

$$c_2^2=\frac{\Phi M}{\alpha\varrho_{\mathrm{f}}}=c_1^2-\frac{M}{\varrho_{\mathrm{sf}}}\left(B-\frac{\Phi}{\alpha}\right)^2.$$

These results agree with those of Refs. 20 and 21. One chooses $c_{\mathrm{P2}}<c_{\mathrm{P1}}$ in accordance with the usual notations of the dedicated literature. Therefore, $m=7$ and $R_0=6$, $R_{\mathrm{S}^+}=R_{\mathrm{S}^-}=2$ with the notations of Sec. II B. The associated eigenvectors on E are

$$\mathbf{b}_r=\left(0,0,\sqrt{2\mu_0}\mathbf{K}(\hat{\mathbf{z}}_r)\hat{\mathbf{z}}_r,\sqrt{\frac{\mu_0}{2}}\mathbf{M}(\hat{\mathbf{z}}_r)\hat{\mathbf{z}}_r,0,0\right),$$

$$\mathbf{b}_3=(0,0,2\sqrt{\mu_0}\mathbf{K}(\hat{\mathbf{z}}_1)\hat{\mathbf{z}}_2,\sqrt{\mu_0}\mathbf{M}(\hat{\mathbf{z}}_1)\hat{\mathbf{z}}_2,0,0),$$

$$\begin{aligned} \mathbf{b}_4=\Bigg(&0,0,\sqrt{\frac{2\lambda_0\mu_0}{\lambda_0+2\mu_0}}\mathbf{K}(\hat{\mathbf{k}})\hat{\mathbf{k}},\sqrt{\frac{\lambda_0\mu_0}{2(\lambda_0+2\mu_0)}}\mathbf{M}(\hat{\mathbf{k}})\hat{\mathbf{k}}, \\ &-\sqrt{\frac{2\lambda_0\mu_0}{\lambda_0+2\mu_0}},0\Bigg), \\ \mathbf{b}_{4+r}=\Bigg(&0,\sqrt{\frac{\Phi}{\alpha\varrho_{\mathrm{f}}}}\hat{\mathbf{z}}_r,0,0,0,0\Bigg), \\ \mathbf{b}_{\mathrm{S}r}^{\pm}=\Bigg(&\frac{\hat{\mathbf{z}}_r}{\sqrt{2\varrho_{\mathrm{sf}}}},-\frac{\Phi\hat{\mathbf{z}}_r}{\alpha\sqrt{2\varrho_{\mathrm{sf}}}},\mp\sqrt{2\mu_0}\mathbf{K}(\hat{\mathbf{k}})\hat{\mathbf{z}}_r, \\ &\mp\sqrt{\frac{\mu_0}{2}}\mathbf{M}(\hat{\mathbf{k}})\hat{\mathbf{z}}_r,0,0\Bigg), \\ \mathbf{b}_{\mathrm{P}j}^{\pm}=\Bigg(&\frac{\hat{\mathbf{k}}}{\sqrt{2\varrho_{\mathrm{sf}}}},-\frac{\Phi\hat{\mathbf{k}}}{\alpha\sqrt{2\varrho_{\mathrm{sf}}}},\mp\frac{2\mu_0\mathbf{K}(\hat{\mathbf{k}})\hat{\mathbf{k}}}{c_{\mathrm{P}_j}\sqrt{2\varrho_{\mathrm{sf}}}},\mp\frac{\mu_0\mathbf{M}(\hat{\mathbf{k}})\hat{\mathbf{k}}}{c_{\mathrm{P}_j}\sqrt{2\varrho_{\mathrm{sf}}}}, \\ &\mp\frac{\lambda_0+MB\left(B-\frac{\Phi}{\alpha}\right)}{c_{\mathrm{P}j}\sqrt{2\varrho_{\mathrm{sf}}}},\mp\frac{M\left(B-\frac{\Phi}{\alpha}\right)}{c_{\mathrm{P}j}\sqrt{2\varrho_{\mathrm{sf}}}}\Bigg), \end{aligned} \tag{25}$$

for $r,j\in\{1,2\}$. Eigenvectors $\mathbf{b}_5$ and $\mathbf{b}_6$ correspond to acoustic transverse advection modes of the fluid phase, and are similar to the nonpropagating modes obtained for a purely acoustic medium.[5] Eigenmodes $\mathbf{b}_1$ through $\mathbf{b}_4$ are their counterparts for a purely elastic medium. The main changes introduced by porosity compared to the elastic case are (i) the emergence of two compressional modes, instead of a single one, and (ii) the introduction of an effective density ϱ_{sf} in the definition of the shear and compression modes, which are similar to the elastic modes omitting this modification. The purely elastic medium is recovered if one lets $\Phi=0$ and $\alpha=1$, as one also has $B=0$ in this limit. Finally, one notes that these modes are nondispersive in accordance with some experimental observations done for two different porous materials (plastic foams).[22]

Regarding the slow-wave speed c_{P2}, it is null for $\Phi=0$ and its value for $\Phi\neq0$ and $\varrho_{\mathrm{f}}/\varrho_{\mathrm{s}}\ll1$ or $K_{\mathrm{f}}/K_{\mathrm{b}}\ll1$ approaches the modified velocity of sound in the air filling the pores $c_{\mathrm{f}}=\sqrt{K_{\mathrm{f}}/\alpha\varrho_{\mathrm{f}}}=c/\sqrt{\alpha}$, in accordance with the high-frequency asymptotic estimate given in Refs. 20 and 21. Thus, one has $c_{\mathrm{P2}}\leqslant c$. For such ratios of fluid to solid phase

bulk modulii, it may be observed that the coupling coefficient B has almost no influence and could be fixed at 1.

Now, let $w_{Pj}^{\pm}(\mathbf{x},\mathbf{k},t)$, $j\in\{1,2\}$, be the specific intensities for compressional waves corresponding to eigenvectors $\mathbf{b}_{Pj}^{\pm}(\mathbf{k})$. They then satisfy the following transport equations:

$$\partial_t w_{Pj}^{\pm} \pm c_{Pj}\hat{\mathbf{k}}\cdot\nabla_{\mathbf{x}} w_{Pj}^{\pm}=0, \quad j\in\{1,2\}. \tag{26}$$

Let $\mathbf{W}_S^{\pm}(\mathbf{x},\mathbf{k},t)$ be the 2×2 coherence matrices of specific intensities corresponding to eigenvectors $\mathbf{b}_{Sr}^{\pm}(\mathbf{k})$, $r\in\{1,2\}$. Thus, they satisfy the following transport equations:

$$\partial_t \mathbf{W}_S^{\pm} \pm c_S\hat{\mathbf{k}}\cdot\nabla_{\mathbf{x}}\mathbf{W}_S^{\pm}=\mathbf{0}, \tag{27}$$

since rotation matrix $\mathbf{N}_S$ is null for homogeneous parameters, and the contributions of the convolution operator $\mathbf{\Omega}$ cancel out in the limit $\epsilon\to0$. Finally, the 6×6 coherence matrix $\mathbf{W}_0(\mathbf{x},\mathbf{k},t)$ satisfies $\partial_t\mathbf{W}_0=\mathbf{0}$, since here again $\mathbf{N}_0=\mathbf{0}$; thus, $\mathbf{W}_0(\mathbf{x},\mathbf{k},t)=\mathbf{0}$ if it is initially null. One observes that $w_{Pj}^{-}(\mathbf{x},\mathbf{k},t)=w_{Pj}^{+}(\mathbf{x},-\mathbf{k},t)$, $j\in\{1,2\}$, and $\mathbf{W}_S^{-}(\mathbf{x},\mathbf{k},t)=\mathbf{W}_S^{+}(\mathbf{x},-\mathbf{k},t)$, so that the energy density is estimated by

$$\mathcal{E}(\mathbf{x},t)=\sum_{j=1,2}\int_{\mathbb{R}^3} w_{Pj}^{+}(\mathbf{x},\mathbf{k},t)\,d\mathbf{k}+\int_{\mathbb{R}^3}\mathrm{Tr}\,\mathbf{W}_S^{+}(\mathbf{x},\mathbf{k},t)\,d\mathbf{k} +\frac{1}{2}\int_{\mathbb{R}^3}\mathrm{Tr}\,\mathbf{W}_0(\mathbf{x},\mathbf{k},t)\,d\mathbf{k}, \tag{28}$$

while the power flow density vector is estimated by

$$\mathbf{\Pi}(\mathbf{x},t)=\sum_{j=1,2} c_{Pj}\int_{\mathbb{R}^3} w_{Pj}^{+}(\mathbf{x},\mathbf{k},t)\hat{\mathbf{k}}\,d\mathbf{k} +c_S\int_{\mathbb{R}^3}\mathrm{Tr}\,\mathbf{W}_S^{+}(\mathbf{x},\mathbf{k},t)\hat{\mathbf{k}}\,d\mathbf{k}, \tag{29}$$

since the modes $\mathbf{b}_1(\mathbf{k})$ through $\mathbf{b}_6(\mathbf{k})$ do not propagate.

IV. THE RANDOM CASE: RADIATIVE TRANSFER

A. General considerations

In this section one considers some randomness of the mechanical properties of the medium and follows again the developments of Refs. 5 and 6. In their work these authors introduce a random perturbation $(\mathbf{V}(\mathbf{x}),\mathbf{x}\in\mathbb{R}^3)$ of matrix $\mathbf{A}$ of the form

$$\mathbf{A}(\mathbf{x})=\underline{\mathbf{A}}(\mathbf{x})\left[\mathbf{I}_n+\sqrt{\epsilon}\mathbf{V}\left(\frac{\mathbf{x}}{\epsilon}\right)\right]. \tag{30}$$

The magnitude $\sqrt{\epsilon}$ of perturbations is chosen such that they have comparable effects on the unperturbed background medium as characterized by the mean matrix of parameters $\underline{\mathbf{A}}(\mathbf{x})$. However, they need not be too strong, for in this case localization rather than transport may occur. Their correlation lengths are chosen to be comparable to wavelengths ϵ to allow strong interactions between the waves and the heterogeneities. For the system (1) to remain symmetric hyperbolic, the perturbation matrix has to satisfy the condition $\underline{\mathbf{A}}(\mathbf{x})\mathbf{V}(\mathbf{y})=\mathbf{V}^*(\mathbf{y})\underline{\mathbf{A}}(\mathbf{x})$ for all $\mathbf{x}$ and $\mathbf{y}$ of $\mathbb{R}^3$. Parameter fluctuations are modeled with a statistically homogeneous, second-order matrix-valued random process with mean zero whose covariance functions are denoted by $R_{jklm}(\mathbf{x}-\mathbf{y}):=\mathbb{E}\{V_{jk}(\mathbf{x})V_{lm}(\mathbf{y})\}$, such that

$$R_{jklm}(\mathbf{x}-\mathbf{y})=\int_{\mathbb{R}^3} e^{-i\mathbf{k}\cdot(\mathbf{x}-\mathbf{y})}S_{jklm}(\mathbf{k})\,d\mathbf{k}. \tag{31}$$

$\mathbb{E}\{\cdot\}$ stands for mathematical expectation and $\mathbf{k}\mapsto S_{jklm}(\mathbf{k})$ are the spectral density functions of random matrix $(\mathbf{V}(\mathbf{x}),\mathbf{x}\in\mathbb{R}^3)$ for j, k, l, $m\in\{1,2,...,n\}$. They satisfy the usual properties of spectral densities of homogeneous stochastic processes: $S_{jklm}(\mathbf{k})=\overline{S_{lmjk}(\mathbf{k})}=S_{lmjk}(-\mathbf{k})$. Let $\mathbf{S}$ be the fourth-order spectral density tensor corresponding to the above spectral density functions. Thus, it can be shown that the coherence matrices $\mathbf{W}_\alpha$, $1\leqslant\alpha\leqslant m$, satisfy coupled radiative transfer equations given by

$$\partial_t\mathbf{W}_\alpha+\{\omega_\alpha,\mathbf{W}_\alpha\}+[\mathbf{W}_\alpha,\mathbf{N}_\alpha] =\sum_{\beta=1}^{m}\int_{\mathbb{R}^3}[\sigma_{\alpha\beta}(\mathbf{k},\mathbf{k}')\mathbf{W}_\beta(\mathbf{k}')]\,\delta(\omega_\alpha(\mathbf{k})-\omega_\beta(\mathbf{k}'))\,d\mathbf{k}' -[\mathbf{\Sigma}_\alpha,\mathbf{W}_\alpha]. \tag{32}$$

The differential scattering operators $\sigma_{\alpha\beta}(\mathbf{k},\mathbf{k}')$ are defined by

$$[\sigma_{\alpha\beta}(\mathbf{k},\mathbf{k}')\mathbf{W}_\beta(\mathbf{k}')]=2\pi\omega_\alpha^2(\mathbf{k})\times\mathcal{S}_{\alpha\beta}(\mathbf{k},\mathbf{k}')\mathbf{W}_\beta(\mathbf{k}'), \tag{33}$$

where $\mathcal{S}_{\alpha\beta}(\mathbf{k},\mathbf{k}')$ is a fourth-order tensor with elements

$$\mathcal{S}_{rstu}^{\alpha\beta}(\mathbf{k},\mathbf{k}')=\mathbf{c}_{\alpha_r}(\mathbf{k})\otimes\mathbf{b}_{\beta_t}(\mathbf{k}'):\mathbf{S}(\mathbf{k}-\mathbf{k}'):\mathbf{c}_{\alpha_s}(\mathbf{k})\otimes\mathbf{b}_{\beta_u}(\mathbf{k}'). \tag{34}$$

The total scattering operators $\mathbf{\Sigma}_\alpha(\mathbf{k})$ are defined by

$$\mathbf{\Sigma}_\alpha(\mathbf{k})=\frac{1}{2}\sum_{\beta=1}^{m}\left\{\int_{\mathbb{R}^3}[\sigma_{\alpha\beta}(\mathbf{k},\mathbf{k}')\mathbf{I}_{R_\beta}]\,\delta(\omega_\alpha(\mathbf{k})-\omega_\beta(\mathbf{k}'))\,d\mathbf{k}' -i\int_{\mathbb{R}^3}\frac{1}{\omega_\alpha(\mathbf{k})-\omega_\beta(\mathbf{k}')}[\sigma_{\alpha\beta}(\mathbf{k},\mathbf{k}')\mathbf{I}_{R_\beta}]\,d\mathbf{k}'\right\}. \tag{35}$$

$\mathbf{b}_{\alpha_r}$ are the right eigenvectors of the mean dispersion matrix and $\mathbf{c}_{\alpha_r}$ are its left eigenvectors, given by $\mathbf{c}_{\alpha_r}=\underline{\mathbf{A}}\mathbf{b}_{\alpha_r}$, for $1\leqslant\alpha\leqslant m$ and $1\leqslant r\leqslant R_\alpha$. The differential scattering operator $\sigma_{\alpha\beta}(\mathbf{x},\mathbf{k},\mathbf{k}')$ gives the rate at which the energy mode β with wave vector $\mathbf{k}'$ is converted to the energy mode α with wave vector $\mathbf{k}$, at position $\mathbf{x}$; the total scattering operator $\mathbf{\Sigma}_\alpha(\mathbf{x},\mathbf{k})$ gives the rate at which the energy mode α with wave vector $\mathbf{k}$ at position $\mathbf{x}$ is converted to all other modes or wave numbers (thus, it can account for intrinsic damping). The conservation of energy without dissipation is implied by the fact that one formally has $\sigma_{\alpha\beta}(\mathbf{k},\mathbf{k}')\equiv\sigma_{\beta\alpha}(\mathbf{k}',\mathbf{k})$.

B. Radiative transfer equations for isotropic porous media

Let us consider a random perturbation matrix $(\mathbf{V}(\mathbf{x}),\mathbf{x}\in\mathbb{R}^3)$ having the form

$$\mathbf{V}=\begin{bmatrix} \tilde{\varrho}\mathbf{I}_3 & -\frac{\alpha}{\Phi}\tilde{F}\mathbf{I}_3 & 0 & 0 & 0 & 0 \\ -\frac{\Phi}{\alpha}\tilde{\varrho}\mathbf{I}_3 & \frac{\alpha\varrho}{\Phi\varrho_{\mathrm{f}}}\tilde{F}\mathbf{I}_3 & 0 & 0 & 0 & 0 \\ 0 & 0 & \tilde{\mu}\mathbf{I}_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & \tilde{\mu}\mathbf{I}_3 & 0 & 0 \\ 0 & 0 & 0 & 0 & \tilde{\lambda} & -B\tilde{\lambda} \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

where $(\tilde{\varrho}(\mathbf{x}),\mathbf{x}\in\mathbb{R}^3)$, $(\tilde{F}(\mathbf{x}),\mathbf{x}\in\mathbb{R}^3)$, $(\tilde{\mu}(\mathbf{x}),\mathbf{x}\in\mathbb{R}^3)$, and $(\tilde{\lambda}(\mathbf{x}),\mathbf{x}\in\mathbb{R}^3)$ are four real-valued, statistically homogeneous stochastic processes whose spectral measures are characterized by their power spectral density functions $\mathbf{k}\mapsto S_\varrho(\mathbf{k})$, $\mathbf{k}\mapsto S_F(\mathbf{k})$, $\mathbf{k}\mapsto S_\mu(\mathbf{k})$, and $\mathbf{k}\mapsto S_\lambda(\mathbf{k})$, respectively, and cross-spectral density functions $\mathbf{k}\mapsto S_{\varrho F}(\mathbf{k})$, $\mathbf{k}\mapsto S_{\varrho\mu}(\mathbf{k})$, $\mathbf{k}\mapsto S_{\varrho\lambda}(\mathbf{k})$, $\mathbf{k}\mapsto S_{F\mu}(\mathbf{k})$, $\mathbf{k}\mapsto S_{F\lambda}(\mathbf{k})$, and $\mathbf{k}\mapsto S_{\lambda\mu}(\mathbf{k})$. This model of perturbations corresponds to random fluctuations of the mean density ϱ, the mean inverse formation factor F^{-1} (defined by $F=\alpha/\Phi$), and the mean inverse Lamé's moduli μ_0^{-1} and λ_0^{-1}. Therefore, the differential scattering operator for $\mathrm{P}j$ to $\mathrm{P}k$ mode conversion, $j\neq k\in\{1,2\}$, is a simple multiplicative operator

$$[\sigma_{\mathrm{P}j\mathrm{P}k}(\mathbf{k},\mathbf{k}')\mathrm{w}_{\mathrm{P}k}(\mathbf{k}')]=\sigma_{\mathrm{pp}}(\mathbf{k},\mathbf{k}')\mathrm{w}_k(\mathbf{k}'), \tag{36}$$

characterized by the (scalar) scattering cross section

$$\begin{aligned}\sigma_{\mathrm{pp}}(\mathbf{k},\mathbf{k}')=\frac{\pi}{2}c_{\mathrm{P}}^2|\mathbf{k}||\mathbf{k}'|\Bigg[&\frac{c_2}{c_{\mathrm{P}}}(\hat{\mathbf{k}}\cdot\hat{\mathbf{k}}')^2(S_\varrho(\mathbf{k}-\mathbf{k}')+S_F(\mathbf{k}-\mathbf{k}')+2\,\mathrm{Re}\,S_{\varrho F}(\mathbf{k}-\mathbf{k}'))+4\frac{c_{\mathrm{S}}^4}{c_{\mathrm{P}}^3c_2}(\hat{\mathbf{k}}\cdot\hat{\mathbf{k}}')^4S_\mu(\mathbf{k}-\mathbf{k}')\\&+4\frac{c_{\mathrm{S}}^2}{c_{\mathrm{P}}^2}(\hat{\mathbf{k}}\cdot\hat{\mathbf{k}}')^3(\mathrm{Re}\,S_{\varrho\mu}(\mathbf{k}-\mathbf{k}')+\mathrm{Re}\,S_{F\mu}(\mathbf{k}-\mathbf{k}'))+2\left(1-2\frac{c_{\mathrm{S}}^2}{c_{\mathrm{P}}^2}\right)(\hat{\mathbf{k}}\cdot\hat{\mathbf{k}}')(\mathrm{Re}\,S_{\varrho\lambda}(\mathbf{k}-\mathbf{k}')\\&+\mathrm{Re}\,S_{F\lambda}(\mathbf{k}-\mathbf{k}'))+\frac{4c_{\mathrm{S}}^2}{c_{\mathrm{P}}c_2}\left(1-2\frac{c_{\mathrm{S}}^2}{c_{\mathrm{P}}^2}\right)(\hat{\mathbf{k}}\cdot\hat{\mathbf{k}}')^2\,\mathrm{Re}\,S_{\lambda\mu}(\mathbf{k}-\mathbf{k}')+\frac{c_{\mathrm{P}}}{c_2}\left(1-2\frac{c_{\mathrm{S}}^2}{c_{\mathrm{P}}^2}\right)^2S_\lambda(\mathbf{k}-\mathbf{k}')\Bigg].\end{aligned} \tag{37}$$

The differential scattering operator for $\mathrm{P}j$ scattering without mode conversion, $j\in\{1,2\}$, is also a multiplicative operator

$$[\sigma_{\mathrm{P}j\mathrm{P}j}(\mathbf{k},\mathbf{k}')\mathrm{w}_{\mathrm{P}j}(\mathbf{k}')]=\sigma_{\mathrm{pp}}^j(\mathbf{k},\mathbf{k}')\mathrm{w}_{\mathrm{P}j}(\mathbf{k}'), \tag{38}$$

characterized by the scattering cross section

$$\begin{aligned}\sigma_{\mathrm{pp}}^j(\mathbf{k},\mathbf{k}')=\frac{\pi}{2}c_{\mathrm{P}j}^2|\mathbf{k}|^2\Bigg[&(\hat{\mathbf{k}}\cdot\hat{\mathbf{k}}')^2(S_\varrho(\mathbf{k}-\mathbf{k}')+S_F(\mathbf{k}-\mathbf{k}')+2\,\mathrm{Re}\,S_{\varrho F}(\mathbf{k}-\mathbf{k}'))+4\frac{c_{\mathrm{S}}^4}{c_{\mathrm{P}j}^4}(\hat{\mathbf{k}}\cdot\hat{\mathbf{k}}')^4S_\mu(\mathbf{k}-\mathbf{k}')\\&+4\frac{c_{\mathrm{S}}^2}{c_{\mathrm{P}j}^2}(\hat{\mathbf{k}}\cdot\hat{\mathbf{k}}')^3(\mathrm{Re}\,S_{\varrho\mu}(\mathbf{k}-\mathbf{k}')+\mathrm{Re}\,S_{F\mu}(\mathbf{k}-\mathbf{k}'))+2\frac{c_{\mathrm{P}}^2}{c_{\mathrm{P}j}^2}\left(1-2\frac{c_{\mathrm{S}}^2}{c_{\mathrm{P}}^2}\right)(\hat{\mathbf{k}}\cdot\hat{\mathbf{k}}')(\mathrm{Re}\,S_{\varrho\lambda}(\mathbf{k}-\mathbf{k}')\\&+\mathrm{Re}\,S_{F\lambda}(\mathbf{k}-\mathbf{k}'))+4\frac{c_{\mathrm{S}}^2c_{\mathrm{P}}^2}{c_{\mathrm{P}j}^4}\left(1-2\frac{c_{\mathrm{S}}^2}{c_{\mathrm{P}}^2}\right)(\hat{\mathbf{k}}\cdot\hat{\mathbf{k}}')^2\,\mathrm{Re}\,S_{\lambda\mu}(\mathbf{k}-\mathbf{k}')+\frac{c_{\mathrm{P}}^4}{c_{\mathrm{P}j}^4}\left(1-2\frac{c_{\mathrm{S}}^2}{c_{\mathrm{P}}^2}\right)^2S_\lambda(\mathbf{k}-\mathbf{k}')\Bigg].\end{aligned} \tag{39}$$

The differential scattering operator for $\mathrm{P}j$ to S mode conversion, $j\in\{1,2\}$, is given by

$$[\sigma_{\mathrm{P}j\mathrm{S}}(\mathbf{k},\mathbf{k}')\mathbf{W}_{\mathrm{S}}(\mathbf{k}')]=\sigma_{\mathrm{ps}}(\mathbf{k},\mathbf{k}')\mathbf{G}(\mathbf{k},\mathbf{k}'):\mathbf{W}_{\mathrm{S}}(\mathbf{k}'), \tag{40}$$

where

$$\begin{aligned}\sigma_{\mathrm{ps}}(\mathbf{k},\mathbf{k}')=\frac{\pi}{2}c_{\mathrm{S}}^2[&|\mathbf{k}'|^2(S_\varrho(\mathbf{k}-\mathbf{k}')+S_F(\mathbf{k}-\mathbf{k}')\\&+2\,\mathrm{Re}\,S_{\varrho F}(\mathbf{k}-\mathbf{k}'))+4|\mathbf{k}|^2(\hat{\mathbf{k}}\cdot\hat{\mathbf{k}}')^2S_\mu(\mathbf{k}\\&-\mathbf{k}')+4|\mathbf{k}||\mathbf{k}'|(\hat{\mathbf{k}}\cdot\hat{\mathbf{k}}')(\mathrm{Re}\,S_{\varrho\mu}(\mathbf{k}-\mathbf{k}')\\&+\mathrm{Re}\,S_{F\mu}(\mathbf{k}-\mathbf{k}'))],\end{aligned} \tag{41}$$

and $\mathbf{G}(\mathbf{k},\mathbf{k}')$ is the 2×2 real matrix defined by $G_{rs}(\mathbf{k},\mathbf{k}')\coloneqq(\hat{\mathbf{k}}\cdot\hat{\mathbf{z}}_r(\mathbf{k}'))(\hat{\mathbf{k}}\cdot\hat{\mathbf{z}}_s(\mathbf{k}'))$. Conversely, the differential scattering operator for S to $\mathrm{P}j$ mode conversion, $j\in\{1,2\}$, is

$$[\sigma_{\mathrm{SP}j}(\mathbf{k},\mathbf{k}')\mathrm{w}_{\mathrm{P}j}(\mathbf{k}')]=\sigma_{\mathrm{ps}}(\mathbf{k}',\mathbf{k})\mathbf{G}(\mathbf{k}',\mathbf{k})\mathrm{w}_{\mathrm{P}j}(\mathbf{k}'). \tag{42}$$

Finally, the differential scattering operator for S scattering without mode conversion is given by

$$\begin{aligned}[\sigma_{\mathrm{SS}}(\mathbf{k},\mathbf{k}')\mathbf{W}_{\mathrm{S}}(\mathbf{k}')]=&\sigma_{\mathrm{ss}}^{\mathrm{TT}}(\mathbf{k},\mathbf{k}')\mathbf{T}(\mathbf{k},\mathbf{k}')\mathbf{W}_{\mathrm{S}}(\mathbf{k}')\mathbf{T}(\mathbf{k}',\mathbf{k})\\&+\sigma_{\mathrm{ss}}^{\Gamma\Gamma}(\mathbf{k},\mathbf{k}')\mathbf{\Gamma}(\mathbf{k},\mathbf{k}')\mathbf{W}_{\mathrm{S}}(\mathbf{k}')\mathbf{\Gamma}(\mathbf{k}',\mathbf{k})\\&+\sigma_{\mathrm{ss}}^{\mathrm{T}\Gamma}(\mathbf{k},\mathbf{k}')\mathbf{T}(\mathbf{k},\mathbf{k}')\mathbf{W}_{\mathrm{S}}(\mathbf{k}')\mathbf{\Gamma}(\mathbf{k}',\mathbf{k})\\&+\sigma_{\mathrm{ss}}^{\mathrm{T}\Gamma}(\mathbf{k}',\mathbf{k})\mathbf{\Gamma}(\mathbf{k},\mathbf{k}')\mathbf{W}_{\mathrm{S}}(\mathbf{k}')\mathbf{T}(\mathbf{k}',\mathbf{k}),\end{aligned} \tag{43}$$

where

$$\sigma_{ss}^{TT}(\mathbf{k},\mathbf{k}') = \frac{\pi}{2} c_S^2 |\mathbf{k}|^2 (S_\varrho(\mathbf{k}-\mathbf{k}') + S_F(\mathbf{k}-\mathbf{k}') + 2\,\mathrm{Re}\, S_{\varrho F}(\mathbf{k}-\mathbf{k}')),$$
$$\sigma_{ss}^{\Gamma\Gamma}(\mathbf{k},\mathbf{k}') = \frac{\pi}{2} c_S^2 |\mathbf{k}|^2 S_\mu(\mathbf{k}-\mathbf{k}'), \tag{44}$$
$$\sigma_{ss}^{T\Gamma}(\mathbf{k},\mathbf{k}') = \frac{\pi}{2} c_S^2 |\mathbf{k}|^2 (\mathrm{Re}\, S_{\varrho\mu}(\mathbf{k}-\mathbf{k}') + \mathrm{Re}\, S_{F\mu}(\mathbf{k}-\mathbf{k}')).$$

$\mathbf{T}(\mathbf{k},\mathbf{k}')$ and $\mathbf{\Gamma}(\mathbf{k},\mathbf{k}')$ are the 2×2 real matrices defined by $T_{rs}(\mathbf{k},\mathbf{k}') := \hat{\mathbf{z}}_r(\mathbf{k})\cdot\hat{\mathbf{z}}_s(\mathbf{k}')$ and $\Gamma_{rs}(\mathbf{k},\mathbf{k}') := (\hat{\mathbf{k}}\cdot\hat{\mathbf{k}}')(\hat{\mathbf{z}}_r(\mathbf{k})\cdot\hat{\mathbf{z}}_s(\mathbf{k}')) + (\hat{\mathbf{k}}\cdot\hat{\mathbf{z}}_s(\mathbf{k}'))(\hat{\mathbf{k}}'\cdot\hat{\mathbf{z}}_r(\mathbf{k}))$. These results show that the forward ("+") and backward ("−") waves are uncoupled in the random case.

V. DIFFUSION LIMIT

It is now well established that in the diffusion limit, a heat equation can be recovered from a radiative transfer equation; see for example Ref. 23 for scalar waves or Ref. 5 for polarized waves. This regime is reached after long propagation distances with respect to the scattering mean-free path of the medium, the latter being the mean distance traveled by waves before their direction of propagation is significantly altered. Waves in this regime have lost their initial direction of propagation as well as their initial polarization. From such a diffusion equation, one can also obtain proportionality relationships between the energy densities of particular propagating modes and their associated velocities, independent of the scattering mechanism. The scattering mean-free path is defined for a particular propagating mode α by

$$\ell_\alpha(\mathbf{x},\mathbf{k}) = \frac{|\boldsymbol{\nabla}_{\mathbf{k}}\omega_\alpha|}{\Sigma_\alpha(\mathbf{x},\mathbf{k})}. \tag{45}$$

The derivation of the radiative transfer equations has been carried out so far for high frequencies such that $\Lambda \ll c_\alpha t$ and it remains applicable as long as $\Lambda \ll \ell_\alpha$, where Λ is a typical wavelength. The diffusion limit considers that propagation distances are also much longer than the scattering mean-free path: $c_\alpha t \gg \ell_\alpha$. Parameter fluctuations are also assumed to be statistically isotropic. This implies that the differential scattering cross sections are rotationally invariant inasmuch as they depend on $|\mathbf{k}|$ and $\hat{\mathbf{k}}\cdot\hat{\mathbf{k}}'$ solely. The former may be written under these conditions $\sigma_{\alpha\beta}(\mathbf{k},\mathbf{k}') := c_\beta |\mathbf{k}'|^{-2} \times \tilde{\sigma}_{\alpha\beta}(|\mathbf{k}|,|\mathbf{k}'|,\hat{\mathbf{k}}\cdot\hat{\mathbf{k}}')$ and the total scattering cross sections, depending on $|\mathbf{k}|$ solely, are

$$\Sigma_{\alpha\beta}(|\mathbf{k}|) = \frac{R_\beta}{R_\alpha}\left(\frac{c_\alpha}{c_\beta}\right)^3 \Sigma_{\beta\alpha}\left(\frac{c_\alpha}{c_\beta}|\mathbf{k}|\right) = \int_{S^2} \tilde{\sigma}_{\alpha\beta}\left(|\mathbf{k}|, \frac{c_\alpha}{c_\beta}|\mathbf{k}|, \hat{\mathbf{k}}\cdot\hat{\mathbf{k}}'\right) \mathrm{d}\hat{\mathbf{k}}',$$

where $\mathrm{d}\hat{\mathbf{k}}$ is the elementary solid angle of $\mathbb{R}^3$ and S^2 is its unit sphere.

Based on these assumptions, Eq. (32) yields the following diffusive regime for the present model of porous media. Coherence matrix $\mathbf{W}_S^+$ and specific intensities w_{Pj}^+ become independent of $\hat{\mathbf{k}}$ and have the form

$$\mathbf{W}_S^+(\mathbf{x},\mathbf{k},t) = \psi(\mathbf{x},|\mathbf{k}|,t)\mathbf{I}_2, \tag{46a}$$
$$\mathrm{w}_{Pj}^+(\mathbf{x},\mathbf{k},t) = \psi\left(\mathbf{x}, \frac{c_{Pj}}{c_S}|\mathbf{k}|, t\right), \quad j\in\{1,2\}, \tag{46b}$$

where $\psi(\mathbf{x},|\mathbf{k}|,t)$ is a scalar function satisfying the following diffusion equation:

$$\partial_t \psi(\mathbf{x},|\mathbf{k}|,t) = \boldsymbol{\nabla}_{\mathbf{x}}\cdot[D(|\mathbf{k}|)\boldsymbol{\nabla}_{\mathbf{x}}\psi(\mathbf{x},|\mathbf{k}|,t)]. \tag{47}$$

The diffusion constant D is

$$D(|\mathbf{k}|) = \frac{2D_S(|\mathbf{k}|) + \left(\frac{c_S}{c_{P1}}\right)^3 D_{P1}\left(\frac{c_S}{c_{P1}}|\mathbf{k}|\right) + \left(\frac{c_S}{c_{P2}}\right)^3 D_{P2}\left(\frac{c_S}{c_{P2}}|\mathbf{k}|\right)}{2 + \left(\frac{c_S}{c_{P1}}\right)^3 + \left(\frac{c_S}{c_{P2}}\right)^3}, \tag{48}$$

where the $D_\alpha(|\mathbf{k}|)$'s are partial diffusion coefficients which have the usual expression $D_\alpha = \frac{1}{3} c_\alpha \ell_\alpha^\star$, $\alpha\in\{S,P1,P2\}$. The diffusion constant characterizes the growth speed of the diffusive halo in a disordered medium. The so-called transport mean-free paths $\ell_\alpha^\star$ satisfy

$$\mathbf{T}^{-1}(|\mathbf{k}|)\mathbf{\Lambda}\star(|\mathbf{k}|) = \mathbf{C}, \tag{49}$$

where $\mathbf{\Lambda}\star(|\mathbf{k}|) = (\ell_S^\star(|\mathbf{k}|), \ell_{P1}^\star(c_S/c_{P1}|\mathbf{k}|), \ell_{P2}^\star(c_S/c_{P2}|\mathbf{k}|))$, $\mathbf{C} = (c_S, c_{P1}, c_{P2})$, and

$$\mathbf{T}^{-1}(|\mathbf{k}|) = \begin{bmatrix} \Sigma_S(|\mathbf{k}|) - \Sigma_{SS}^\star(|\mathbf{k}|) & -\Sigma_{SP1}^\star(|\mathbf{k}|) & -\Sigma_{SP2}^\star(|\mathbf{k}|) \\ -\Sigma_{P1S}^\star\left(\frac{c_S}{c_{P1}}|\mathbf{k}|\right) & \Sigma_{P1}\left(\frac{c_S}{c_{P1}}|\mathbf{k}|\right) - \Sigma_{P1P1}^\star\left(\frac{c_S}{c_{P1}}|\mathbf{k}|\right) & -\Sigma_{P1P2}^\star\left(\frac{c_S}{c_{P1}}|\mathbf{k}|\right) \\ -\Sigma_{P2S}^\star\left(\frac{c_S}{c_{P2}}|\mathbf{k}|\right) & -\Sigma_{P2P1}^\star\left(\frac{c_S}{c_{P2}}|\mathbf{k}|\right) & \Sigma_{P2}\left(\frac{c_S}{c_{P2}}|\mathbf{k}|\right) - \Sigma_{P2P2}^\star\left(\frac{c_S}{c_{P2}}|\mathbf{k}|\right) \end{bmatrix},$$

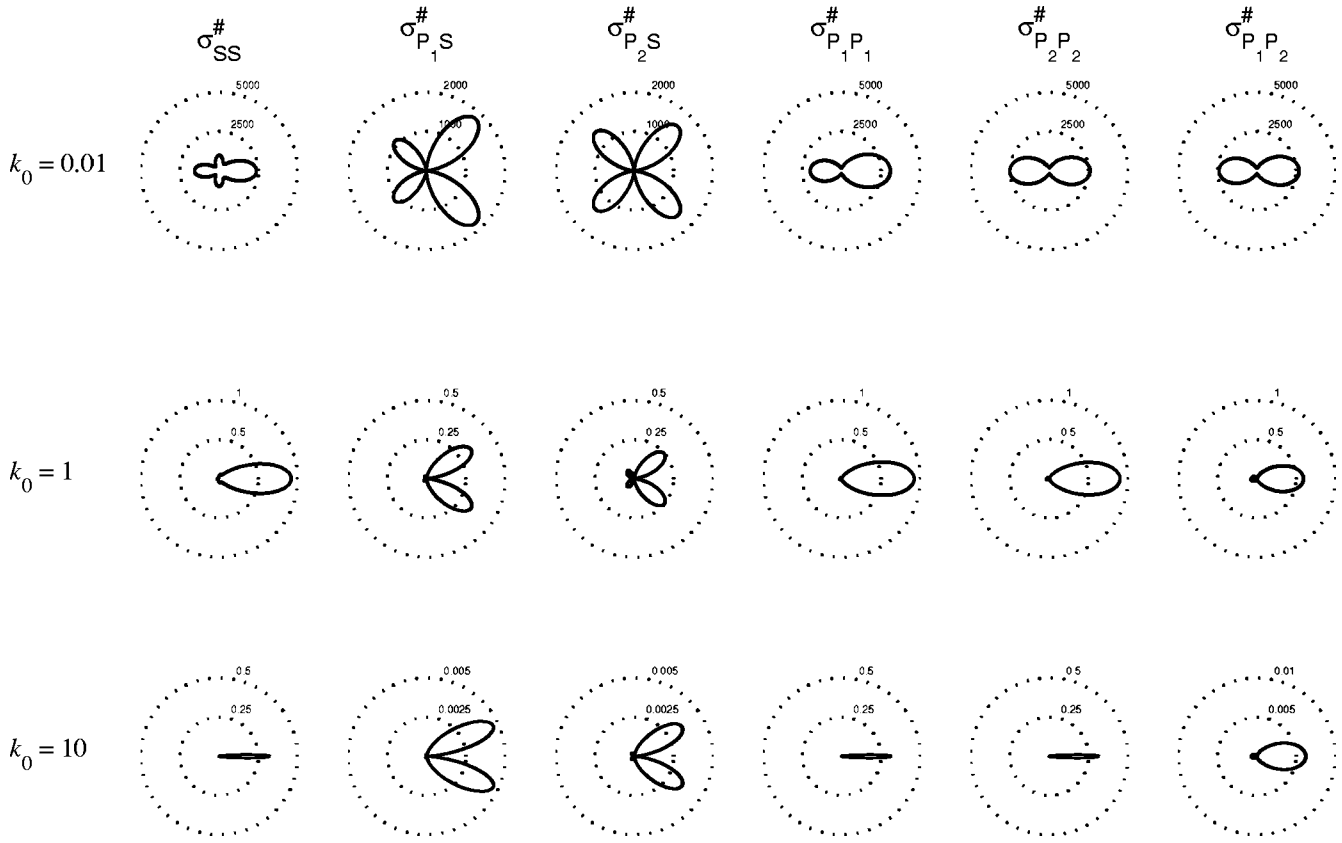


FIG. 1. Normalized scattering cross sections $\sigma^{\#}_{\alpha\beta}$, α, $\beta \in \{S,P_1,P_2\}$, as functions of the angle Θ between $\mathbf{k}$ and $\mathbf{k}'$, for various values of the dimensionless wave number $k_0 = \gamma^{-1}|\mathbf{k}|$, and for a porous medium with Gaussian correlations of perturbations; $h_\varrho = h_F = 0.1$ and $h_\lambda = h_\mu = 1$.

with

$$\Sigma^{\star}_{\alpha\beta}(|\mathbf{k}|) = \frac{R_\beta}{R_\alpha}\left(\frac{c_\alpha}{c_\beta}\right)^3 \Sigma^{\star}_{\beta\alpha}\left(\frac{c_\alpha}{c_\beta}|\mathbf{k}|\right) = \int_{S^2} \tilde{\sigma}_{\alpha\beta}\left(|\mathbf{k}|, \frac{c_\alpha}{c_\beta}|\mathbf{k}|, \hat{\mathbf{k}}\cdot\hat{\mathbf{k}}'\right)\hat{\mathbf{k}}\cdot\hat{\mathbf{k}}'\,\mathrm{d}\hat{\mathbf{k}}',$$

for α, $\beta \in \{S,P1,P2\}$. These coefficients characterize the tendency of total scattering to be in the forward direction. It is worth noting at this stage that analytical expressions of the diffusion parameters like the mean-free paths or the diffusion coefficients are useful because these quantities can be estimated from experimental data while correlation lengths and standard deviations of parameter fluctuations are usually not observable; see for example Refs. 24–28 for such experimental analyses.

At last, the space-time energy densities $\mathcal{E}_S(\mathbf{x},t) = \int_{\mathbb{R}^3} \mathrm{Tr}\,\mathbf{W}^+_S(\mathbf{x},\mathbf{k},t)\mathrm{d}\mathbf{k}$ and $\mathcal{E}_{Pj}(\mathbf{x},t) = \int_{\mathbb{R}^3} w^+_{Pj}(\mathbf{x},\mathbf{k},t)\mathrm{d}\mathbf{k}$, for $j \in \{1,2\}$, satisfy

$$\frac{\mathcal{E}_{P1}(\mathbf{x},t)}{\mathcal{E}_{P2}(\mathbf{x},t)} = \left(\frac{c_{P2}}{c_{P1}}\right)^3, \quad \frac{\mathcal{E}_S(\mathbf{x},t)}{\mathcal{E}_{Pj}(\mathbf{x},t)} = 2\left(\frac{c_{Pj}}{c_S}\right)^3, \tag{50}$$

in the diffusion limit, whatever the underlying (random) heterogeneity of the medium or the source may be. The above equipartition rules are indicators of the relevance of the diffusion approximation: the latter is acceptable as soon as these relations are fulfilled by the energy densities, at each location $\mathbf{x}$ of the medium and at each time t. Furthermore, they show that in this regime, most of the energy in the skeleton is transported by the S modes, no matter how it was distributed initially, because $c_{P2} < c_S < c_{P1}$. This characteristic feature can be said to be ''universal'' because, again, it does not depend on the details of the heterogeneities and/or the source, and it is observed everywhere in the medium at any time independent of any averaging process.

VI. NUMERICAL EXAMPLE

The compressional wave speeds for energy transport are computed by Eq. (24). One can use for example the data provided in Ref. 29 for the experiments carried on in Ref. 30 for a water-saturated porous solid composed of sintered glass spheres: $\varrho_f = 1000\ \mathrm{kg/m^3}$, $\varrho_s = 2480\ \mathrm{kg/m^3}$, $K_f = 2.25\,10^9$ Pa, $K_b = 15.2\,10^9$ Pa, $K_s = 40.7\,10^9$ Pa, $\mu_0 = 10.75\,10^9$ Pa, $\Phi = 0.283$, and $\alpha = 1.82$. One obtains $c_{P1} = 4053$ m/s and $c_{P2} = 1045$ m/s, which compares remarkably well with the experimental estimations $c_{P1} \simeq 4050$ m/s and $c_{P2} \simeq 1040$ m/s.[30] For the diffusion parameters, a statistically isotropic, Gaussian model of correlation of the fluctuations of ϱ, F^{-1}, μ_0^{-1}, and λ_0^{-1} is considered

$$S_{ab}(|\mathbf{k}|) = \frac{h_a h_b r_{ab}}{8\sqrt{\pi^3}\gamma^3}\exp\left(-\frac{|\mathbf{k}|^2}{4\gamma^2}\right), \quad a,b \in \{\varrho,F,\mu,\lambda\}, \tag{51}$$

where γ^{-1} is the correlation length of these parameters. The h_a's, a, $b \in \{\varrho,F,\mu,\lambda\}$, are their fluctuation amplitudes, and the r_{ab}'s are coherency coefficients such that $|r_{ab}| \leqslant 1$ and $r_{aa} = 1$. The normalized scattering cross sections

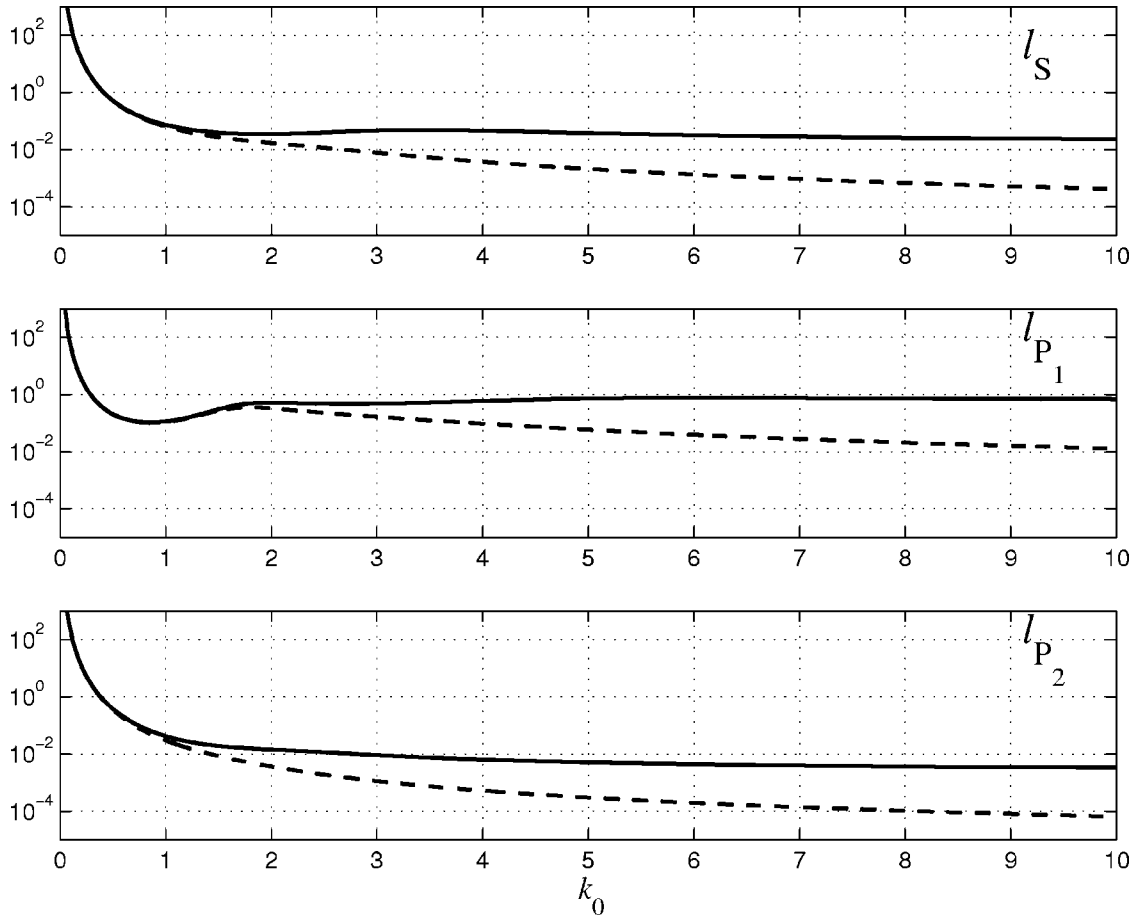


FIG. 2. Normalized scattering (dashed lines) and transport (solid lines) mean-free paths $\gamma\ell_\alpha$ and $\gamma\ell_\alpha^\star$, α=S, P_1 or P_2, as functions of the dimensionless wave number $k_0=\gamma^{-1}|\mathbf{k}|$, for a porous medium with Gaussian correlations of perturbations; $h_\varrho=h_F=0.1$ and $h_\lambda=h_\mu=1$.

$$\sigma^{\#}_{\alpha\beta}(|\mathbf{k}|,\Theta)=\left(\frac{c_\alpha}{c_\beta}\right)^3\frac{\sigma_{\alpha\beta}(\mathbf{k},\mathbf{k}')}{\gamma c_\alpha\Sigma_{\alpha\beta}(|\mathbf{k}|)},\quad \alpha,\beta\in\{\mathrm{S},\mathrm{P}_1,\mathrm{P}_2\},$$

where $\cos\Theta=\hat{\mathbf{k}}\cdot\hat{\mathbf{k}}'$, are plotted in polar coordinates in Fig. 1 as functions of Θ for various values of the dimensionless wave number $k_0:=\gamma^{-1}|\mathbf{k}|$, and for the above data with $h_\varrho=h_F=0.1$, $h_\mu=h_\lambda=1$ (strongly fluctuating inverse Lamé's moduli with respect to the density and inverse formation factor), $r_{\varrho F}=r_{\mu\lambda}=0.8$, and all other coherency coefficients being equal to 0.2. We outline their overall strong anisotropy—even at low frequencies—despite the assumption of statistical isotropy. Also, forward scattering dominates as the frequency (wave number) increases. Figure 2 displays the normalized scattering mean-free paths $\gamma\ell_\alpha$ (dashed lines) and the normalized transport mean-free paths $\gamma\ell_\alpha^*$ (solid lines), α=S, P_1, or P_2, as functions of the dimensionless wave number. Transport mean-free paths get larger than the scattering mean-free paths as frequency increases, as expected from the increased anisotropy of scattering in the forward direction usually observed as frequency increases. One also notes that the former have nonzero, frequency-independent values over a broad frequency range: this is the so-called Rayleigh–Gans regime. Figure 3 displays the normalized diffusion constant $\gamma D/c_S$ as a function of the dimensionless wave number k_0, for the same data and set of fluctuation amplitudes and coherency coefficients. The fact that these diffusion parameters have very low asymptotic values indicates that weak localization may dominate in porous media at the higher frequencies. That is the reason why such materials are used for acoustic insulation, among other possible applications. Finally, the author wishes to point out that choosing different values for the fluctuation amplitudes and coherency coefficients does not modify the diffusion parameters significantly, both from a quantitative and a qualitative point of view.

VII. CONCLUSIONS

In this work the acoustic diffusion properties of a mechanically impacted fluid-saturated poro-visco-elastic medium in the higher frequency range of vibration have been obtained. They are associated with the diffusion equation which must be satisfied by the energy density (in space and time) at late times or after long propagation distances in a randomly perturbed medium. This diffusive regime is also characterized by universal equipartion rules between the modes of propagation, S shear waves of the bulk solid phase, and P1 and P2 compressional waves of the solid and fluid phases. In the course of the derivation one has also obtained analytical expressions of the associated energy speeds and observed that they are nondispersive in the higher frequency range of vibration.

One important issue in the proposed theory remains unsolved: what are the boundary conditions for the energy den-

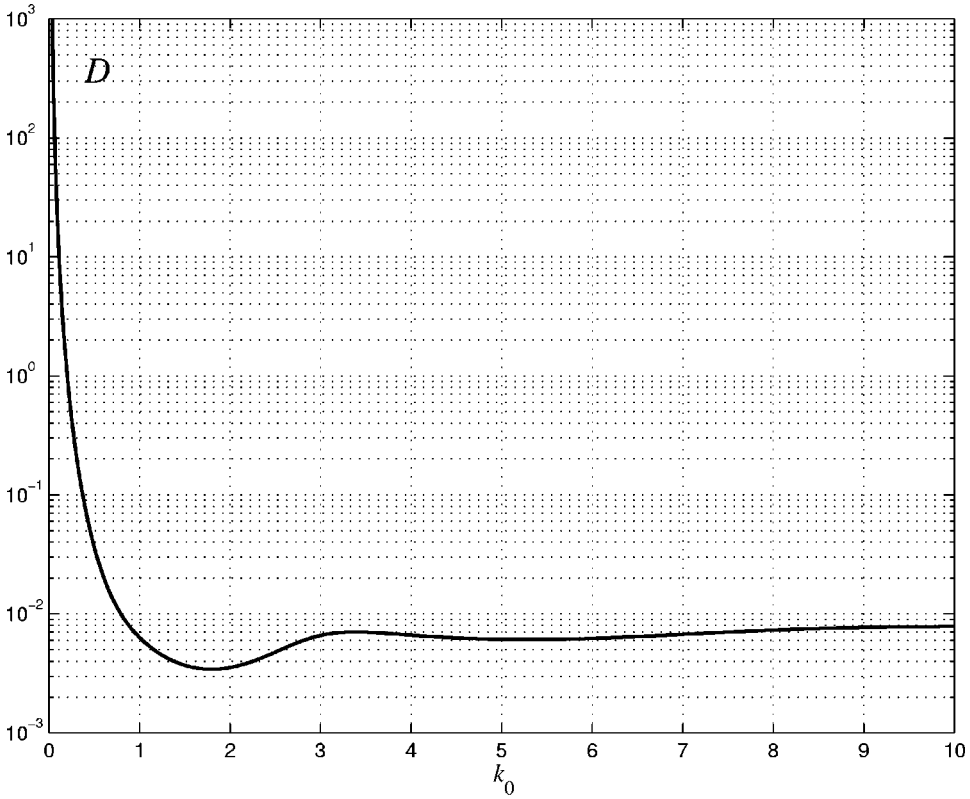


FIG. 3. Normalized diffusion coefficient $\gamma D/c_S$ as a function of the dimensionless wave number $k_0 = \gamma^{-1}|\mathbf{k}|$, for a porous medium with Gaussian correlations of perturbations; $h_\varrho = h_F = 0.1$ and $h_\lambda = h_\mu = 1$.

sity with respect to the boundary conditions imposed to the original displacement field of the medium? Further theoretical work on this approach shall be devoted to this aspect of the model. Also, some work shall be done to possibly link the above results to the acoustic insulation properties of porous materials, such as plastic or metallic foams. One has not explored that possibility yet, though it may have some important applications in the aerospace or automotive industry.

[1] S. Chandrasekhar, *Radiative Transfer* (Dover, New York, 1960).

[2] U. Frisch, "Wave propagation in random media," in *Probabilistic Methods in Applied Mechanics*, edited by A. T. Bharucha-Reid (Academic, London, 1968), Vol. 1, pp. 75–198.

[3] P.-L. Lions and T. Paul, "Sur les mesures de Wigner," Rev. Mat. Iberoam. **9**(3), 553–618 (1993).

[4] P. Gérard, P. A. Markowich, N. J. Mauser, and F. Poupaud, "Homogenization limits and Wigner transforms," Commun. Pure Appl. Math. **L**(4), 323–379 (1997).

[5] G. C. Papanicolaou and L. V. Ryzhik, "Waves and transport," in *Hyperbolic Equations and Frequency Interactions*, edited by L. Caffarelli and Weinan E (American Mathematical Society, IAS/Park City Mathematics Series, Providence, MA, 1999), Vol. 5, pp. 305–382.

[6] M. Guo and X.-P. Wang, "Transport equations for a general class of evolution equations with random perturbations," J. Math. Phys. **40**(10), 4828–4858 (1999).

[7] J.-L. Akian, "Wigner measures for high-frequency energy propagation in visco-elastic materials" (Tech. Report RT 2/07950 DDSS, ONERA, Châtillon, 2003).

[8] J. G. Berryman, "Scattering by a spherical inhomogeneity in a fluid-saturated porous medium," J. Math. Phys. **26**(6), 1408–1419 (1985).

[9] B. Gurevich, A. P. Sadovnichaja, S. L. Lopatnikov, and S. A. Shapiro, "The Born approximation in the problem of elastic wave scattering by a spherical inhomogeneity in a fluid-saturated porous medium," Appl. Phys. Lett. **61**(11), 1275–1277 (1992).

[10] A. N. Norris, "Radiation from a point source and scattering theory in a fluid-saturated porous solid," J. Acoust. Soc. Am. **77**(6), 2012–2023 (1985).

[11] E. Smith, "Renormalization of quenched randomness in Biot theory," Phys. Rev. B **64**(13), 134202 (2001).

[12] A. Hanyga, "Asymptotic ray theory and scattering in Biot's theory of porous media," in *Computational Mechanics: New Trends and Applications, Proceedings of the 4th World Congress on Computational Mechanics, Buenos Aires, 29 June–2 July, 1998*, edited by S. Idelsohn, E. Oñate, and E. Dvorkin (CIMNE, Barcelona, 1998).

[13] E. Savin, "Transient transport equations for high-frequency power flow in heterogeneous cylindrical shells," Waves Random Media **14**(3), 303–325 (2004).

[14] N. Antonić, "H-measure applied to symmetric systems," Proc. - R. Soc. Edinburgh, Sect. A: Math. **126**(6), 1133–1155 (1996).

[15] 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**(2), 168–191 (1956).

[16] O. Coussy, *Mechanics of Porous Media* (in French) (Éditions Technip, Paris, 1991).

[17] Y. C. Fung, *Foundations of Solid Mechanics* (Prentice Hall, Englewood Cliffs, NJ, 1965).

[18] Z. E. A. Fellah and C. Depollier, "Transient acoustic wave propagation in rigid porous media: A time-domain approach," J. Acoust. Soc. Am. **107**(2), 683–688 (2000).

[19] M. A. Biot, "General theory of acoustic propagation in porous dissipative media," J. Acoust. Soc. Am. **34**(5), 1254–1264 (1962).

[20] J. G. Berryman, "Confirmation of Biot's theory," Appl. Phys. Lett. **37**(4), 382–384 (1980).

[21] N. C. Dutta, "Theoretical analysis of observed second bulk compressional wave in a fluid-saturated porous solid at ultrasonic frequencies," Appl. Phys. Lett. **37**(10), 898–900 (1980).

[22] Z. E. A. Fellah, S. Berger, W. Lauriks, C. Depollier, C. Aristégui, and J.-Y. Chapelon, "Measuring the porosity and the tortuosity of porous materials via reflected waves at oblique incidence," J. Acoust. Soc. Am. **113**(5), 2424–2433 (2003).

[23] K. M. Case and P. F. Zweifel, *Linear Transport Theory* (Addison-Wesley, Reading, MA, 1967).

[24] H. P. Schriemer, M. L. Cowan, J. H. Page, P. Sheng, Z. Liu, and D. A. Weitz, "Energy velocity of diffusing waves in strongly scattering media," Phys. Rev. Lett. **79**(17), 3166–3169 (1997).

[25] R. L. Weaver, "Ultrasonics in an aluminium foam," Ultrasonics **36**(1–5), 435–442 (1998).

[26] A. Tourin, A. Derode, A. Peyre, and M. Fink, "Transport parameters for an ultrasonic pulsed wave propagating in a multiple scattering medium," J. Acoust. Soc. Am. **108**(2), 503–512 (2000).

[27] P. Anugonda, J. S. Wiehn, and J. A. Turner, ''Diffusion of ultrasound in concrete,'' Ultrasonics **39**(6), 429–435 (2001).
[28] J. A. Scales and A. E. Malcolm, ''Laser characterization of ultrasonic wave propagation in random media,'' Phys. Rev. E **67**(4), 046618 (2003).
[29] B. Gurevich, O. Kelder, and D. M. J. Smeulders, ''Validation of the slow compressional wave in porous media: Comparison of experiments and numerical simulations,'' Transp. Porous Media **36**(2), 149–160 (1999).
[30] T. J. Plona, ''Observation of a second bulk compressional wave in a porous medium at ultrasonic frequencies,'' Appl. Phys. Lett. **36**(4), 259–261 (1980).

Distribution of wave packet sizes in microbarom wave trains observed in Alaska[a)]

John V. Olson[b)] and Curt A. L. Szuberla
Geophysical Institute, University of Alaska, 903 Koyukuk Drive, Fairbanks, Alaska 99775-7320

(Received 25 April 2003; revised 18 November 2004; accepted 5 December 2004)

This work reports on a study of the distribution of wave packet sizes contained in intervals of continuous microbarom activity. Microbaroms are a class of atmospheric infrasound that is characterized by narrow-band, nearly sinusoidal, waveforms with periods near 5 s. They are known to be generated by marine storms, presumably through a nonlinear interaction of surface waves, however the detailed analysis of the process is still incomplete. The data analyzed were obtained using the University of Alaska infrasound array of four microphones located in central Alaska. Because of the narrow-band feature of the microbarom signals, the Hilbert transform is applicable as a method for finding phase breaks in the signal. The phase breaks are interpreted as the demarcation of the boundaries of wave packets. When applied to long sequences of microbaroms a broad distribution of packet lengths is found that diminishes monotonically with length and has a mean near 10 cycles and a variance nearly as large. The distribution function decreases exponentially with packet length. The distribution of packet sizes is influenced by the presence of multiple sources and multiple propagation paths between the sources and the sensor array. Identification of individual packets should open the way to a more detailed analysis of microbarom wave trains. After separating the wave train into individual wavelets the intermicrophone correlation is estimated as a function of microphone separation. As has been observed in earlier microbarom studies, a decrease in correlation was observed for microphone pairs orthogonal to the direction of propagation when compared to correlations between microphones spaced along the direction of wave propagation. [DOI: 10.1121/1.1854651]

PACS numbers: 43.28.Dm [LCS] Pages: 1032–1037

I. INTRODUCTION

Microbaroms are a class of atmospheric acoustic waves that are associated with marine storms. They are observed worldwide with periods in the 3–8-s range and amplitudes of a few tenths of a Pascal. Benioff and Gutenberg (1939) first reported microbaroms and their similarity to microseisms and discussed their association with low pressure areas over the ocean. Saxer (1945) showed the first relationship of microbaroms with wave height in ocean storms and microbarom amplitudes. Details of this early work can be found in Posmentier (1967) and Gossard and Hooke (1975). Microbaroms appear to be generated by the nonlinear interaction of standing ocean waves with the atmosphere. Seismic waves called microseisms are also radiated by standing ocean waves interacting with the sea bottom. Donn and Naini (1973) and Rind (1980) have shown that often microbaroms and microseisms often share a common source. Theoretical studies by Daniels (1952, 1953), Posmentier (1967), Bretherton (1969), and recently by Arendt and Fritts (2000) have investigated the coupling of between ocean waves and atmospheric waves. Studies such as these have shown that the coupling can produce propagating atmospheric waves only when nonlinear terms are included in the formulation.

It is also known from the study of arrival directions that the microbarom source region is distributed in space. Studies of the variations of signal coherence across an array of sensors have shown that the ''beamwidth'' of signals can be as large as 0.2 rad (Gossard and Sailors, 1970). The intersensor coherence can also be affected by the arrival of waves from different elevation angles due to refraction in the atmosphere. This leads to the variation in estimated horizontal trace velocities of the waves and to what Gossard and Hooke (1975) term the ''velocity bandwidth.'' The beam width can be shown to cause decorrelation of the microbarom signals transverse to the direction of propagation while the velocity bandwidth causes decorrelation along the direction of propagation.

In our study we treat each microbarom time series that is observed at a single sensor as an independent realization of the microbarom wave train. We employ the Hilbert transform as an approach to estimating the spatial coherence of the microbarom wave packets. The Hilbert transform of a single time series returns what is termed the ''analytic signal.'' This is a complex function whose amplitude is the envelope of the waveform and whose phase is proportional to the instantaneous frequency. By tracking the instantaneous frequency derived from the Hilbert transform we can detect sudden changes in frequency that are associated with sudden changes in the phase of the original signal. Since phase coherence is expected across a wave packet we interpret these changes in phase to mark the boundaries of the individual wave packets in the wave train.

[a)]Portions of this work were presented in ''Studies of microbaroms using multiple infrasound arrays,'' at the 23rd Annual DTRA/NNSA Seismic Research Review, Jackson Hole, WY, 1–5 October, 2001.

[b)]Electronic mail: jvo@gi.alaska.edu

TABLE I. Locations of the UAF array relative to the central element.

Site	North (km)	East (km)
APEX	0.0	0.0
BALL	0.2587	0.4978
CIGO	0.3486	−0.6468
DEER	−0.9314	0.1339

Separations
APEX-BALL: 0.5687 km
APEX-CIGO: 0.7297
APEX-DEER: 0.9548
BALL-CIGO: 1.1493
BALL-DEER: 1.2652
CIGO-DEER: 1.5059

II. INSTRUMENTATION AND ANALYSIS

The Geophysical Institute of the University of Alaska, through the support of the Department of Energy, installed a four-element infrasound array in a wooded area on the Fairbanks campus in 1998. The array is comprised of four Chaparral model 4 microphones that have four ''soaker'' hoses attached to serve as a wind-noise filter. The sensor array is in the shape of a triangle with a central element. The relative sensor locations are given in Table I and a map showing their locations on the University of Alaska, Fairbanks Campus is shown in Fig. 1. This array geometry includes intersensor distances ranging from 0.561 to 1.449 km. Pressure variations measured at each sensor are collected at 100 samples/second and time-stamped using a GPS clock. For analysis of low-frequency variations, the data are low-pass filtered and resampled at 10 samples/second.

In the analysis below we derive our estimates of azimuth and trace velocity of the signal from a least-squares fit of a plane wave arrival to the sequence of time delays estimated using cross-correlations between sensors. This is a standard technique described, for example, by Claerbout (1976). Once the delay times are found a least-squares fit to the arrival of a plane wave can be made. For N sensors located at $\vec{x}_i$, $i=1,...,N$ a plane wave from azimuth Θ traveling with horizontal velocity V will arrive at the individual sensors according to $\tau = Xf$ where τ is a vector of time delays, X is the location matrix of the sensors, and f is the ''slowness vector:''

$$\tau = \begin{pmatrix} \tau_1 \\ \tau_2 \\ \vdots \\ \tau_N \end{pmatrix}, \quad X = \begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_N & y_N \end{pmatrix}, \quad f = \begin{pmatrix} \frac{1}{V}\sin\Theta \\ \frac{1}{V}\cos\Theta \end{pmatrix}. \quad (1)$$

When the measured values of τ_i are used the least squares estimator for f is

$$\hat{f} = (X^T X)^{-1} X^T \tau \quad (2)$$

and from $\hat{f}$ we find the estimates for velocity and azimuth of arrival as $\hat{V} = \sqrt{f_1^2 + f_2^2}$ and $\hat{\Theta} = \tan^{-1}(f_1/f_2)$. The uncertainties in the estimates are obtained in a straightforward way (see Szuberla and Olson, 2004).

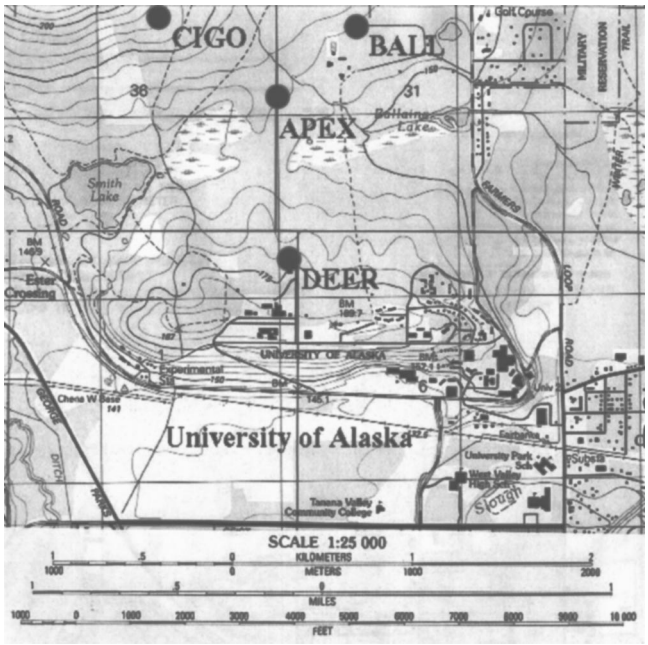


FIG. 1. This figure shows a map of the UAF campus with the locations of the four infrasound sensors. The locations of the sensors and their separations are given in Table I.

III. OBSERVATIONS

Microbarom signals are commonly observed during the winter months in Alaska because of the presence of strong storms moving across the Gulf of Alaska. During the intervals when microbaroms are present in the data they can persist for up to tens of hours and can reach amplitudes of a few tenths of a Pascal. Figure 2 shows an example of microbarom waveforms from each of the four microphones in the University of Alaska Fairbanks (UAF) array during a 360-s interval near 0500 UT on 6 April 2001. The data here are taken from the 10 sample/second records and band-pass filtered between 0.1 and 1.0 Hz using a fifth-order Butterworth band-pass filter. The primary oscillations lie at a frequency just below 0.2 Hz and show large amplitude modulations. We interpret these modulations as the result of the superposition of many wave-packets propagating across the array.

The data from the four microphones shown in the top four panels of Fig. 2 are replotted as a phase-aligned overlay in the bottom panel. In this analysis the cross-correlation of each pair of data sequences was computed and searched for a maximum. When the lag distance between maxima is converted to units of time we obtain a set of six interstation time lags that represent the passage of the signal across the array. The data in the bottom panel are plotted using these time offsets as an indication of the quality of the estimation of the phase delays. The estimated velocity and azimuth, along with the various intersensor coherences, are shown at the top of Fig. 2. In this instance the azimuth of arrival, 263° clockwise from North, is found to point to the center of a large, low-pressure center of a storm off the coast of Alaska. Of course, this direction represents an average over many wave packets.

In order to remove the incoherent portion of the signal and leave only the coherent microbarom signal we have

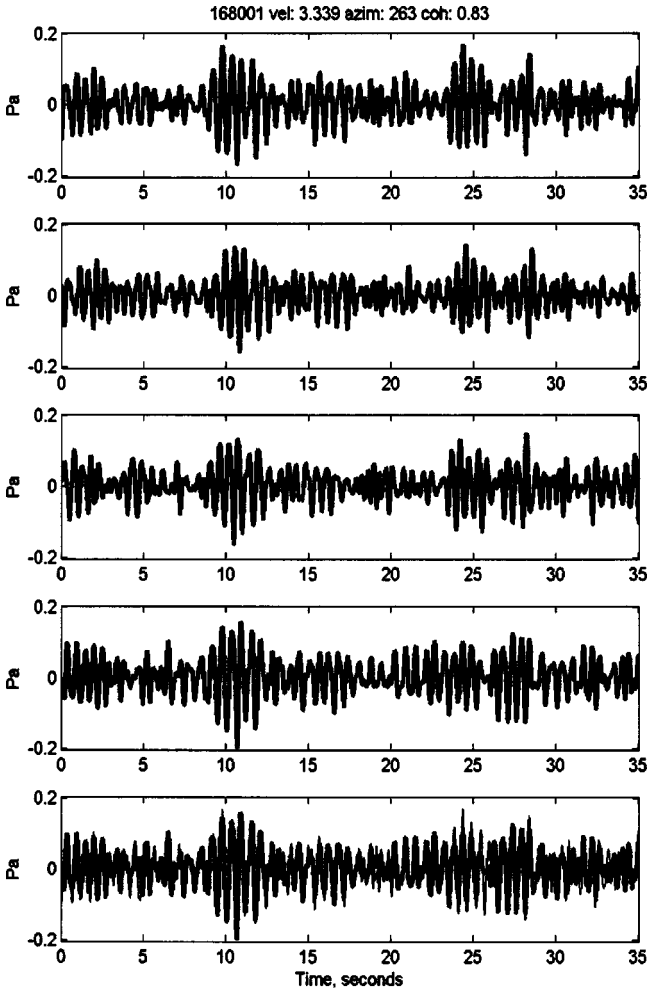


FIG. 2. A sequence of data from the four University of Alaska infrasound sensors showing a strong microbarom wave train. The analysis shows the wavelets are arriving from an average azimuth of 263° clockwise from North at a near-acoustic speed of 0.334 km/s. The data were band-pass filtered between 0.1 and 1.0 Hz. The bottom panel shows the time-shifted overlay of the data. See the text for details.

found it useful to submit the band-pass-filtered data to the pure-state filter. This filter, described in detail by Olson (1982), is designed to estimate the incoherent signal contribution to each frequency estimate in a spectrum and to reduce the contribution of that spectral element accordingly. In effect, the filter selects for signals that are coherent across the sensor array. As an example, Fig. 3 shows the power spectrum for the set of band-pass filtered microbaroms from Fig. 2 before and after pure-state filtering. The band-pass filter is a fast Fourier transform (FFT) based filter with passband from 0.1 to 1.0 Hz. It is clear that the broadband noise has been reduced by several orders of magnitude while the central microbarom peak is preserved.

As an indication of the persistence of the microbarom signals, Fig. 4 shows the results of the least-squares analysis when applied to 24 h of data. The band-pass-filtered data are divided into 100-s intervals and the least-squares estimate of velocity, azimuth, and coherences is found for each interval and plotted. In Fig. 4 the mean value of the intersensor correlation is plotted in the bottom panel. Note that when the coherence is high the least-squares estimator produces consistent results and as the coherence diminishes towards the end of the day the quality of the estimates diminishes although the signal can still be tracked. As mentioned in the description of Fig. 2 the arrival azimuth is approximately 263° and holds steady throughout the event.

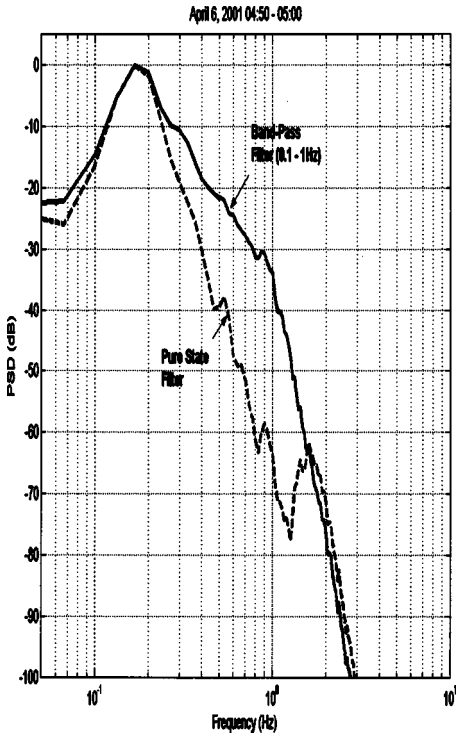


FIG. 3. This figure shows the power spectral density of the data from the APEX sensor shown in Fig. 1 before and after application of the pure-state filter. The filter reduces the amplitude of the incoherent portions of the signal and results in a drop of over 20 dB in the portions of the spectrum away from the central microbarom peak.

Notice that the apparent speed of the phase fronts across the array changes abruptly near 10 and 16 UT. Changes in apparent speed of arriving signals can reflect changes in conditions along the propagation path of the waves. For example, a shift in the wind direction at the altitude that the waves are reflected as they propagate can change the apparent arrival speed and direction. However, the changes observed in this data set are curious since they occur very abruptly, within intervals of a few minutes. We have inspected the data near these changes and find no obvious change in the character of the data themselves. Close inspection shows that the jump in velocity is nearly 70 m/s going from an average near 330 m/s to approximately 400 m/s withing a span of a few minutes. Velocity shifts of this magnitude are much larger than the resolution of the array and the effects that are produced when analyzing digital data. The variance associated with velocity estimates using our array is approximately 6° (Szuberla and Olson, 2004) and the effects of digitization are much smaller. Although we do not understand the cause of the shifts observed, we believe they are a real aspect of the data. Similar shifts have been observed by Wilson (1999a, b) in his analysis of the historic microbarom data sets from Alaska and Antarctica.

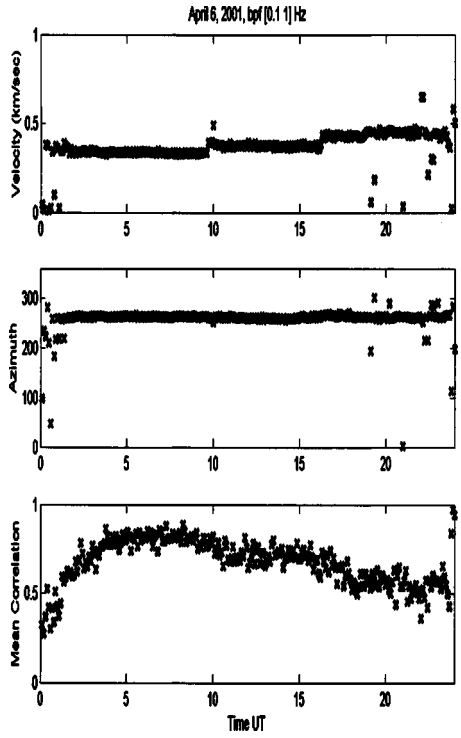


FIG. 4. This plot shows the estimates of velocity, azimuth and mean intersensor correlation for the data taken from the University of Alaska infrasound array on 6 April 2001. The data were band-pass filtered between 0.1 and 1.0 Hz. These estimates were made using a least-squares fit to the time lags observed as the waves propagated across the array.

IV. THE HILBERT TRANSFORM AND MICROBAROM WAVE PACKETS

The Hilbert transform is a useful tool in extracting amplitude and frequency information from real time series and is discussed in many standard texts (e.g., Bracewell, 1986). Given a real sequence $\vec{x}(t)$, it can be combined with its Hilbert transform $\vec{x}_H(t)$ to produce the *analytic signal*: $\vec{z}(t) = \vec{x}(t) + i\vec{x}_H(t)$. The magnitude of $\vec{z}(t)$ is the instantaneous amplitude, or envelope, of the analytic signal and the phase of the analytic signal can be shown to be proportional to the instantaneous frequency. In fact, for a narrow-band signal, it can be shown that the instantaneous frequency is given by $\omega = \mathrm{Im}[d \ln \vec{z}/dt] = \mathrm{Im}[(1/\vec{z}) d\vec{z}/dt]$, where Im is the imaginary part.

The typical microbarom wave train is band-limited with a bandwidth that is smaller than the center frequency and thus is a good candidate for analysis using the Hilbert transform techniques. We make the assumption that the amplitude variations in the microbarom wave train are due to the superposition of individual wave packets and that the individual wave packets can be separated by identifying the locations of the phase breaks between packets. That is, we consider each wave packet to be phase coherent; then the boundary between neighboring packets can be identified by the sudden change in phase that occurs in moving from one packet to the next. As described above, the change in phase corresponds to a change in the instantaneous frequency calculated from the analytic signal developed from the microbarom wave train.

Once the analytic signal is produced and the instantaneous frequency calculated for the entire length of the wave train we use the following construction to establish a threshold for identification of a packet boundary. Using the sequence of instantaneous frequencies we form an estimate of the probability distribution function (pdf) of the estimated frequencies and, from that pdf, we establish thresholds at the 95% confidence level for deviations in frequency that will be considered significant. From the total sequence of frequencies we can then identify which frequencies exceed the threshold and identify their temporal location as the location of a packet boundary. An example is given in Fig. 5.

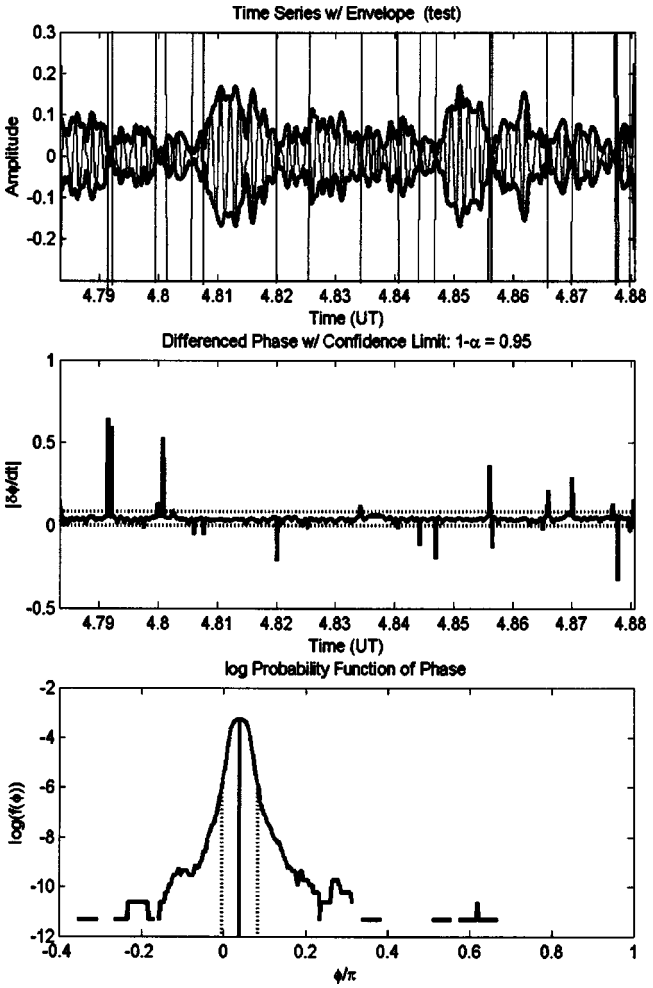


FIG. 5. This figure shows the result of the Hilbert transform technique for identifying phase breaks between packets. The original data are shown the top panel along with the envelope of the analytic function and vertical lines showing locations of phase breaks. The second panel shows the estimates of the instantaneous frequency along with the 95% confidence limits. The bottom panels show the distribution of frequencies from which the confidence limits were derived.

In Fig. 5 we show three panels. The top panel shows the microbarom signal derived from a single sensor. The instantaneous envelope of the analytic signal is also shown. The vertical lines, whose locations are determined in the lower panels, identify possible packet boundaries. The second panel shows the magnitude of the instantaneous frequency of the analytic signal. The dashed lines are the threshold levels determined from the pdf in the bottom panel. Using the collection of instantaneous frequencies shown in the middle panel the pdf of the instantaneous frequencies is calculated

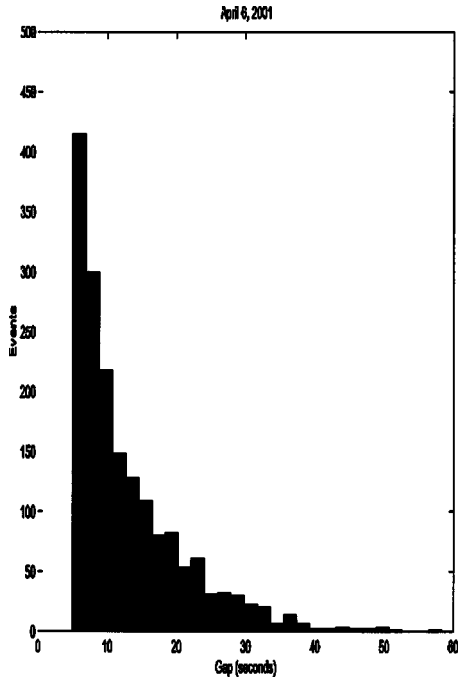


FIG. 6. This figure shows a histogram of the frequency of occurrence of the time intervals between phase breaks in the data taken from the APEX sensor during the interval 0000 UT to 0800 UT, 6 April 2001. Only time intervals above 5 s (approximately 1 wave period) were included in this distribution.

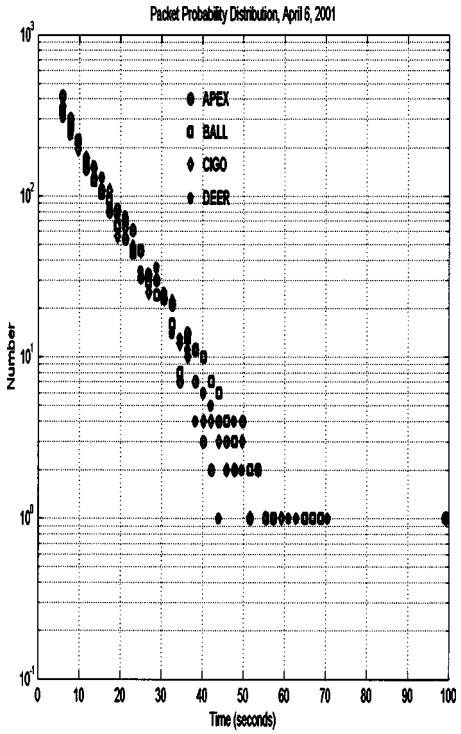


FIG. 7. This figure shows the probability distribution function values replotted on a semilog scale. Data from all four sensors are shown. These data represent averages over 8 h of microbarom signals from 0000 to 0800 UT on 6 April 2001. This shows that the pdf is characterized by a power law with exponent approximately -1.5.

and displayed in the lower panel. Once the pdf is formed, confidence limits are selected and shown on the plot as vertical dashed lines. Those values are then transferred to the middle panel to show where the instantaneous frequencies exceed the threshold. At those locations where the threshold is exceeded the time is noted and transferred back to the top panel where vertical lines are drawn as indicators of packet boundaries. In our analysis of various microbarom signals we have found that setting the threshold levels to bound 95% of the pdf for the instantaneous frequency leads to packet boundaries that are also consistent across the four sensor signals. This is consistent with our assumption that the wave train does not change its form as it propagates across the sensor array.

Once we have determined the temporal locations of packet boundaries we procede to characterize the interval spanned between each pair of boundaries as an indication of packet length. In our analysis we found it useful to construct the distribution using data intervals 5 to 10 min in length. This usually insured that a visual inspection was possible to verify the accuracy of the algorithm. Excluding intervals that are less than one wave period we can obtain a distribution of packet lengths from each of our sensors. The result for 8 h of data on 6 April 2001 is shown in Fig. 6. There we have constructed an estimate of the probability distribution function of packet lengths for each sensor. The distribution functions shown in Fig. 6 were constructed from analyses spanning 1-h intervals. The distribution functions did not change significantly from hour to hour throughout the day. In fact, the distribution function maintained the same shape regardless of the size of the data interval for intervals between 1 and 8 h.

Distributions of packet sizes can be obtained directly from the distribution of packet lengths at a constant frequency for acoustic waves by a simple numerical scaling. The monotonic behavior of the distribution function is readily observed when the data are plotted on a semilogarithmic scale. Figure 7 shows the data from Fig. 6 plotted on a semi-logarithmic scale. The linear decrease in the pdf values with packet size indicates an exponential decrease of the form $P \simeq \exp(-T/T_P)$ where T_p is the average packet length. These data represent averages over 8 h of microbarom signals from 0000–0800 UT on 6 April 2001. This shows that the pdf is characterized by a power law with exponent approximately -1.5. While we have not carried out an exhaustive study of all microbarom intervals observed by our array we have found this relationship to hold for the eight microbarom intervals that showed the largest signal-to-noise levels during the winter of 2000/2001. The values of T_P vary from event to event ranging from approximately 8 s to 15 s with variances of the same order.

V. APPLICATION OF PACKET ANALYSIS TO MICROBAROM COHERENCE

Once the demarcation between packets has been identified it is possible to investigate the intermicrophone correlation of each packet and for the ensemble of packets that comprise the wave train. In Fig. 8 we show the results of the correlation estimates for each microphone pair over 216 packets taken from an interval of the 6 April 2001 data that had a large signal-to-noise ratio. We have plotted the correlation as a function of the intermicrophone separation and

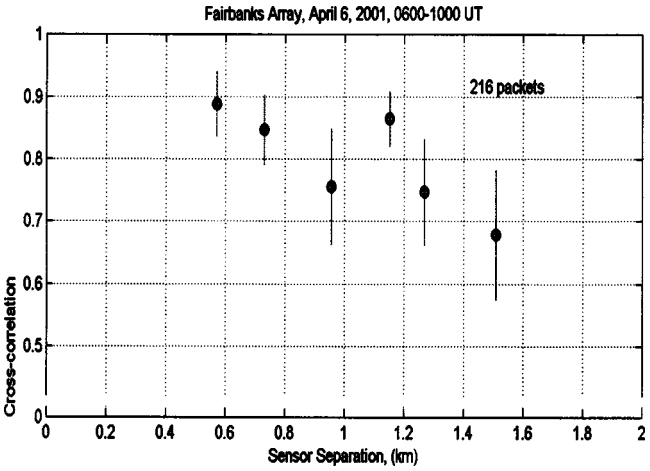


FIG. 8. This figure shows the intermicrophone correlation averaged over 216 wave packets plotted as a function of the separation between the microphone pairs. Note the correlations that involve microphone 4 show lower correlations than the other pairs. This is interpreted as an indication of the decrease in correlation transverse to the direction of wave propagation.

see that there is a general decrease in correlation with increasing microphone distance. However, the decrease is not monotonic. In fact, it is seen that the correlations are all low for pairs that include microphone 4. In the array used here, microphone 4 was located in the southeast quadrant and the wave propagation vector pointed slightly north of east. Thus, each microphone pair that included microphone 4 would have separation vectors nearly perpendicular to the direction of the wave front. The resulting decrease in the correlations for these pairs can be understood in terms of the transverse correlation length that has been described by Postmentier (1967) and Gossard and Sailors (1970).

VI. DISCUSSION

With the view that the modulated waveform that is characteristic of microbarom wave trains is produced by the superposition of coherent wave packets, we have sought to characterize the distribution of packet sizes that are observed in our array in central Alaska. We have analyzed microbaroms from several intervals with a strong microbarom signal-to-noise level from the winter of 2000/2001. Since the microbaroms represent narrow-band fluctuations it is possible to identify boundaries between coherent wave packets using a Hilbert transform technique. Our analysis shows that the distribution of packet lengths is usually limited with very few packets longer than about ten oscillations. The mean of the distributions was found to be consistently near 10 s, representing only two oscillations. The variance in the distribution was large, usually of the order of 5 to 10 s.

Of course, the mean packet length is an indication of the bandwidth of interfering signals. Since our mean $T_P \simeq 10\,\mathrm{s}$, or two wavelengths, we must be dealing with a wave spectrum that has an average bandwidth of approximately $2\pi/T_P \simeq 0.6\,\mathrm{Hz}$ indicating that $\delta f \simeq f$ for the microbarom sequences studied here. Arendt and Fritts (2000) developed a generation mechanism for microbaroms based upon a nonlinear interaction of wind and ocean waves and showed that for a given spectrum of ocean waves the acoustic spectrum generated would be shifted to higher wave numbers. It is well known that the microbarom spectrum is relatively narrow band centered at periods near 5 s. It is difficult to imagine that all marine storms generate surface waves with nearly the same spectrum so as to produce microbaroms near the same period.

The distribution of packet sizes may depend upon the distance of the source, the presence of multiple sources, propagation paths for each of the wave trains, etc. Certainly, as one adds more microbarom sources, each with its own distinctive spectrum, the phase structure of the resulting wave train becomes more complex. Source distance and path variations can affect the amplitude of waves that are superimposed to form the observed wave train and influence the apparent coherence of packets also. We plan to begin to use the technique developed here to isolate individual wave packets in order to study their properties. Through this we may be able to begin to identify multiple sources and/or multiple propagation paths.

ACKNOWLEDGMENTS

UAF infrasound array was constructed through Contract No. DE-FC04-98AL79802 from the Department of Energy. The authors wish to express their appreciation to Dr. C. Wilson for his guidance and advice in preparing this manuscript. In addition we would like to acknowledge the referees and the editor for comments that helped improve the manuscript.

Arendt, S., and Fritts, D. C. (**2000**). ''Acoustic radiation by ocean surface waves,'' J. Fluid Mech. **415**, 1–21.
Benioff, H., and Gutenberg, B. (**1939**). ''Waves and currents recorded by electromagnetic barographs,'' Bull. Am. Meteorol. Soc. **20**, 421–429.
Bracewell, R. N. (**1986**). *The Fourier Transform and Its Applications*, 2nd ed. (McGraw-Hill, New York).
Bretherton, F. (**1969**). ''Waves and turbulence in stably stratified fluids,'' Radio Sci. **4**, 1279–1287.
Claerbout, F. (**1976**). *Fundamentals of Geophysical Data Processing* (McGraw-Hill, New York).
Daniels, F. (**1952**). ''Acoustical energy radiated by ocean waves,'' J. Acoust. Soc. Am. **24**, 83.
Daniels, F. (**1953**). ''The mechanism of generation of infrasound by ocean waves,'' J. Acoust. Soc. Am. **25**, 796.
Donn, W., and Naini, B. (**1973**). ''Sea wave origin of microbaroms and microseisms,'' J. Geophys. Res. **78**, 4482–4488.
Gossard, E., and Hooke, W. (**1975**). *Waves in the Atmosphere* (Elsevier, Amsterdam).
Gossard, E., and Sailors, D. (**1970**). ''Dispersion bandwidth deduced from coherency of wave recordings from spatially separated sites,'' J. Geophys. Res. **75**, 1324–1329.
Olson, J. (**1982**). ''Noise suppression using data-adaptive polarization filters: Applications to infrasonic array data,'' J. Acoust. Soc. Am. **72**, 1456–1460.
Posmentier, E. (**1967**). ''A theory of microbaroms,'' Geophys. J. R. Astron. Soc. **13**, 487–501.
Rind, D. (**1980**). ''Mocroseisms and palisades 3. Microseisms and microbaroms,'' J. Geophys. Res. **85**, 4854–4862.
Saxer, L. (**1945**). ''Elektrische messung kleiner atmospharischer druckschwankungen,'' Helv. Phys. Acta **18**, 527.
Szuberla, C. A. L., and Olson, J. V. (**2004**). ''Uncertainties associated with parameter estimation in atmospheric infrasound arrays,'' J. Acoust. Soc. Am. **115**, 253–258.
Wilson, C. R. (**1999a**). ''Microbarom infrasound at Fairbanks, Alaska 1984, Identification of coherent infrasonic signals in an historical data set,'' Report No. Seven, Defense Threat Reduction Agency.
Wilson, C. R. (**1999b**). ''Identification of coherent infrasonic signals in an historical data set,'' Report No. 9, Defense Threat Reduction Agency.

Analysis of multipath scintillations from long range acoustic transmissions on the New England continental slope and shelf

Andrew Fredricks, John A. Colosi,[a] James F. Lynch, and Glen Gawarkiewicz
Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

Ching-Sang Chiu
Naval Postgraduate School, Monterey, California 93943

Philip Abbot
OASIS Inc., Lexington, Massachusetts 02421

(Received 24 November 2003; revised 28 October 2004; accepted 31 October 2004)

Observations of intensity fluctuations for broadband, 400-Hz multipath arrivals observed during the 1996 New England shelfbreak PRIMER study are described. Acoustic signals were generated by two bottom-mounted sources located on the continental slope in roughly 290-m water depth and were received on a 52-m-long vertical-line-array (VLA) located in 93-m water depth. Propagation ranges were 42.2 and 59.6 km. The bathymetry, oceanography, and bottom geology of the PRIMER site are described. Acoustic observables of point intensity, peak intensity, and integrated energy over the VLA are treated in terms of the scintillation index, log-intensity variance, and intensity probability density functions (PDFs). Variability of the observables are decomposed into high and low frequency components with time scales less than and greater than 2 h, to facilitate correlation to ocean processes at different timescales. Parabolic equation numerical simulations using a quasi-random undular tidal bore model are able to reproduce many of the observed intensity fluctuation to within a factor of 2, and they allow investigation of scintillation behavior as a function of range. [DOI: 10.1121/1.1835510]

PACS numbers: 43.30.Re, 43.30.Ft, 43.20.Fn [WLS] Pages: 1038–1057

I. INTRODUCTION

In the last decade a monumental effort has been underway to understand shallow water acoustic propagation with its attendant complex oceanographic, biological, and geologic environment (Pace and Jensen, 2003; Duda and Farmer, 1999). In particular for ''long-range'' propagation applications, frequencies in the several hundreds of Hz region have been studied since lower frequencies are more effectively attenuated by geoacoustic interaction, and higher frequencies suffer from stronger scattering. The typical arrival pattern from a long-range broadband transmission, in shallow water, shows a complex multimode, multipath interference pattern, and the focus of experimental analysis to date has been on the travel time statistics of acoustic normal modes which make up this interference pattern (Headrick *et al.*, 2000a, b; Sperry *et al.*, 2003). Theoretical and numerical studies have examined the statistics of normal mode amplitudes in shallow water environments (Creamer, 1996; Teilburger *et al.*, 1997; Duda and Preisig, 1999), but little attention has been given to the intensity statistics of the total acoustical field in these cases (Pasewark *et al.*, 2002). In this paper we present a comprehensive analysis of observed acoustic intensity fluctuations for the Shelfbreak PRIMER study, and a numerical model is used along the lines of Teilburger *et al.* (1997) to interpret the observations.

The key to understanding long-range shallow water acoustic fields involves understanding the evolution of deterministic and stochastic multipath interference. Multipath interference leads to acoustical fields that are characterized by random walk phases and intensities with occasional deep fades and rare high intensity peaks. Theories of multipath interference effects on wave field statistics have been treated (Dyer, 1970; Dyson *et al.*, 1976; Flatté *et al.*, 1979) and recently expanded to include variable bandwidth (Colosi and Baggeroer, 2004). At short time scales the acoustical fields (phase and intensity) are dominated by fade statistics yielding rapid phase changes and a rms intensity variation of ~5.6 dB (Dyson *et al.*, 1976; Dyer, 1970). At long time scales the acoustical variability is not dominated by fades, and phase undergoes a ''random walk'' behavior while intensity is modulated by large scale ocean processes. Single frequency propagation in deep water over long ranges (Dyson *et al.*, 1976) seems to confirm this kinematical model, but shallow water cases have not been treated. In this paper, observed intensity fluctuations in this multipath regime are described in a shallow water oceanographic setting, and using wideband transmissions. The data consist of 9 days of 400-Hz broadband transmissions from the slope region up to the New England continental shelf (ranges of 42 and 60 km) during the summer shelfbreak PRIMER study. In this environment the acoustic field is a complex interference pattern of coupled modes and/or scattered rays (Sperry *et al.*, 2003), and the analysis focuses on the variability of the total acoustic field, and not on any particular wave front or normal mode arrival. In this case the acoustic field is affected by both deterministic and stochastic multipathing caused by the

[a]Author to whom correspondence should be addressed. Electronic mail: jcolosi@whoi.edu

complex propagation environment. In addition, as pointed out by Creamer (1996), since significant acoustical interaction with the seabed occurs, the effects of bottom loss must also be considered to be important, though the specific ocean model used by Creamer is not directly relevant to the PRIMER observations.

To facilitate identification with ocean processes, the acoustic field variability is separated into two time scale ranges with high/low frequency acoustic variability delineated by a transition frequency of 0.5 cph; this cutoff separates the well-known oceanographic frequencies such as the energetic internal-tide band ($f \sim 0.08$ cph) and mesoscale band ($f < 1$ cpd) from the equally energetic soliton and high frequency internal-wave band ($1 \text{ cph} < f < 10 \text{ cph}$) (Colosi *et al.*, 2001). This separation also delineates the acoustical regimes dominated by fade statistics (high frequency) and large scale modulation (low frequency) as discussed previously. This approach was taken over the direct spectral method because the transmissions were not continuous; the duty cycle of the transmissions was 4 min on, and 11 min off.

Most of the observed acoustic field variability occurs at high frequencies where rms log-intensity is 5.7 dB as compared to 2.5 dB for the low frequencies. Normalized intensity variance or scintillation index,

$$\mathrm{SI} = \frac{\langle I^2 \rangle}{\langle I \rangle^2} - 1, \tag{1}$$

gives another measure of intensity variability and again the high frequency acoustic variability dominates with SI of 1.1 compared to 0.3 for the low frequency. Scintillation index of 1 and rms log-intensity of 5.6 dB are trademarks of full saturation (Dyer, 1970; Flatté *et al.*, 1979; Colosi and Baggeroer, 2004), where the complex field is interpreted as the sum of a large number of random independent contributions; the central limit theorem then dictates that the real and imaginary components will approach zero mean Gaussian random variables. The low frequency acoustic variability, on the other hand, acts as a weak modulation of this Gaussian process.

Because of the impact of energy losses due to bottom interaction, another important acoustic observable is integrated energy (IE) over the receivers' vertical aperture and over the pulse duration; this observable quantifies the amount of energy being detected at the array. It is found that the IE has very small fluctuations indeed, with rms variations between 1.9 and 2.2 dB; this implies that losses from acoustic bottom interactions are rather steady over the 9-day experiment duration, in spite of large oceanographic modulations. Similar results have been recently reported by Duda *et al.* (2004) for long-range 400-Hz transmissions in the South China Sea.

Finally, to aid in the interpretation of the observations and to estimate the range variability of the acoustic fluctuation quantities, parabolic equation numerical simulations are performed through time evolving quasi-stochastic fields of nonlinear internal tides (solibores). Because the shelfbreak PRIMER environment was rather well characterized in terms of the mesoscale oceanography, the shelfbreak front, and the internal wave fluctuations (and to some extent the bottom geology), a somewhat realistic simulation can be carried out. The simulations are able to reproduce many of the observed intensity fluctuations within a factor of 2, and they indicate that near-saturation phenomena occur roughly within the first 10–20 km of propagation, with bottom loss playing a major role.

The organization of this paper is as follows: In Sec. II the PRIMER experiment physical setting and the acoustical parameters and geometry are described. Section II also has a description of the internal waves and parabolic equation numerical simulations. Section III presents the analysis of acoustical integrated energy, while analysis of point intensity is done in Sec. IV. Sections V and VI have discussion and summary.

II. THE NEW ENGLAND SHELFBREAK PRIMER EXPERIMENT

The summer 1996 New England shelfbreak PRIMER experiment was an integrated acoustics/physical oceanography deployment aimed at (1) understanding low frequency (200–500 Hz) acoustic propagation, coherence, and inverse methods in the presence of a strong shelfbreak front and (2) understanding the physical oceanographic dynamics of shelfbreak frontal systems. As seen in Fig. 1, the experiment was physically comprised of a long term mooring array line and an ''intense survey'' region; in this paper, the concentration will be on those measurements made in the intense survey area. These include high resolution hydrography with the SeaSoar undulating CTD, acoustic transmission measurements made with a source/receiver array, and bottom geoacoustic property measurements made with explosive sources. These data allow for the detailed study of the characteristics of sound propagation up the slope, through the front, and onto the shelf. A description of acoustic propagation through the front and large scale ocean features have been described in a Ph.D. thesis by Sperry (1999) and in other literature (Sperry *et al.*, 2003; Lynch *et al.*, 2003). The main aim of the present paper is to address the acoustic intensity fluctuations which were observed in the multipath arrivals. Since much of the experimental overview has been previously published (Sperry *et al.*, 2003), only a brief discussion will be given here.

A. Setting: Physical oceanography

SeaSoar surveys provided 1 km along track horizontal resolution of the 50×50-km intense study region once per day, with a parallel horizontal track spacing of 10 km. Moored temperature and current sensors gave numerous observations of both the mesoscale and finescale oceanography, including the important internal-wave field (Colosi *et al.* 2001). It was found that the PRIMER sound speed fluctuation fields with time scales between a few minutes to a day, were dominated by the existence of primarily mode 1 nonlinear internal tides and high frequency solitary-like waves (Colosi *et al.* 2001). These observations are the prime motivation for the sound-speed fluctuation model presented in Sec. II D. The cross-shelf structure of the large scale sound-speed field for the two main acoustic tracks on 31 July 1996

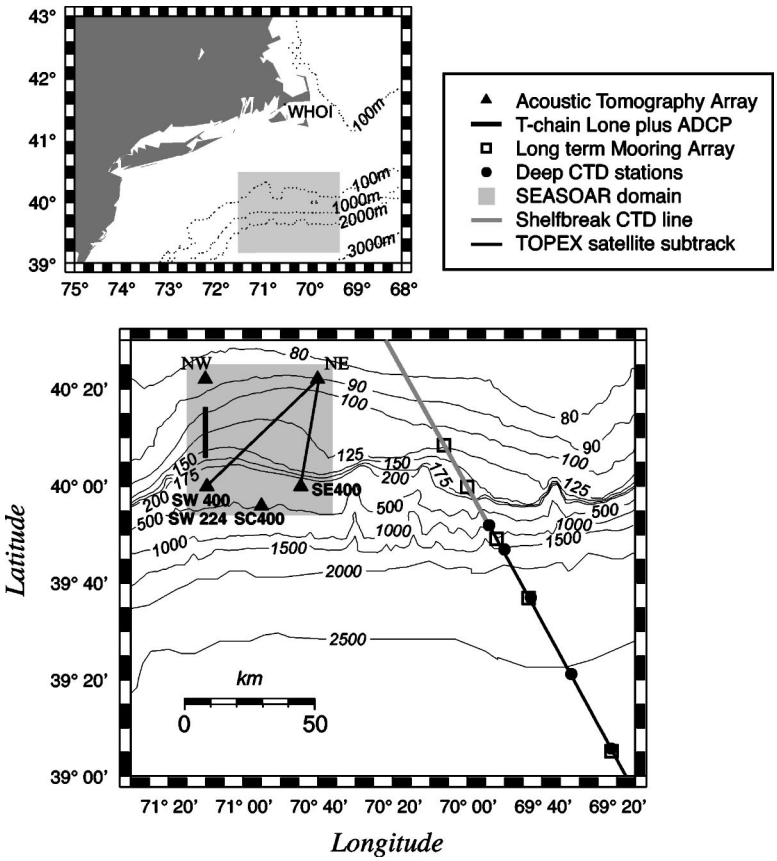


FIG. 1. Maps showing the experimental region of the New England Summer Shelfbreak Primer study, which took place between July and August 1996 on the shelfbreak just South of Cape Cod. Sound sources used in this study are in the SW and SE corners of the experimental region and the acoustic receptions discussed here were obtained on the NE receiver.

(Julian day 213) are shown in Fig. 2 as well as the alongtrack bathymetry. However, the sound speed field varied considerably on daily time scales. The dominant mode of variability was the westward propagation of a large amplitude frontal meander (Gawarkiewicz *et al.* 2004; Colosi *et al.* 2001). The westward propagation speed was 0.1 m/s and the wavelength was roughly 40 km. The cross-shelf amplitude of the meander was roughly 30 km. A key aspect of this meander relative to the acoustic intensity fluctuations was that during meander crest passage (maximum offshore/southward extent of cold shelf water) there was a mid-depth duct (low temperature/sound speed) that extended offshore of the southwest source. During the meander trough passage there was warm saline slope water above the southwest source, and the shelfbreak front was located between the source and receiver. Beginning on Julian day 213 (31 July), the meander crest passed over the southwest source. We will see that this impacted the intensity field for the low frequency fluctuations.

B. Setting: Subbottom geology

Turning to the geology, geoacoustic parameters were obtained from inversions of observed sound fields generated by SUS charges and the electronic sources (Potty *et al.*, 2000, 2003). These inverses were made using acoustic receptions at the northeast (NE) VLA receiver, and on the whole represent geoacoustic parameters in the shelf region, not the slope. A simplified bottom geoacoustic model based on these inverses is shown in Table I.

C. Setting: Acoustic propagation

Three source moorings were deployed to the south on the continental slope (see Fig. 1; SW, SC, and SE) and two vertical line array receivers were deployed to the north on the shelf (NW and NE). These instruments were continuously operational over an approximately 11-day period while the oceanographic environment was being measured by SeaSoar and other oceanographic devices. The SW source sent out both 224- and 400-Hz pseudorandom codes while the SC and SE sources only transmitted 400-Hz pseudorandom codes. These signals were match-filtered at the receiver to provide useful pulse compression and signal gain ($\simeq$25 dB). This study is concerned with the transmissions from the southeast (SE) and southwest (SW) 400-Hz sources to the northeast (NE) receiver array. The NE array had 16 hydrophones at depths of 44.4, 47.6, 49.8, 53.1, 56.4, 59.7,...,92.7 m with a water depth of 93 m; the shallowest few phones had separations of 3.2 and 2.2 m but the deeper phones all have 3.3-m separation. These sources (SE, SC, and SW) transmitted for 4 continuous minutes every 15 minutes (27% duty cycle), sending repeated 511-digit m-sequences each of 5.11-s length, which corresponds to the 10-ms digit length (or cor-

TABLE I. Geoacoustic parameters based on (Potty *et al.*, 2000, 2003) used in the parabolic equation acoustic simulation.

	Thickness (m)	Sound speed (m/s)	Density (kg/m^3)	β, Attenuation (dB/λ)
Layer 1	5	1525	1400	0.5
Layer 2	295	1750	1800	0.5

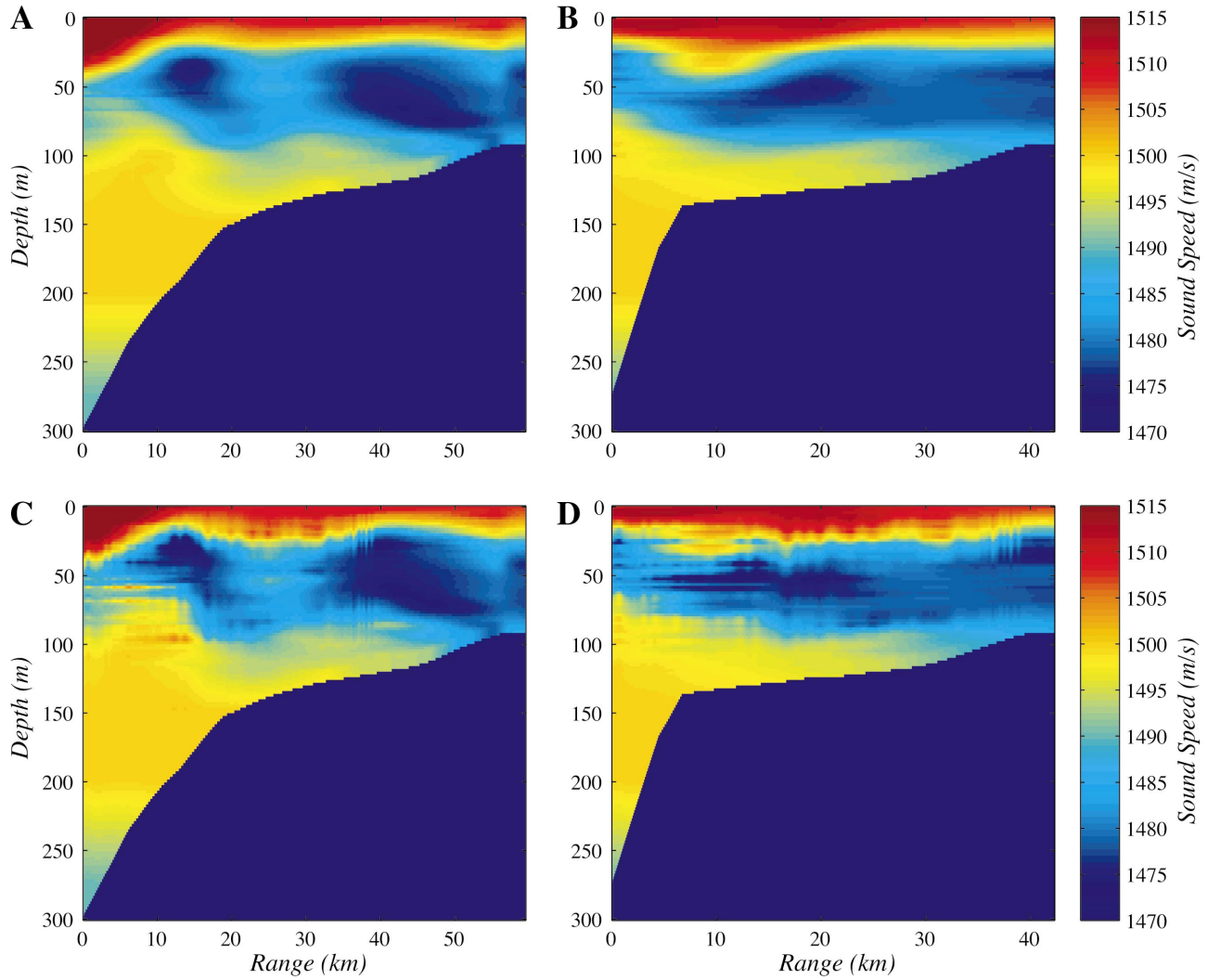


FIG. 2. Sound speed fields obtained by SeaSoar for 31 July 1996 (Julian day 213) are shown in panels (a) and (b) for the SW and SE acoustic propagation paths. Panels (c) and (d) show the SeaSoar fields for the SW and SE paths plus sound speed perturbation from the KdV internal tide model. The undular bore structure of the internal tide is evident near 40-km range on the two paths.

respondingly, the 100-Hz bandwidth) of the source. When match filtered, the m-sequences impulse response is obtained. Figure 3 shows a spectrogram of the SW and SE sources over a 10-min portion of a 15-min transmission cycle. The correlation time of the intensity at 400 Hz is seen to be very short (of the order a few minutes), while the correlation time of the 224-Hz data is much longer. The upper panels of Fig. 4 show examples of received 400-Hz timefronts from the SW and SE transmission paths, and the multipath interference pattern is very evident in the arrival structure; no time separated mode or wavefront arrivals are seen. The signal-to-noise (SNR) level is seen to be very good, and was typically 20–25 dB.

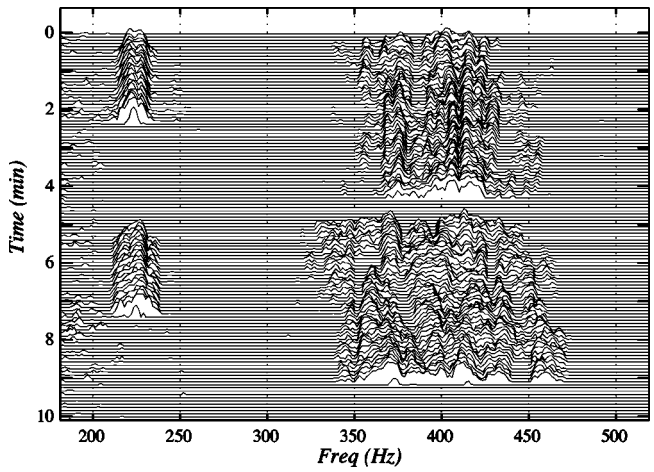


FIG. 3. Spectrogram (linear intensity) of a typical 10-min duration over the 15-min transmission cycle at the NE receiver. Spectral slices are 5.11 apart. The upper 350–450-Hz arrival is from the SW source while the lower one is from the SE source. The 224-Hz receptions, also seen, are not examined in this paper. Typical SNRs for the SE and SW sources were 24 and 20 dB, respectively.

D. Numerical models

To aid in the interpretation of the acoustic observations and to estimate the range evolution of scintillation behavior, a parabolic equation acoustic propagation model (RAM) is used (Collins, 1993, 1997). This model utilizes the large scale SeaSoar sound-speed fields plus an internal tide model

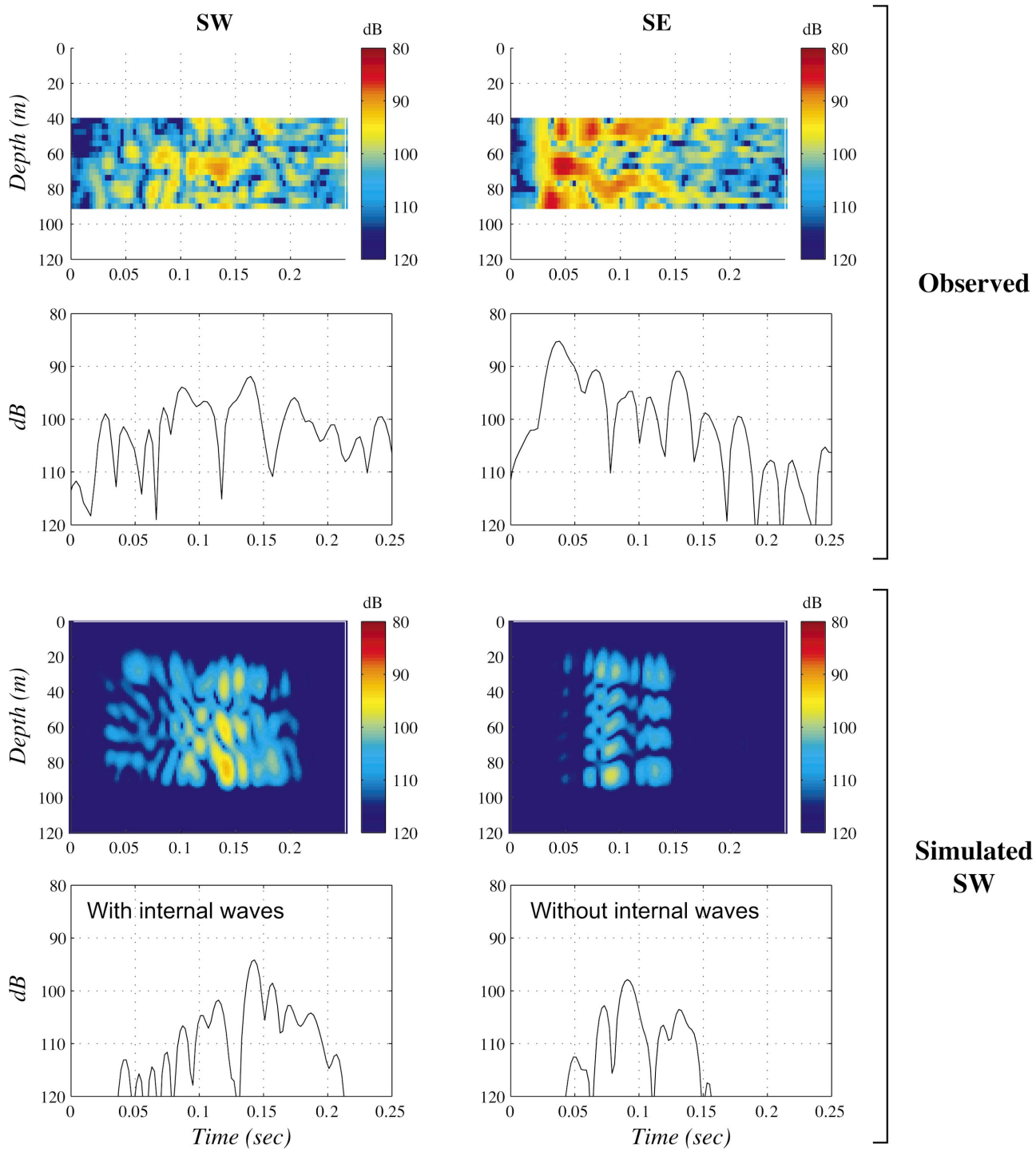


FIG. 4. Examples of observed time fronts for the SW and SE propagation paths (upper panels) from Julian day 205. In the lower panels are simulated time fronts with (left) and without (right) internal-wave sound-speed perturbations. Accompanying color timefront displays are ''line'' plots of the arrival pattern for the deepest hydrophone.

based on first-order Korteweg–deVries (KDV) internal-wave dynamics. From the PRIMER oceanographic observations it was determined that the sound speed fluctuation field was dominated by mode 1 internal tides and solitons, and therefore in the following analysis the ''stochastic'' background spectrum of internal waves is largely ignored or deemphasized. The issue of establishing the different scattering mechanisms and relative importance of the nonlinear versus stochastic internal wave fields is of intense interest (Sperry *et al.*, 2003, appendix) but is beyond the scope of the present analysis. While the internal tides and solitons in the PRIMER experiment were surely not obeying first-order KDV dynamics, an empirical approach was taken such that the KDV model parameters were tuned to produce model output that mimiced the PRIMER internal-wave observations, much like the dynamically incorrect ''epicycles'' celestial system of Ptolemy which still mimicked the planetary motions rather well.

TABLE II. KDV parameters for the internal tide simulation. The domain size is 107.3 km and the internal tide wavelength is 26.8 km.

c_0 (m/s)	α (s^{-1})	γ (m^3/s)	ϵ (m^2/s)	ν (s^{-1})	ζ_0 (m)
0.60	−0.03	650	0.05	2.0×10^{-5}	10

1. Nonlinear internal tides

Internal-wave-induced sound speed fluctuations are modeled as a vertical advection of local sound speed structure so that

$$\delta c(t,x,z)=\left(\frac{\partial c}{\partial z}\right)_p(x,z)\zeta(t,x,z), \tag{2}$$

where the first term is the local gradient of potential sound speed and the second term is the vertical displacement of a water parcel caused by internal waves. In conditions of weak nonlinearity, shear, and rotation, the first-order KDV theory applies (Apel *et al.*, 1995) and the internal-wave displacement is written as a separable function of time and range (t,x) and depth (z) coordinates,

$$\zeta(t,x,z)=A(t,x)B(z), \tag{3}$$

where B is the linear vertical mode function of the internal wave, and the time/range function A evolves according to the KDV equation,

$$\frac{\partial A}{\partial t}+c_0\frac{\partial A}{\partial x}+\alpha A\frac{\partial A}{\partial x}+\gamma\frac{\partial^3 A}{\partial x^3}=\frac{\epsilon}{2}\frac{\partial^2 A}{\partial x^2}-\nu A+F(t,x). \tag{4}$$

In the KDV model c_0 is the linear phase speed, α is a nonlinearity parameter, γ is a dispersion parameter, ϵ is a horizontal eddy viscosity, ν is a shoaling parameter (Liu, 1988), and F is a forcing function. The vertical mode function B is the lowest order internal tide mode obtained from the linear mode equation (Liu, 1988). Equation (4) is solved numerically using a second-order leapfrog method (Press *et al.*, 1992) over a domain of length four times the internal tide wavelength ($\lambda_{IT}=2\pi/k_{IT}=26.8$ km). Quasi-stochastic internal tides are generated using the forcing function F which oscillates one end of the computational domain at the tidal frequency ($\omega_{IT}=2\pi/12.42$ hrs), and a broad spectrum of randomly phased frequencies is added to represent a stochastic perturbation. The functional form of F is

$$F(t,x)=\exp(-x^2/\Delta^2)\left(F_0\sin[k_{IT}x-\omega_{IT}t]+\mathrm{Real}\sum_{n=1}^{N}a(k_n)e^{i[k_n x-\omega(k_n)t]}\right), \tag{5}$$

where F_0 is the internal tide forcing amplitude, $\omega(k)$ is the KDV dispersion relation (cubic), $N=256$, and $a(k_n)$ are Gaussian random deviates shaped using a Lorentzian wave number spectrum

$$S(k_n)=S_0\frac{k_0}{\pi}\frac{1}{k_n^2+k_0^2} \tag{6}$$

with $k_0=2\pi/10\,(\mathrm{km}^{-1})$, and $2\pi/107\leqslant k_n\leqslant 2\pi/0.4(\mathrm{km}^{-1})$. Consistent with the PRIMER observations (Colosi *et al.*, 2001) the forcing function and random deviates were normalized to put ten times more energy into the internal tide as compared to the random component. The exponential function at the beginning of Eq. (5) limits the influence of the forcing function to a few hundred meters at the end of the computational domain near the origin ($\Delta\sim300$ m). At the other end of the computational domain, a sponge layer is used to absorb all wave energy. The parameters c_0, α, γ, and ν may be computed directly from a knowledge of the density profile and the bottom slope (Liu, 1988); however, it is known that several (if not all) of the conditions for the validity of the first-order KDV are violated in the PRIMER experiment; therefore a more empirical approach is taken. The parameters of the KDV equation are tuned to the PRIMER internal wave observations discussed by Colosi

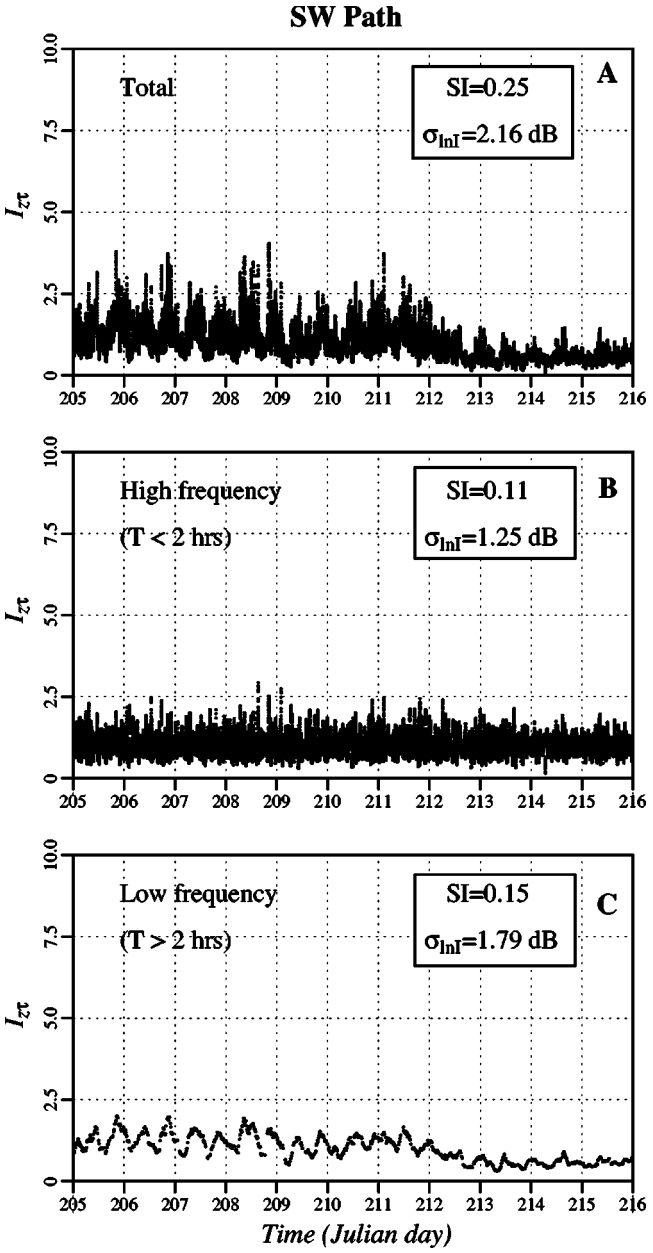


FIG. 5. Time series of depth and arrival time integrated energy, $\mathcal{I}_{z\tau}(l)$, for the SW path. In panel (a) the unfiltered time series is shown, while panels (b) and (c) display the high and low frequency components, respectively. In each case the series are normalized so that $\langle\mathcal{I}_{z\tau}\rangle=1$.

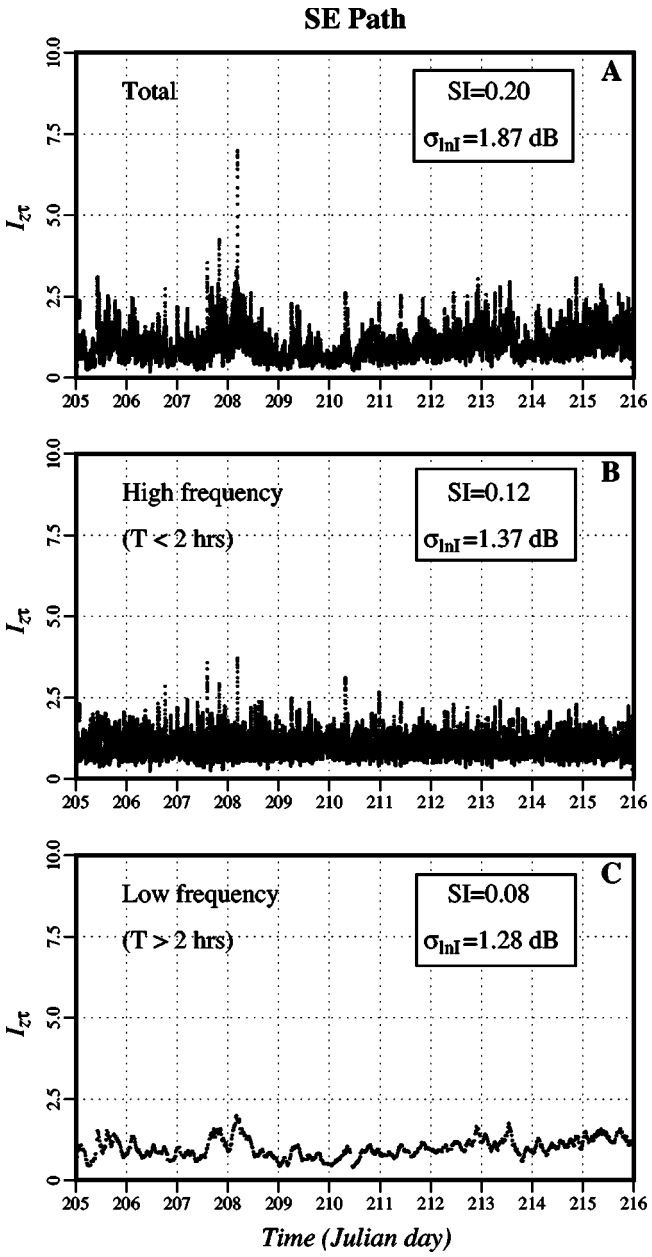


FIG. 6. Time series of depth and arrival time integrated energy, $\mathcal{I}_{z\tau}(l)$ for the SE path. In panel (a) the unfiltered time series is shown, while panels (b) and (c) display the high and low frequency components, respectively. In each case the series are normalized so that $\langle \mathcal{I}_{z\tau} \rangle = 1$.

et al. (2001). In particular, there is a rapid nonlinear evolution of the internal tide between the 150 and 120 m isobaths; thus the nonlinear and shoaling parameters α and ν are made functions of x and they increase from small values near the wave source to their final values near these isobaths. The nonlinear evolution of the internal tide is accompanied by the generation of high frequency internal waves. The rate of generation of these waves and their horizontal dimension are governed by the dispersion and viscosity parameters γ and ϵ, which are also chosen to mimic the PRIMER observations. Finally, the initial amplitude of the wave (ζ_0) is chosen to be 10 m, again in agreement with the PRIMER data. Table II shows a list of parameter values used in the simulations. The internal tide is assumed to propagate as a plane wave in the due north direction (x direction) so the displacement field is

TABLE III. Statistics for integrated energy, $\mathcal{I}_{z\tau}$. High-pass and low-pass variances of the log-intensity sum to give the total (table shows rms values), while SI values cannot be summed in this way. Simulation values are shown in parentheses ( ).

	Total		High pass		Low pass	
	$\sigma_{\ln I}$ (dB)	SI	$\sigma_{\ln I}$ (dB)	SI	$\sigma_{\ln I}$ (dB)	SI
SW	2.16	0.25	1.25	0.11	1.79	0.15
	(2.30)	(0.30)	(1.58)	(0.13)	(1.65)	(0.15)
SE	1.87	0.20	1.37	0.12	1.28	0.08

projected onto the acoustic propagation paths (x,y,z) using

$$\zeta(t,x,y,z) = A(t,x/\cos(\theta))B(z;x), \tag{7}$$

where θ is the angle relative to due north. For the SW and SE sources the angles are 45° and 8°, respectively. Further, the internal wave mode $B(z;x)$ is computed using the local water depth, so that the proper vertical structure is obtained. Internal tide temporal evolution was simulated over ten tidal cycles, with an internal-wave field sampled every minute; examples of the sound speed fields for the SW and SE paths are shown in Fig. 2.

2. Acoustic propagation

Both broadband and single-frequency acoustic propagation calculations were carried out using the RAM simulation code (Collins, 1993, 1997). Due to computational limitations, simulations were only carried out for the longer SW propagation path. For the calculations with internal wave perturbations a simulation was carried out every 15 min over the ten tidal cycles of simulated internal waves. The source was placed at a depth of 285.5 m, roughly 10 m above the bottom and the RAM self-starter was used to simulate a point source over the frequency band 350–450 Hz in increments of $\frac{1}{5}$ Hz. For the pulse calculations the frequencies were synthesized using a Gaussian window, yielding a source Q (rms bandwidth over center frequency) of roughly 0.11. For the SW simulation the final range was 59.6 km and the final water depth was 93 m. The bottom was simulated using a two-layer model (see Table I). Below the second layer a 100-m-thick artificial sponge layer was used to maintain computational stability. The value of bottom attenuation, β, in Table I is somewhat larger than the value obtained by geoacoustic inversion (Potty *et al.*, 2000, 2003) because larger values were needed to obtain better agreement with the PRIMER mean intensity and scintillations. This modification of the attenuation is justified given that the geoacoustic inversion estimates came from long-shelf acoustic data near the NW and NE receiver arrays, while the PRIMER data analyzed here samples the distinct slope and deep shelf regions. Figure 4 shows an example of the simulated broadband arrival pattern from the SW source both with and without the internal wave sound speed perturbation from Fig. 2. Note that while the peak intensities of the SW simulation with internal waves and SW observations are within a few dB, the SW observations show significantly more energy spread out in time. Further, it is seen that the calculation with internal waves has significantly more overall energy in the

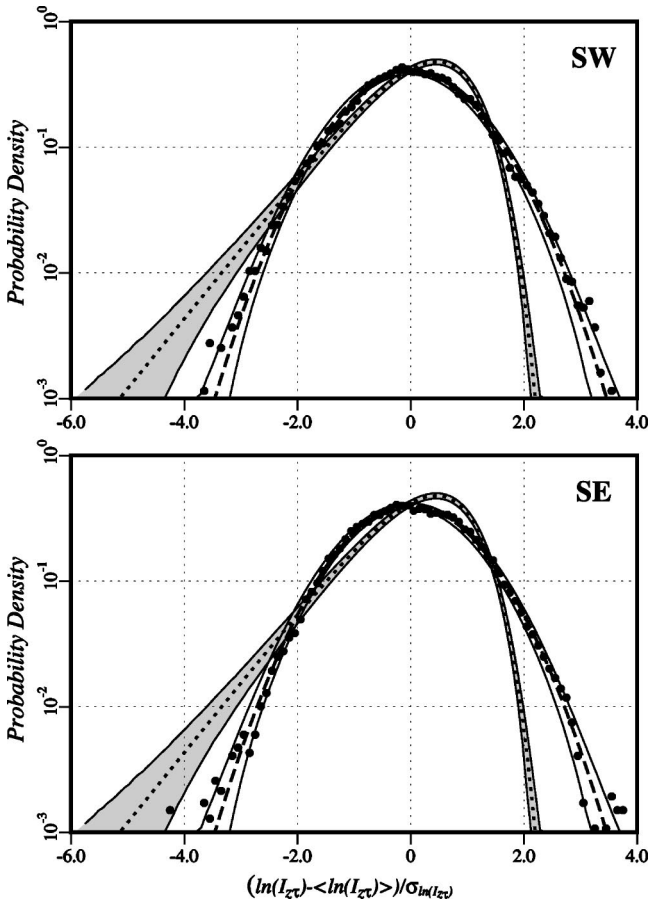


FIG. 7. Probability density functions of the high frequency component of $\mathcal{I}_{z\tau}(l)$ for the SW and SE transmission paths (●) and theoretical log-normal and exponential distributions (gray in-filling). Sampling error confidence intervals (95%) on the theoretical distributions (dash) are bound by solid lines.

arrival compared to the simulation without internal waves. These aspects will be discussed in more detail in Secs. IV A and C.

III. INTEGRATED ENERGY STATISTICS

The PRIMER intensity data are derived from the complex demodulates of the match filter output ψ and are given by $I(\tau,z,l)=|\psi(\tau,z,l)|^2$ where the three dependent variables are l the transmission number (or Julian day), z the hydrophone depth, and τ the arrival time. Figure 4 shows an example of the intensity as a function of the arrival time τ and depth z for one transmission (l) on Julian day 205 from the SW and SE sources. The multipath arrival shows a rapid onset of energy which is almost independent of depth followed by a peak and a slow decay of intensity into the noise. The time duration of the pulse is roughly 200 ms.

A. Integrated energy

An important observable is the total acoustic energy detected at the array as a function of transmission number l or Julian day, that is, intensity integrated over depth and arrival time.

$$\mathcal{I}_{z\tau}(l)=\int dz \int d\tau I(\tau,z,l), \tag{8}$$

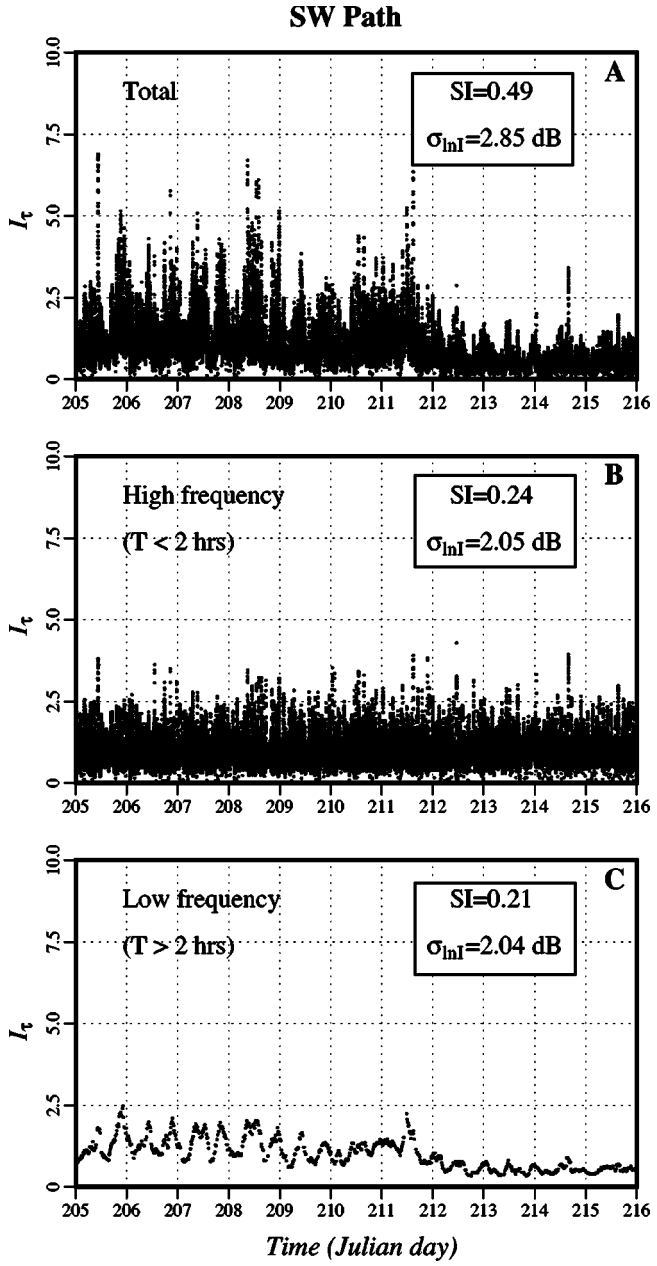


FIG. 8. Time series of arrival time integrated energy, $\mathcal{I}_\tau(z,l)$, for the SW path and the deepest hydrophone. In panel (a) the unfiltered series is shown, while panels (b) and (c) display the high and low frequency components, respectively. In each case the series are normalized so that $\langle\mathcal{I}_\tau\rangle=1$.

where the depth integral is done over the 16-element VLA aperture. The τ integral is done only over a subregion of the energetic part of the pulse (typically 400–500 ms) and this value is corrected for the expected noise contribution based on the noise levels immediately preceding and following the energetic region. Due to intermittent and often intense local shipping noise, some of the transmissions were excluded from this analysis. The time series of $\mathcal{I}_{z\tau}$ for the SE and SW transmissions are shown in the upper panels of Figs. 5 and 6, where the mean of $\mathcal{I}_{z\tau}$ has been normalized to unity. The two time series show very different characteristics; for the SW path the statistics of $\mathcal{I}_{z\tau}$ change around Julian day 213, whereas the SE path appears more statistically stationary (except for a significant event around Julian day 208). The change in SW path statistics around Julian day 213 coincides

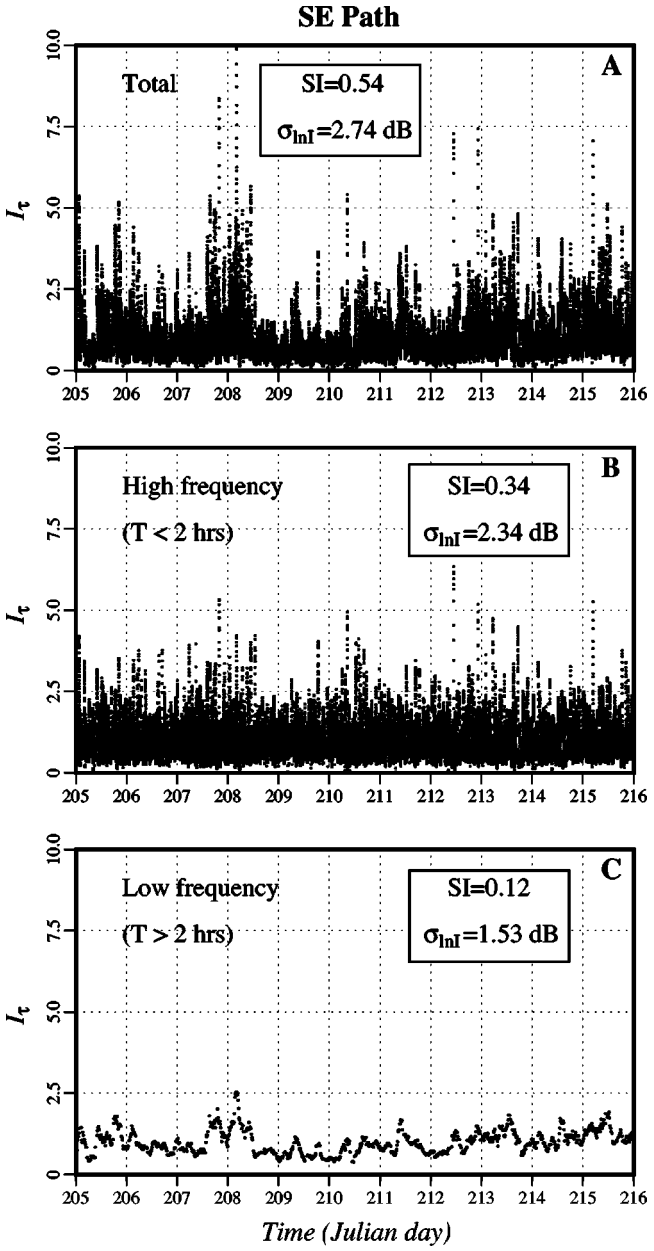


FIG. 9. Time series of arrival time integrated energy, $\mathcal{I}_\tau(z,l)$, for the SE path and the deepest hydrophone. In panel (a) the unfiltered series is shown, while panels (b) and (c) display the high and low frequency components, respectively. In each case the series are normalized so that $\langle \mathcal{I}_\tau \rangle = 1$.

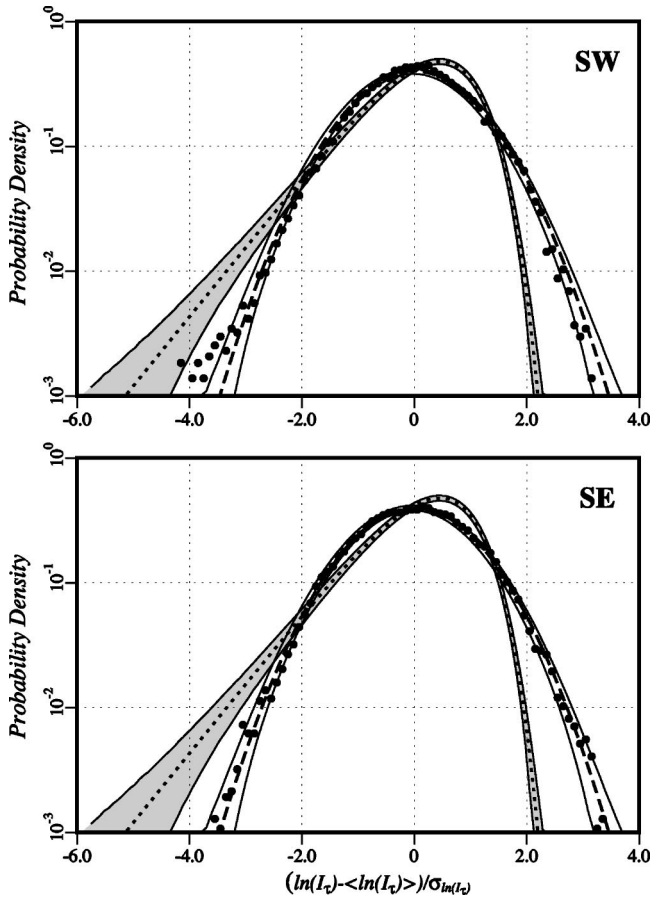


FIG. 10. Probability density functions of the high frequency component of $\mathcal{I}_\tau(z,l)$ for the SW and SE transmission paths (●) and theoretical log-normal and exponential distributions (gray in-filling). Sampling error confidence intervals (95%) on the theoretical distributions (dash) are bound by solid lines.

with the passage of a shelfbreak front meander which brought cool shelf water over the SW source.

The middle and bottom panels of Figs. 5 and 6 show the time series of $\mathcal{I}_{z\tau}$ separated into high and low frequency components with a 0.5 cph cutoff frequency; the low-pass filtering is accomplished using a moving average boxcar filter of 2-h duration. The high-pass time series is simply the unfiltered record minus the lowpass record. The 0.5 cph frequency cutoff for the acoustic fluctuations was used to separate the distinct oceanographic time scales of high frequency internal waves and solitarylike waves from the lower frequency tides and frontal meanders. Because of the irregular time sampling of the PRIMER acoustic data (4 min on, 11 min off), reliable spectra of intensity could not be computed.

For the SW transmissions it is seen that the transition around Julian day 212 is due to a change in the low frequency component and the high frequencies are more-or-less statistically stationary. Also, a clear semidiurnal tide component appears evident in the SW low frequency time series. In contrast, the time series from the SE path show a more

TABLE IV. Statistics for temporally integrated energy, $\mathcal{I}_\tau$. Quoted values are averages over the VLA, and (±) values are variations about the mean. High-pass and low-pass variances of the log-intensity sum to give the total (table shows rms values), while SI values cannot be summed in this way. Simulation values are shown in parentheses ( ).

	Total		High pass		Low pass	
	$\sigma_{\ln I}$ (dB)	SI	$\sigma_{\ln I}$ (dB)	SI	$\sigma_{\ln I}$ (dB)	SI
SW	2.58 ± 0.14	0.41 ± 0.04	1.87 ± 0.09	0.22 ± 0.01	1.84 ± 0.11	0.16 ± 0.03
	(2.99 ± 0.18)	(0.48 ± 0.05)	(2.31 ± 0.08)	(0.25 ± 0.02)	(1.81 ± 0.17)	(0.18 ± 0.03)
SE	2.52 ± 0.12	0.44 ± 0.05	2.15 ± 0.10	0.30 ± 0.02	1.38 ± 0.10	0.10 ± 0.01

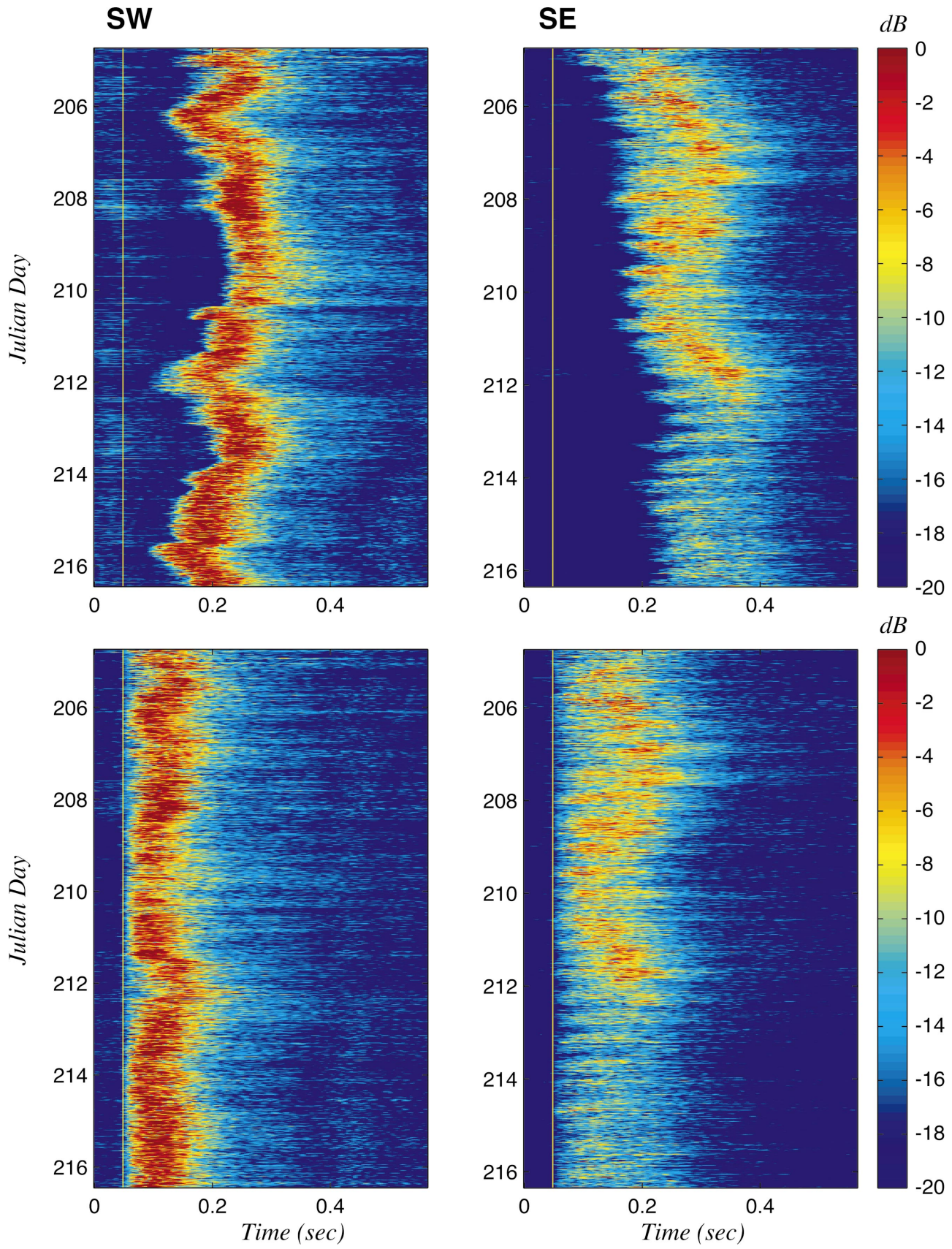


FIG. 11. Arrival patterns for the deepest hydrophone on the SW and SE transmission paths. Upper panels show data uncorrected for wander for SW (left) and SE (right) paths. Wander corrected data are shown in the lower panels.

''broadband'' variability with no clear semidiurnal or low frequency signature. This effect illustrates the dramatic differences between the internal tides and the low frequency oceanography across the two paths. The mesoscale decorrelation length across the PRIMER area was of order 10 km (Gawarkiewicz *et al.*, 2004), so that a difference in low frequency variability between the SW and SE paths is to be expected. Moreover, the internal tide field was shown to be significantly weaker to the east (Colosi *et al.*, 2001; Shearman and Lentz, 2004) and thus the stronger semi-diurnal signal in the SW path is also to be expected.

Table III shows an analysis of variance for the integrated energy for the two transmission paths. In this analysis two measures of variability are used, that is the scintillation index defined by Eq. (1), and the variance of log-intensity given by

$$\sigma^2_{\ln I} = \langle (\ln I)^2 \rangle - \langle \ln I \rangle^2. \tag{9}$$

Scintillation index, being a linear measure of intensity variation, is more sensitive to high intensity values, while the logarithmic distortion of low intensity values makes $\sigma^2_{\ln I}$ more sensitive to fades. It is seen that the temporal variability of $\mathcal{I}_{z\tau}$ has roughly equal contributions from low and high frequencies, and that the variances are not large. This result suggests that lossy bottom interactions must *not* be strongly modulated by the variable oceanography in the PRIMER region; in fact, the numerical simulations in this study suggest that even the small values in Table III must be upper bounds, as some of the energy leaks above the vertical array. Further, the differences of low frequency variances and SI between the SW and SE paths are significant and are likely due to the longer range of the SW path, and the stronger internal tide found to the west. Table III also shows that the PRIMER observations and the numerical model are in very good agreement.

Finally, Fig. 7 shows the probability density functions (PDFs) for the high frequency component of $\mathcal{I}_{z\tau}$ for the two paths. A PDF could not be constructed for the low frequency component because the time series was not long enough to yield a sufficient number of independent samples. For example, taking four independent samples per day yields only 36 degrees of freedom. The PDFs of the high frequency component are seen to be very close to log-normal. The log-normal PDF is expected in conditions under which the wave function evolves in range under a sequence of multiplicative ''filter'' operations, that is

$$\psi = \prod_j F_j \psi_0 \rightarrow \ln\left(\frac{\psi}{\psi_0}\right) = \sum_j \ln F_j, \tag{10}$$

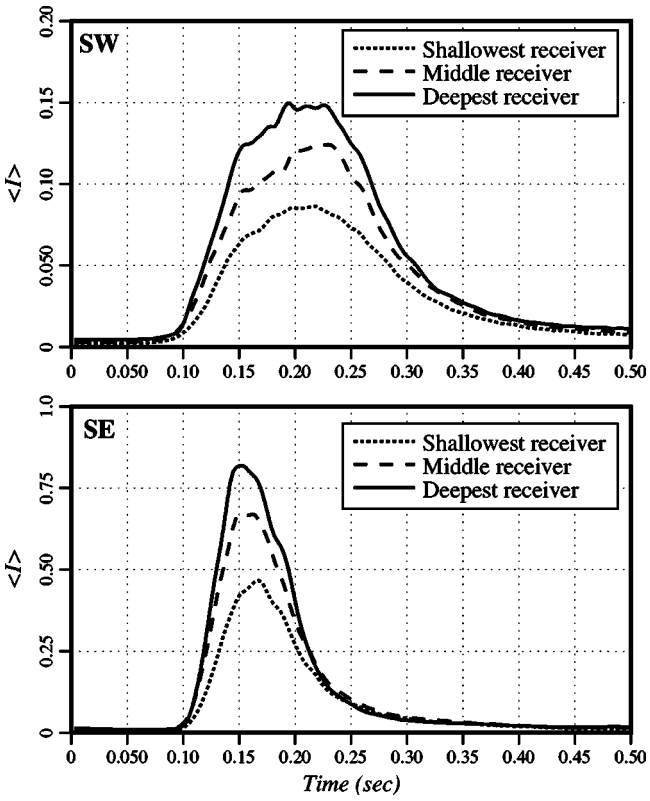


FIG. 12. Mean arrival pattern for three different hydrophone depths for the SW and SE transmission paths.

where ψ_0 is the initial wave function and F_j are operators at the jth range step.[1] By the central limit theorem, the sum of a large number of random variables, $\ln F_j$, will approach a Gaussian distribution and thus ψ has a log-normal distribution. A case of relevance here, perhaps, is propagation though a sequence of lossy slabs so that $\psi_j = \psi_{j-1} \exp(-\Delta x \alpha_j)$, where the attenuation coefficients α_j are random variables.

B. Temporally integrated energy

The second observable to be treated is the temporally integrated energy, defined as

$$\mathcal{I}_\tau(z,l) = \int d\tau I(\tau,z,l) \tag{11}$$

where, as before, the τ integral is only done over an energetic subregion of the pulse (typically 400–500 ms) and this value is corrected for the expected noise contribution based on the noise levels immediately preceding and following the ener-

TABLE V. Statistics of intensity scintillation for an arrival time near the peak of the mean pulse. Quoted values are averages over the VLA, and (±) values are variations about the mean. High-pass and low-pass variances of the log-intensity sum to give the total (table shows rms values), while SI values cannot be summed in this way. Simulation values are shown in parentheses ( ).

	Total		High pass		Low pass	
	$\sigma_{\ln I}$ (dB)	SI	$\sigma_{\ln I}$ (dB)	SI	$\sigma_{\ln I}$ (dB)	SI
SW	6.11 ± 0.07	1.80 ± 0.11	5.65 ± 0.07	1.08 ± 0.03	2.45 ± 0.09	0.30 ± 0.02
	(5.96 ± 0.27)	(1.76 ± 0.21)	(5.46 ± 0.25)	(0.98 ± 0.06)	(2.43 ± 0.26)	(0.34 ± 0.09)
SE	6.28 ± 0.08	1.82 ± 0.13	5.74 ± 0.05	1.05 ± 0.04	2.59 ± 0.12	0.34 ± 0.03

TABLE VI. Statistics of peak intensity $I_p(z,l)$. The first column shows average peak intensity and remaining columns show fluctuations as in Tables III–V. Quoted values are averages over the VLA, and (±) values are variations about the mean. High-pass and low-pass variances of the log-intensity sum to give the total (table shows rms values), while SI values cannot be summed in this way. Simulation values are shown in parentheses ( ). In the first column the two simulation values for $\langle I_p \rangle$ are the cases with and without internal waves.

		Total		High pass		Low pass	
	$\langle I_p \rangle$ (dB *re* 1 μPa)	$\sigma_{\ln I}$ (dB)	SI	$\sigma_{\ln I}$ (dB)	SI	$\sigma_{\ln I}$ (dB)	SI
SW	94.1	2.75 ± 0.10	0.67 ± 0.05	1.88 ± 0.03	0.29 ± 0.01	1.86 ± 0.10	0.24 ± 0.03
	(96.0/89.2)	(3.18 ± 0.14)	(0.58 ± 0.09)	(2.64 ± 0.11)	(0.36 ± 0.03)	(1.69 ± 0.12)	(0.15 ± 0.02)
SE	88.1	2.83 ± 0.10	0.70 ± 0.03	2.31 ± 0.05	0.41 ± 0.01	1.47 ± 0.09	0.17 ± 0.01

getic region. This observable is important because it shows the depth dependence of the observed energy fluctuations; variability seen in $\mathcal{I}_\tau$ and not seen in $\mathcal{I}_{z\tau}$ is associated with vertical energy redistribution processes like mode coupling or ray scattering. Moreover, there is also interest in the statistics of $\mathcal{I}_\tau$ which is the classic ''energy detector'' mode of a sonar system. The time series of $\mathcal{I}_\tau$ for the SE and SW transmissions and for the deepest hydrophone are shown in the top panels of Figs. 8 and 9, where the mean of $\mathcal{I}_\tau$ has been normalized to unity. These time series show more total fluctuation than the $\mathcal{I}_{z\tau}$ series, which suggests that there is some significant vertical redistribution of energy occuring. As in the previous section, high- and low-pass filtered acoustic time series (middle and bottom panels of Figs. 8 and 9) shows more-or-less stationary high frequency acoustic fluctuations and significantly different low frequency acoustic time series on the SW and SE paths. Table IV shows an analysis of variance over the VLA using the temporally integrated energy for the two transmission paths. The mean variances and uncertainties in Table IV are computed over a 5-point subset (phones 1, 4, 8, 12, and 16) of the 16 hydrophone array. The uncertainties represent the variability of the variance estimates over the VLA which are evidently small; thus the time series in Figs. 8 and 9 for the deepest phone are quite representative. For the SW path there are roughly equal contributions from low and high frequency variability, but for the SE path the high frequency portion is dominant. Comparing to Table III, the temporally integrated energy has about 1 dB more fluctuation, but again the variability is not large. Further, there is little variability of the intensity statistics across the array; this suggests that the number of interfering multipaths is not a strong function of depth.

Finally the PDFs of the high frequency temporally integrated energy are shown in Fig. 10; the PDFs are close to log-normal for the same reasons discussed in the previous section.

IV. POINT STATISTICS

Having examined integrated energy statistics, the focus is turned to statistics of ''point'' observations of the broadband intensity $I(\tau,z,l)$. Analysis of normal mode arrivals (Sperry, 1999) and examination of Fig. 4 shows that the arrival pattern is not consistent with a model of time separated normal modes or ray wave front arrivals; instead the pattern is the result of complex multipath interference which suggests the following wave function representation:

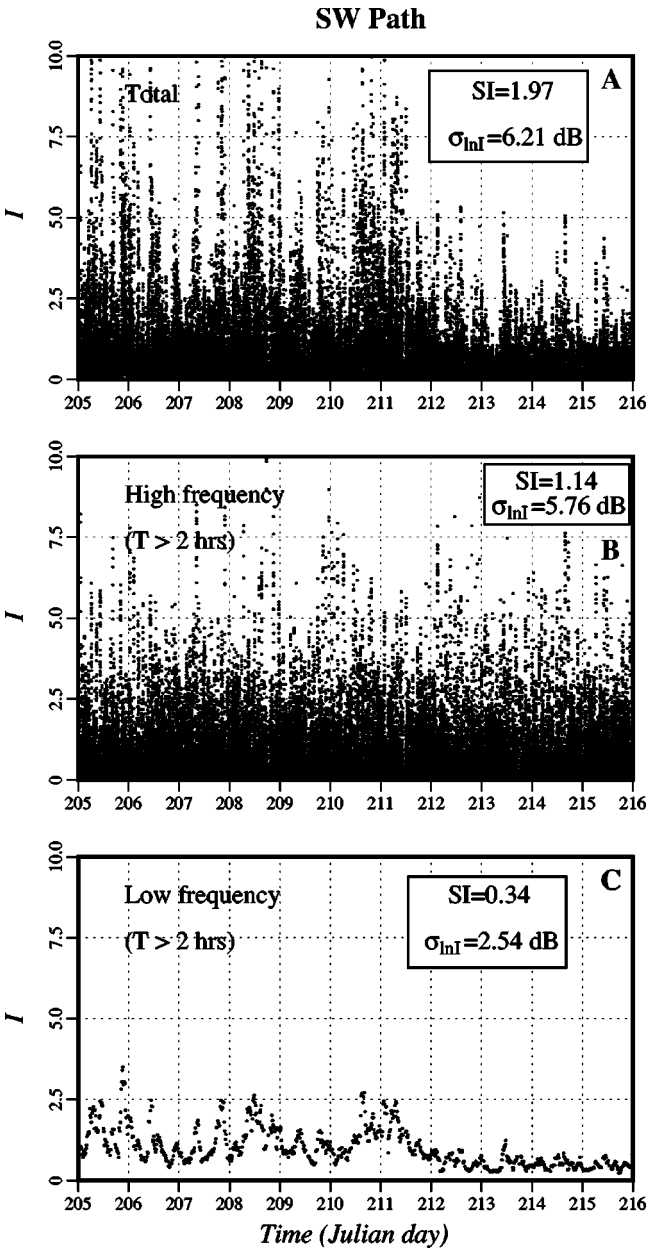


FIG. 13. Time series of intensity, $I(\tau,z,l)$, for the SW path, the deepest hydrophone, and τ near the peak of the mean pulse. In panel (a) the unfiltered series is shown, while panels (b) and (c) display the high and low frequency components, respectively. In each case the series are normalized so that $\langle \mathcal{I} \rangle = 1$.

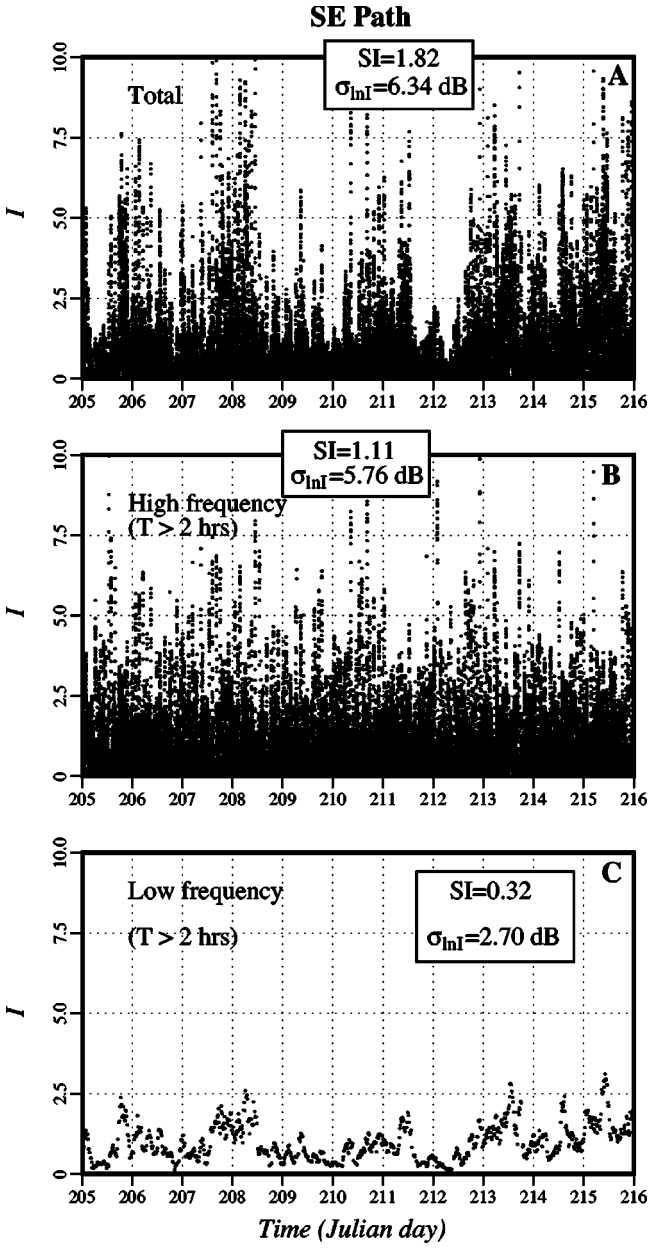


FIG. 14. Time series of intensity, $I(\tau,z,l)$, for the SE path, the deepest hydrophone, and τ near the peak of the mean pulse. In panel (a) the unfiltered series is shown, while panels (b) and (c) display the high and low frequency components, respectively. In each case the series are normalized so that $\langle I\rangle=1$.

$$\psi(\tau)=\sum_{n=1}^{N} a_n(\tau)\mathrm{e}^{i\phi_n(\tau)}. \tag{12}$$

For broadband signals, important parameters are Q, the rms bandwidth divided by the center frequency,[2] Φ^2, the mean square phase variance of the multipaths, and N, the number of independent multipaths. For narrow-band and broadband cases the scintillation index for independent multipaths can be written (Colosi and Baggeroer, 2004)

$$\mathrm{SI}\simeq 1+\frac{1}{N}\left(\frac{\langle a^4\rangle}{\langle a^2\rangle^2}-2\right), \quad \text{narrow band:}$$
$$Q=0, \quad \text{and} \quad \Phi\gg 1, \tag{13}$$

$$\mathrm{SI}\simeq 1+\frac{1}{N}\left(Q\Phi\frac{\langle a^4\rangle}{\langle a^2\rangle^2}-2\right), \quad \text{broadband:} \quad Q\Phi\gg 1. \tag{14}$$

SI is seen to approach the saturation value of 1 in the limit of large N, with the approach to 1, either from above or below, dictated by the sign of the term multiplying $1/N$. The approach to saturation, examined by numerical simulation, will be a key consideration in Sec. V. In the saturation limit, the PDF of intensity approaches the exponential distribution

$$P(I)=\frac{1}{\langle I\rangle}\ \exp\left(-\frac{I}{\langle I\rangle}\right). \tag{15}$$

A consequence of the exponential PDF is that the rms of log-intensity is 5.6 dB (Dyer, 1970). Because of multipath interference, the PRIMER data are expected to have SI near unity, rms log-intensity near 5.6 dB, and an intensity PDF which is near exponential. However, before proceeding with this analysis, the mean pulse must be examined, since intensity moments may be a function of the arrival time τ within the broadband arrival pattern.

A. Mean pulse

Figure 11 shows the intensity for the deepest hydrophone as a function of arrival time τ and Julian day (transmission number l) for both the SW and SE propagation paths. The entire arrival pattern is seen to wander in time due to range average sound speed changes along the transmission paths. This wandering is not of any real significance to the present analysis and so it is removed by lining up the leading edges of the arrivals which represent a pseudo-mode one (fastest mode) arrival. The leading edges are lined up by a combination of objective (edge detection) and subjective techniques. The data so corrected are shown in Fig. 11, where a smaller modulated time spreading of the arrival pattern is now evident (more on this later). Averaging the wander-corrected intensities over Julian day gives the mean pulse shown in Fig. 12. The mean pulses are seen to have a rapid onset followed by a peak and a slow decay back to noise levels. Both SW and SE paths show the largest intensity levels on the deepest hydrophones, which is consistent with a low-mode dominated field in a downward refracting waveguide. The SW and SE paths, however, differ in the total arrival duration which is roughly 250 ms for SW and 150 ms for SE; this effect is due in part to the longer propagation range for the SW path, but because of the connection between scattering and bottom loss, one cannot discount a contribution from different scattering histories.

The SW simulation mean pulse (not shown) is quite different from the observed mean pulse in Fig. 12, both from the standpoint of the arrival pattern and the depth distribution of the energy. The simulated mean pulse shows a slow onset of the mean pulse followed by a rapid cutoff, presumably due to the large attenuation β. Also the simulation shows more energy higher in the water column than near the bottom. In fact, the simulated acoustic field is much more dominated by higher order acoustic modes (see Fig. 4, for example) than the observations. This presents an interesting problem for future work, as the free parameter β was chosen

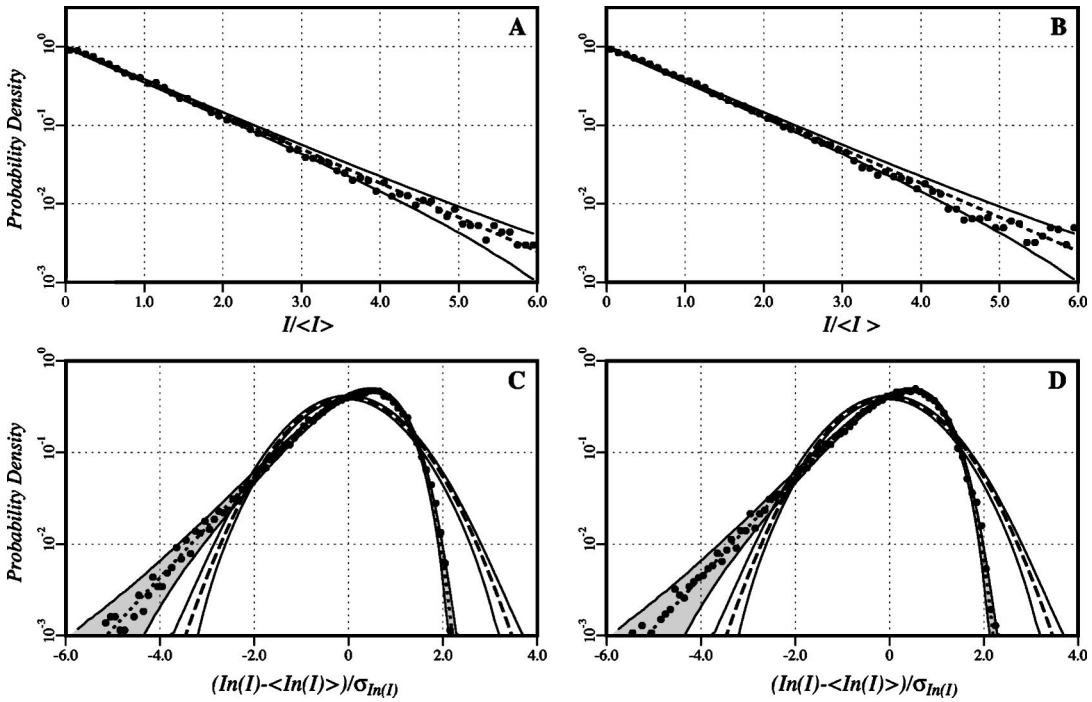


FIG. 15. Probability density functions of the high frequency component of $I(\tau,z,l)$ for the SW (left panels) and SE (right panels) transmission paths (●). The observed PDFs are shown with both logarithmic scales [panels (c) and (d)] and linear scales [panels (a) and (b)], where they are compared to theoretical log-normal and exponential distributions (gray infilling). Sampling error confidence intervals (95%) on the theoretical distributions (dash) are bounded by solid lines.

(increased) to match the scintillation values (Tables IV–VI). The large and small scale oceanography was well measured and so we discount this effect.

B. Point scintillations

Normalizing the instantaneous, wander-corrected arrivals with the mean pulse to give $\langle I\rangle = 1$, one obtains the unit mean intensity fluctuation as a function of arrival time. An arrival time near the peak of the mean pulse is chosen, and time series of intensity fluctuation for the SW and SE paths are shown in the upper panels of Figs. 13 and 14. Breaking the time series into low and high frequency components as before yields the time series shown in the middle and lower panels of Figs. 13 and 14, with VLA-averaged intensity statistics displayed in Table V. As in Table IV the mean variances and uncertainties in Table V are computed over a 5-point subset (phones 1, 4, 8, 12, and 16) of the 16 hydrophone array. The uncertainties represent the variability of the variance estimates over the VLA which are evidently small; thus the time series in Figs. 13 and 14 for the deepest phone are also quite representative. The unfiltered time series of intensity (Figs. 13 and 14) shows for the SW path the now familiar transition near Julian day 212, while the SE path shows a more homogeneous behavior, with occasional bursts of variability. For these point measurements, the intensity variability is indeed large, and is above the saturation limit; for the SW and SE paths the SI values are 1.80 and 1.82, while the rms log-intensity fluctuations are 6.11 and 6.28 dB. As in the integrated energy cases, the high and low frequency time series (Figs. 13 and 14) display a relatively homogeneous, stationary high frequency component (as was confirmed by using the Kolmogorov–Smirnov test) and a nonstationary low frequency component. Semidiurnal variability is still clearly seen in the SW path low frequency series, as is the transition near Julian day 212. For the SE path, the bursts of variability (Fig. 14) are associated with an increased low frequency component. In contrast to the integrated energy observable, the variance of point intensity now is clearly dominated by the high frequency component (Table V). For the SW path the low/high frequency SI values are 0.30/1.08, while the rms log-intensity values are 2.45/5.65 dB. The case is similar for the SE path with low/high frequency SI values of 0.34/1.05, while the rms log-intensity values are 2.59/5.74 dB. Interestingly, for the point statistics the differences between SW and SE propagation paths are not so apparent as in the case of integrated energy; this is because both paths are so close to saturation. Finally, Fig. 15 shows the intensity PDFs of the point intensity high frequency component, and the PDFs are seen to be very close to the exponential PDF [Eq. (15)] indicative of full saturation.

The immediately previous analysis was for a fixed arrival time τ near the peak of the mean pulse, however the statistics as a function of arrival time are also of some interest. Unfortunately, this analysis cannot be carried out adequately due to the diminishing signal-to-noise ratio (SNR) away from the peak. We find that the scintillation statistics are roughly the same over the energy-containing ''centroid'' of the mean pulse, but on the tails these statistics may vary wildly due to SNR problems, and also due to artifacts of the wander correction (see Fig. 12).

C. Peak statistics

The last observable that is treated is the peak intensity,

$$I_p(z,l) = \max_\tau(I(\tau,z,l)), \tag{16}$$

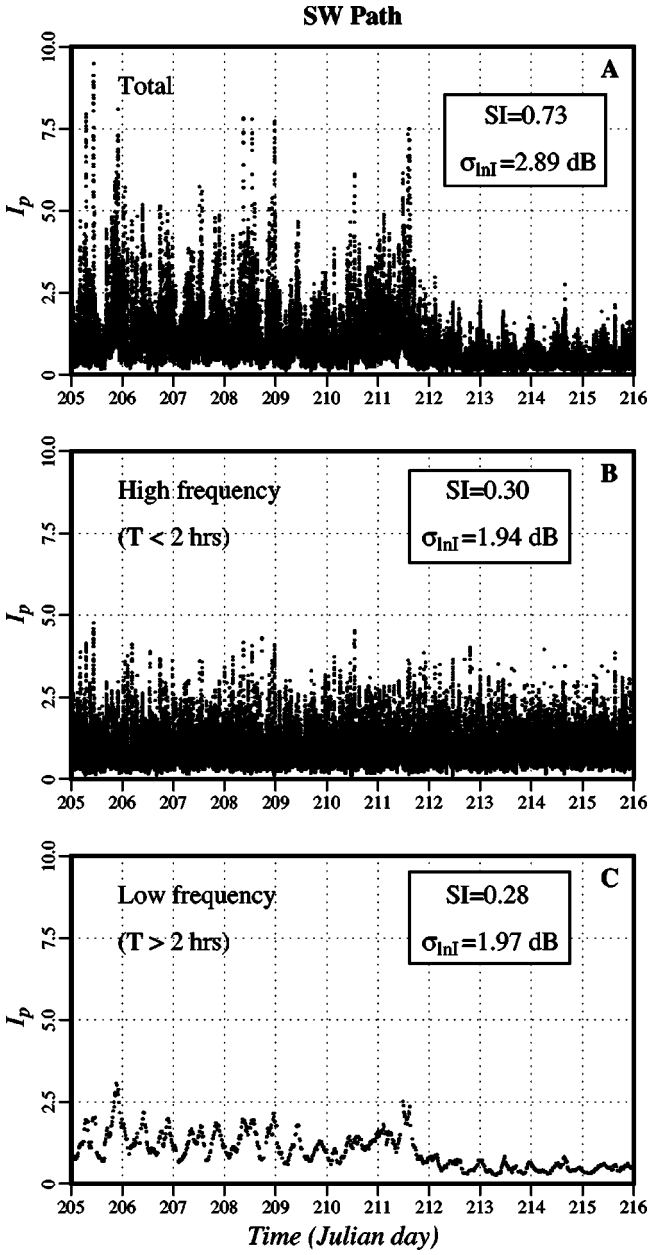


FIG. 16. Time series of peak intensity, $I_p(z,l)$, for the SW path and the deepest hydrophone. In panel (a) the unfiltered series is shown, while panels (b) and (c) display the high and low frequency components, respectively. In each case the series are normalized so that $\langle I \rangle = 1$.

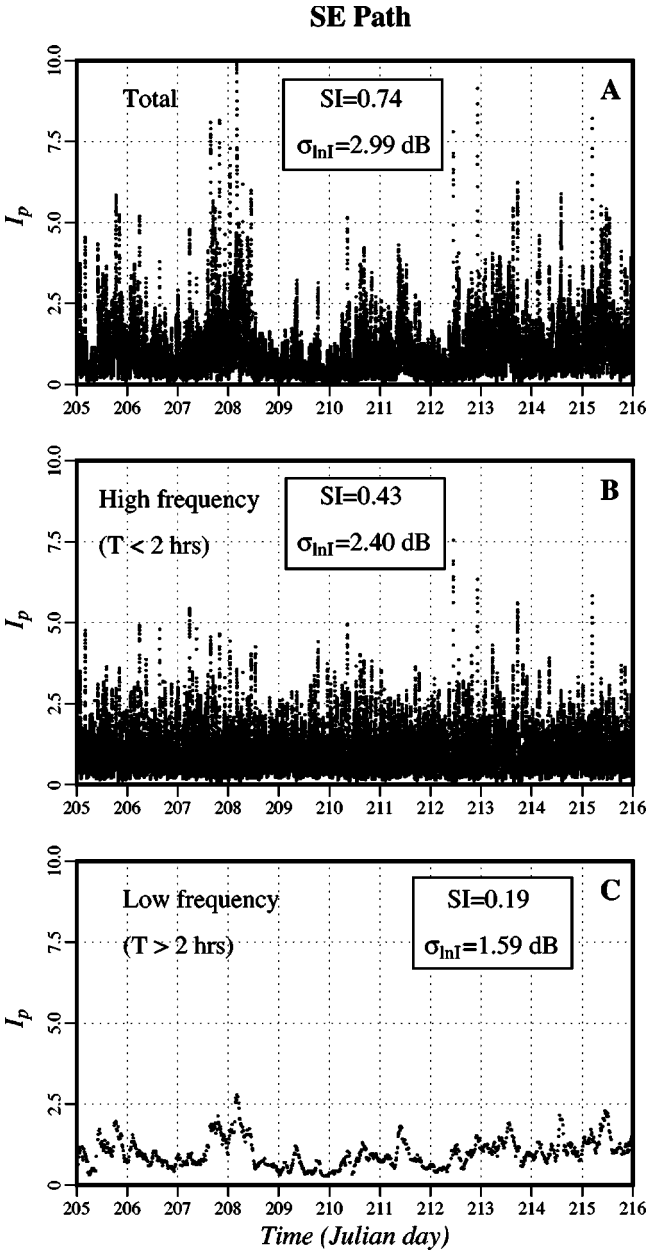


FIG. 17. Time series of peak intensity, $I_p(z,l)$, for the SE path, and the deepest hydrophone. In panel (a) the unfiltered series is shown, while panels (b) and (c) display the high and low frequency components, respectively. In each case the series are normalized so that $\langle I \rangle = 1$.

which is important in many sonar applications. Unlike the analysis of the previous observables, the *mean* peak intensity, which has been derived directly from the calibrated hydrophone outputs, is examined. Table VI shows that the model with internal-wave sound-speed perturbations does a good job of predicting the mean peak intensity, and that the model without internal waves underestimates the mean peak intensity by 6.8 dB. This calculation demonstrates the significant biases that can occur in transmission levels due to the competing effects of mode coupling by internal waves and differential modal attenuation. In this case the initial (source) acoustic field is dominated by high lossy modes, but internal waves couple these higher modes into lower less-lossy modes; the result is that energy at the receiver will be in excess of the no-internal-wave case.

In treating fluctuations, variances of peak intensity are naturally lower than point intensity variance, since deep fades are excluded by definition. For the SW and SE paths, rms log-intensities are found to be 2.75 and 2.83 dB, while SI values are 0.67 and 0.70, respectively (Table VI). The peak intensity variability is seen to be very close to that of time integrated energy ($\mathcal{I}_\tau$) (Table IV). Further, the time series in Figs. 16 and 17 closely resemble the $\mathcal{I}_\tau$ time series shown in Figs. 8 and 9; the correlation coefficients between them are 0.82 and 0.91, respectively. Thus, not surprisingly, higher peak intensity is associated with a larger overall energy in the arrival pattern. Finally, the PDFs of peak intensity (like $\mathcal{I}_\tau$) are seen to be very close to log-normal (Fig. 18).

Another interesting feature of peak arrivals (and thus $\mathcal{I}_\tau$) is the connection to the total arrival time spread. Figure 11

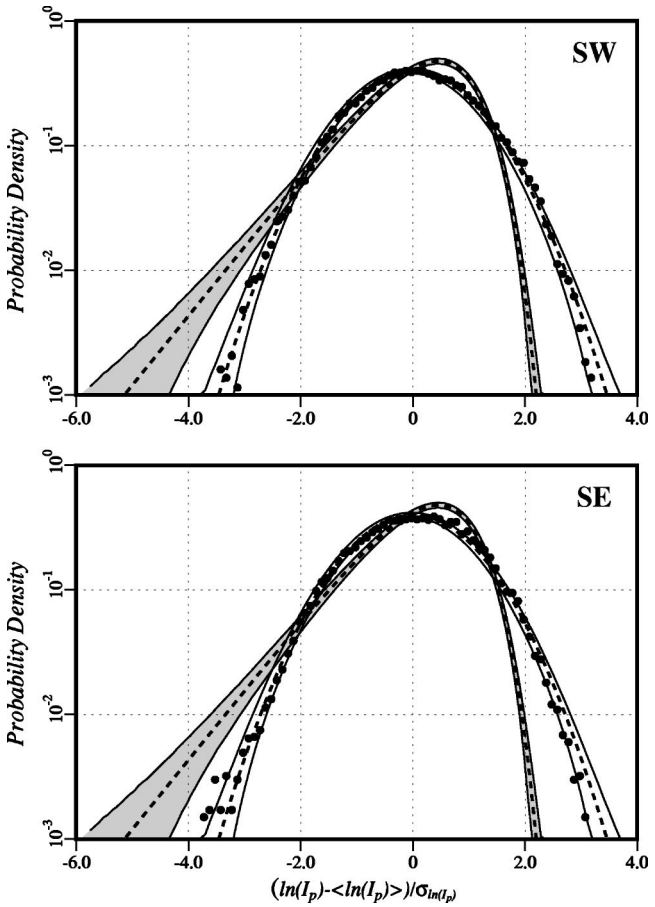


FIG. 18. Probability density functions of the high frequency component of $I_p(z,l)$ for the SW and SE transmission paths (●) and theoretical log-normal and exponential distributions (gray infilling). Sampling error confidence intervals (95%) on the theoretical distributions (dash) are bounded with solid lines.

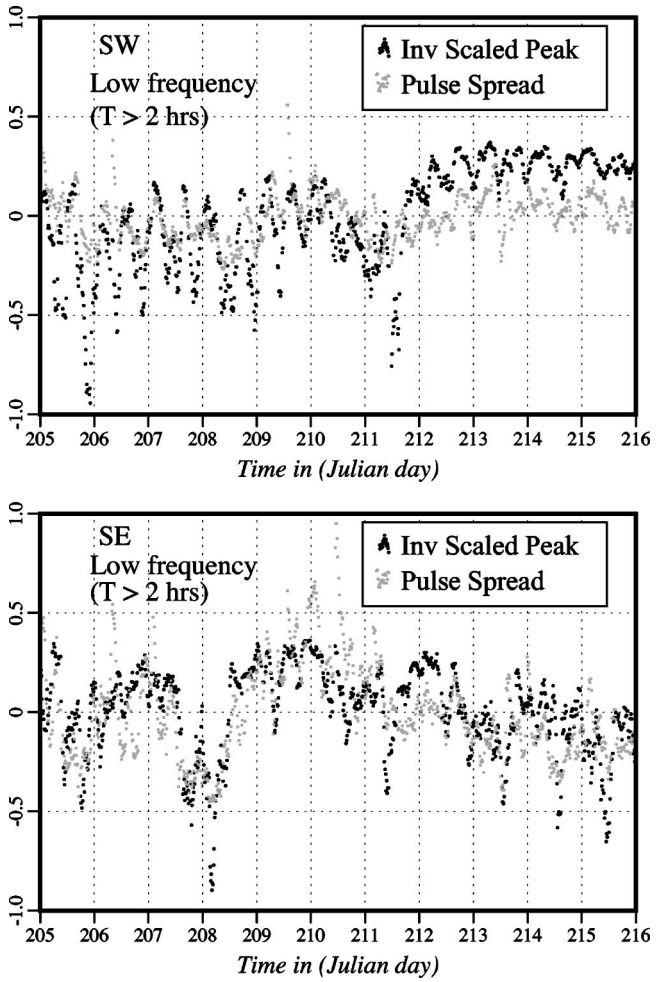


FIG. 19. Low frequency filtered time series of demeaned pulse time spread (black), and −1 times demeaned peak intensity (gray) for both the SW and SE transmission paths. The peak intensity fluctuation is multiplied by −1 to show more clearly the negative correlation between the time spread and peak intensity series. The correlation coefficients between the time spread and peak intensity series are −0.47 and −0.60 for SW and SE paths, respectively. Time-integrated energy (not shown) is also strongly correlated with pulse times spread, with correlation coefficients of −0.50 and −0.58 for SW and SE paths, respectively.

shows that there is considerable variability of the time duration of the arrival pattern. The time spread can be quantified by 90% quartile range of the pulse (Headrick *et al.*, 2000a). Plotting peak intensity against time spread, shown in Fig. 19, it is seen that the two time series are highly anticorrelated. That is, high peak intensity is associated with low time spread. Thus there are two processes competing to change arrival time spread; first there is the effect of higher peak intensity and $\mathcal{I}_\tau$, which raises more of the arrival pattern out of the noise, thus increasing the time spread, and second is a mode coupling effect which can redistribute the energy into higher slower modes, thereby increasing the time spread but decreasing the peak energy. Since the $\mathcal{I}_\tau$ variance is small, the second mechanism wins and thus the signal behaves primarily like a constant energy arrival with a phase-modulation (mode coupling)-induced time spread. This second effect has been noted by Headrick *et al.* (2000a) who showed that soliton scattering near the receiver can account for large time spread of individual modal arrivals. If the modes are statistically independent, then this mechanism will apply equally well for the full field variability described here.

V. DISCUSSION

Having quantified the observed intensity fluctuations we can discuss the implications of these results. In particular, the issue of connecting the acoustic scintillations to the well-measured PRIMER oceanography is taken up, and the range scaling of acoustic scintillation is examined using the numerical model.

A. Connection to oceanography: Intensity timescales

Frequency spectra of sound-speed fluctuations during the PRIMER experiment show a distinct evolution from a semi-diurnal tide dominance at the shelfbreak to a roughly equal partition of energy between tidal and buoyancy frequency contributions at the NW VLA [see Colosi *et al.* (2001) for a PRIMER spectrum]. Single-path weak fluctuation theory (Munk and Zachariasen, 1976; Flatté *et al.*, 1979) predicts that the spectrum of log-intensity and phase will simply be the ocean frequency spectrum divided by one power of the frequency when the scintillation index is less than about 0.3 and the PDF of amplitude is close to log-normal. This behavior can be termed a ''mirror'' effect because the acoustic spectrum ''mirrors'' the ocean spectrum within one factor of frequency. But, the PRIMER observa-

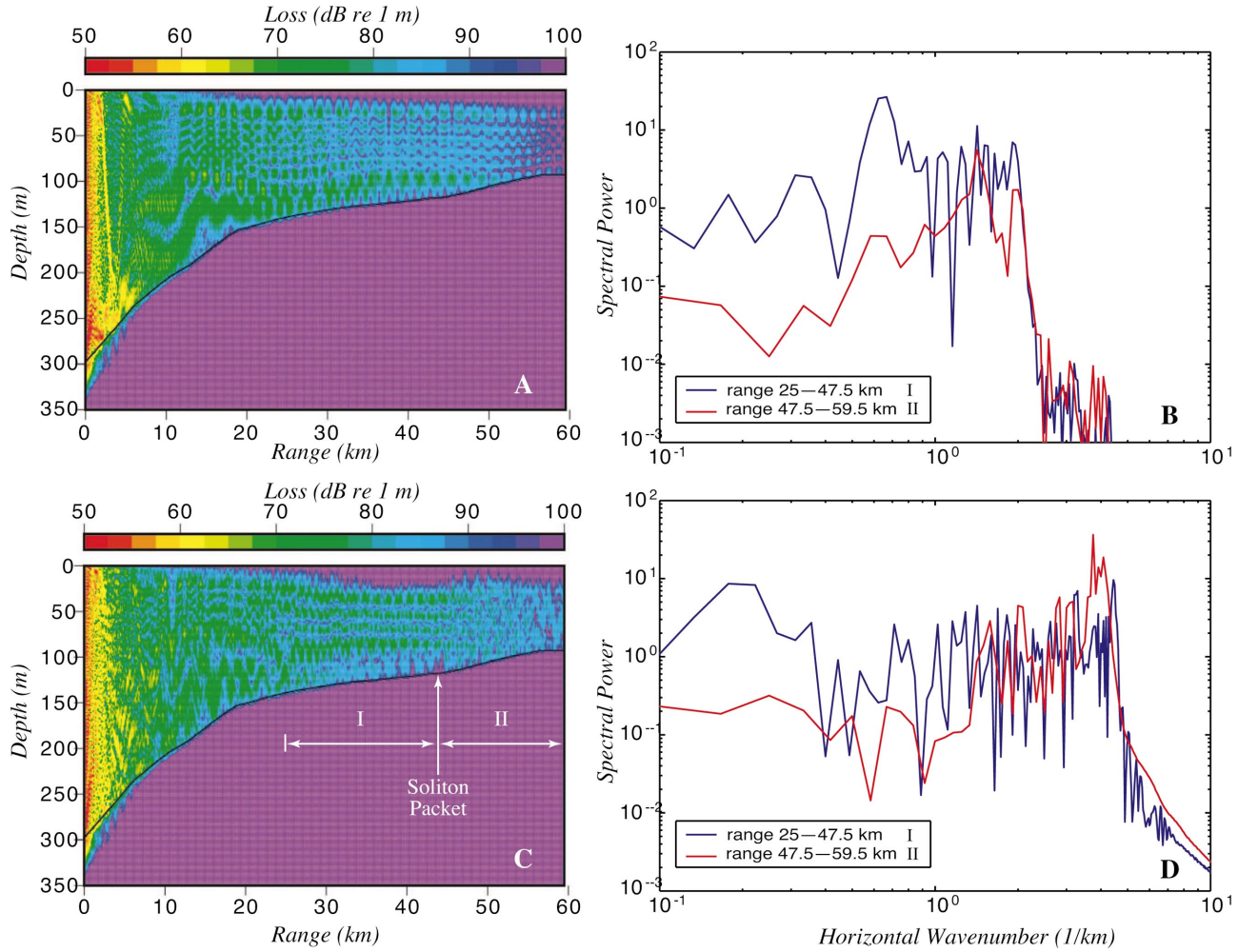


FIG. 20. Simulated transmission loss (TL) for the SW path both with (c) and without (a) internal wave sound speed perturbations. In the right-hand panels (b) and (d) are horizontal wave number spectra of the TL data to the left over two range intervals 25–47.5 km and 47.5–59.5 km which are before and after the internal wave packet at roughly 40-km range. The simulation utilized the sound speed environment from Fig. 2, and the range intervals denote regions before and during acoustic interaction with an internal wave soliton packet.

tions of point intensity have scintillation index near 1 and they are dominated by variability at time scales less than 2 h (close to the buoyancy frequency). This deemphasis of low frequency sound-speed variability was also observed in the Cobb Seamount experiment and the Mid-Ocean Acoustic Transmission experiment (MATE) (Ewart, 1976; Flatté *et al.*, 1979; Ewart and Reynolds, 1984), both of which were high frequency (i.e., kHz), short-range, deep-water experiments with scintillation index significantly above 0.3. Once significant scattering and interference occur, the weak fluctuation ''mirror'' effect of the ocean spectrum dissolves and significant distortion of the intensity frequency spectrum can occur (Macaskill and Ewart, 1996). Further, with strong interference, the spectrum of intensity is dominated by fadeouts whose duration and interval depend critically on the mean square time derivative of phase or simply the mean square phase rate ν^2 (Dyson *et al.*, 1976). In fact, the spectrum of log-intensity is roughly peaked at the frequency ν. Therefore our results are consistent with the physical picture of strong interference leading to intensity variability dominated by fades. The statistics of fades leads to an emphasis of the acoustic variability at higher frequencies.

This effect can be illustrated using our numerical model to demonstrate the increased complexity of the interference pattern when internal waves are added to the observed large scale sound speed fields. For the seven days of the experiment in which we obtained large scale sound speed fields we computed the acoustic fields at 400 Hz both with and without an added perturbation from the internal wave model; an example of the acoustic fields so obtained is shown in Fig. 20. A distinct difference is seen between the calculations with and without internal waves. The calculation without internal waves shows large spatial scale modulations of the interference pattern, but the calculation with internal waves shows significantly more interference in which the field becomes more spatially homogeneous with less large scale changes and higher small scale granularity. This visual picture is confirmed by examining the spatial spectra of a 10-km-long portion of the transmission loss curve on the shelf at 50-m water depth (Fig. 20).

B. Connection to oceanography: Integrated energy

Another matter that requires some extra consideration is the observed and modeled result of small, at most 2 dB rms, fluctuation in the total integrated energy, $\mathcal{I}_{z\tau}$; interestingly,

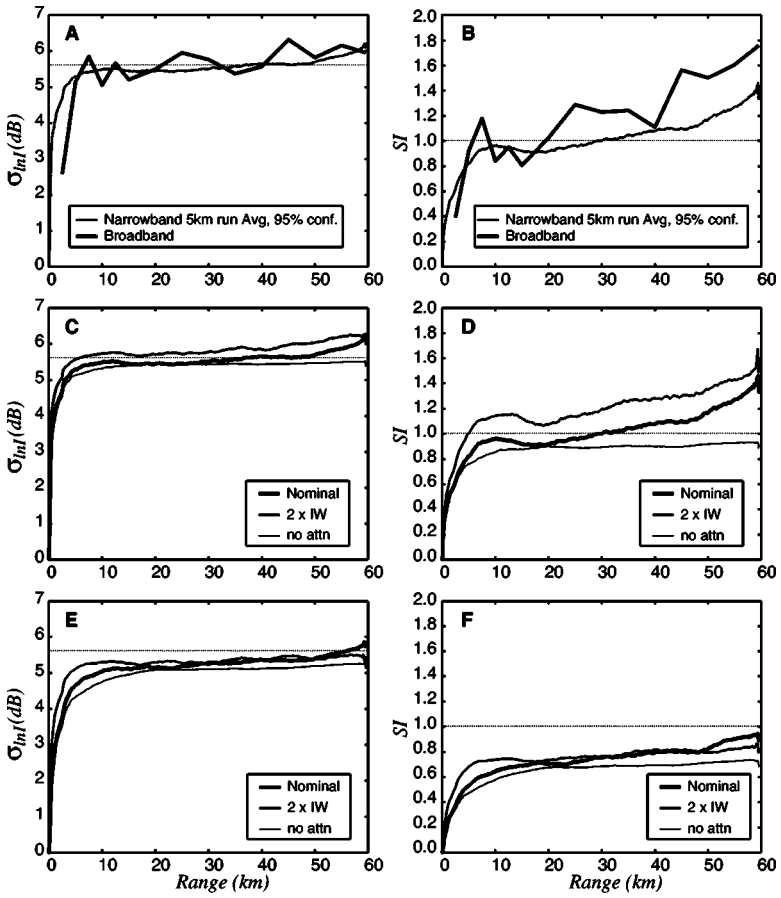


FIG. 21. Scintillation index SI and $\sigma_{\ln I}$ as a function of propagation range for the SW path computed from numerical simulation. In panels (a) and (b) broadband (thick-line) and narrow-band (thin-line) simulations are compared for the total variation (no filtering). In panels (c)–(f) only narrow-band simulations are shown, with panels (c),(d), and (e),(f) showing total variation and high frequency variation, respectively. Three different simulations are presented in panels (c)–(f): thick-line (nominal internal wave energy and bottom attenuation), medium-line (four times internal wave energy, and nominal bottom attenuation), and thin-line (nominal internal wave energy and no bottom attenuation).

similar results have been reported by Duda *et al.* (2004) for long range transmissions in the significantly different geographic location of the South China Sea. Further, it was observed that roughly equal contributions were partitioned between low and high frequencies. The PRIMER area showed intense oceanographic variability, with mesoscale, internal tide, and soliton changes in sound speed of order 15–25, 10–15, and 5–10 m/s, respectively, and the acoustic model of the present study, as well as other models (Sperry, 1999; Sperry *et al.*, 2003), show significant geoacoustic interaction. Therefore, it comes as some surprise that the total energy detected on the array is so stable, especially since the array does not span the entire water column. Observations by Zhou *et al.* (1991), in the Yellow Sea show long duration changes in mean signal level, and numerical simulations by Duda and Preisig (1999) demonstrate 12-dB peak-to-peak (4.2 dB rms) variations from soliton wave packet propagation. We conjecture that the PRIMER result (both observed and modeled) is due to the fact that there are multiple acoustic interactions between the water column inhomogeneities and the sea bed geology. For PRIMER the acoustic cycle distance is roughly 1–2 km (depending on water depth), which gives 20–30 upper (water column interaction) and lower turning (geoacoustic interaction) points. The notion is that oceanographic fluctuations may only change a few of these interactions, thus leaving the sum interaction relatively stable. The situation just described differs markedly from the cases of Zhou and Duda/Preisig as in those cases only a few total oceanographic interactions occur. For the PRIMER experiment, bottom attenuation is a strong function of mode number, and thus it is not the case that the weak variations in received energy are due to weak mode-number-dependent attenuation.

C. Approach to saturation

Table V shows that the observed and modeled intensity variability is very close to the fully saturated regime, and therefore the question arises as to how rapidly this near saturation is established in range. Given the relatively good agreement between our model and the observations, this question can be explored using the model. Simulations for single-frequency (CW) and broadband propagation over the SW path give SI and $\sigma_{\ln I}$ as a function of range as shown in the upper panels of Fig. 21; the relatively good agreement between CW and broadband calculations allows us to ignore the bandwidth and explore the PRIMER saturation behavior using the more computationally efficient CW model. The agreement between CW and broadband computations is not unexpected as the source Q value is of order 0.1 (see Secs. II D 2 and IV). In both cases, SI and $\sigma_{\ln I}$ are seen to rise rapidly in the first 10 km, followed by a more gentle increase over the remaining 50 km. Further, it is seen that the variation of log-intensity reaches its near saturation value slightly earlier in range than does SI. This result demonstrates that the fading mechanism which dominated log-intensity statistics is established earlier in range than the focusing mechanism which dominates the linear statistics like SI.

In the middle and lower panels of Fig. 21, SI and $\sigma_{\ln I}$ are plotted as a function of range for the three simulation cases of (1) nominal internal wave energy and attenuation

($\zeta_0 = 10$ m, $\beta = 0.5$ dB/m), (2) nominal internal wave energy and no attenuation ($\zeta_0 = 10$ m, $\beta = 0$ dB/m), and (3) four times the internal wave energy and nominal attenuation ($\zeta_0 = 20$ m, $\beta = 0.5$ dB/m). The middle panels show total variation while the bottom panels show the high frequency component. Starting with the middle panels, SI and $\sigma_{\ln I}$ for the lossy cases are seen to rise above saturation values of 1 and 5.6 dB and continue to rise with increasing range, while the no-attenuation case shows no rise above 1 but only slow approach from below. The difference in saturation behavior between lossy and attenuation free wave guides has been examined theoretically by Creamer (1996), and numerically by Teilburger *et al.* (1997) who showed that SI can grow indefinitely with range (supersaturation); our results are consistent with this picture, though the growth appears to be quite slow. Our numerical results show that the supersaturation effect is actually quite general and can be obtained in an environment in which the ''random'' component of the internal wave field is quite weak in comparison with the nonlinear field (we choose a factor of 10 difference in energy). Creamer (1996), for example, examined the supersaturation problem only for a fully random internal wave field, while Teilburger *et al.* (1997) did numerical calculations using a strong random component and a weak soliton field.

Turning to the comparisons of the cases which vary the overall internal wave energy (both random and nonlinear), it is seen that aside from a more rapid initial growth of intensity variability, doubling the perturbation strength does not dramatically change the scintillation behavior at longer ranges.

Next the high frequency variability is addressed in the lower panels of Fig. 21. Here it is seen that the cases with and without bottom attenuation are very similar, which shows that the supersaturation effect seen in the middle panels is primarily due to low frequency variability. This low frequency attribution of the supersaturation effect is very interesting and warrants further study. Finally, as seen in the middle panels, doubling the perturbation strength does not dramatically change the scintillation behavior at longer ranges.

These simulation results have important implications for the understanding of wave scattering regimes as depicted by the regimes of unsaturated, partially saturated, and fully saturated propagation (Flatté *et al.*, 1979). They show strikingly that the approach to saturation (or supersaturation) in this shallow water environment is a process of rapid initial growth of scintillation at very short ranges followed by surprisingly slow growth of scintillation up to and beyond the saturation values of SI=1 and $\sigma_{\ln I} = 5.6$ dB. In particular, in spite of the relatively high frequency of 400 Hz, there appears to be no significant partially saturated regime where the acoustic field is dominated by strong focusing caustics which lead to SI values significantly above 1 at short range. In partial saturation SI relaxes to 1 from above which is clearly not the case. The approach to saturation from below 1 also has important implications for the kinematics of the multipath. Colosi and Baggeroer (2004) have shown that the single frequency kinematical condition to get an approach to saturation from above/(below) 1 is $\langle a^4 \rangle / \langle a^2 \rangle^2 > (<) 2$. Thus the multipath amplitude variability in the PRIMER environment is in fact quite weak. The frequency dependence of the saturation process is an interesting issue, but is beyond the scope of this paper.

VI. SUMMARY

Observations of acoustic intensity fluctuations measured by a 52 m VLA for 400-Hz, broadband, long range (42.2 and 59.6 km) propagation in shallow water have been described. Total detected energy at the VLA (obtained by integrating acoustic intensity over the VLA aperture and over the pulse arrival) shows variability of roughly 2 dB, which is small indeed. This demonstrates the surprising result that lossy acoustic bottom interactions were not significantly modulated by the strong observed mesoscale and internal wave variability. Here, since differential modal attenuation was large, we have argued the case that the PRIMER experiment has a sufficiently large number of water column and geoacoustic interactions that water column modulation will be averaged out. Analysis of temporally integrated energy across the arrival structure shows somewhat more variability than the total detected energy (depth and time integrated). This result demonstrates that there is some vertical redistribution of energy occuring, but it is at the fraction-of-a-dB level (see Tables III and IV). Both integrated energy statistics show PDFs close to log-normal.

Regarding statistics of point intensities, it is found that acoustic variability with time scales less than 2 h show fluctuations typical of a fully saturated wave field where SI is close to 1, the $\sigma_{\ln I}$ is close to 5.6 dB rms, and the PDF of intensity is close to exponential; these statistics were found at both the 42.2- and 59.6-km ranges. In addition to this high frequency variability was a low frequency modulation which accounted for roughly 2.5-dB rms change. The observations, therefore, are consistent with the physical picture of Dyson *et al.* (1976) in which the statistics are dominated by fades at short time scales and low frequency ocean modulation at long time scales.

Peak intensity statistics are seen to be highly correlated to temporally integrated energy which shows, not surprisingly, that higher peak intensities are expected for a stronger overall arrival. Further, it was shown that peak intensity (and thus temporally integrated energy) is anticorrelated with the overall pulse time spread so that a small peak intensity is obtained when the arrival structure is spread over a longer duration (and vice versa). The primary mechanism for this behavior appears to be mode-coupling-induced time spread near the receiver, and thus the signal behaves like a constant energy arrival in which mode-coupling-induced phase modulation spreads the arrival out in time. Peak intensity PDFs, like those of integrated energy, show near log-normal forms.

A parabolic equation numerical model has been used with observed large scale sound speed fields, and a numerical quasi-stochastic nonlinear (KdV) internal tide model to obtain estimates of several of the observed quantities. The comparisons between simulation and observation are very encouraging with regard to the scintillation statistics (see Tables III–V), however much more work will be necessary to match the mean pulse shapes.

The numerical model has allowed an examination of the range dependence of scintillation. It is found that intensity variability grows rapidly in the first 10 km of propagation, and thereafter the remaining 50 km of propagation shows slow growth of intensity variance up to and beyond saturation values (supersaturation). This supersaturation behavior is not unexpected in lossy wave guides (Creamer, 1996; Teilburger *et al.*, 1997). Breaking the model variability into low and high frequency components and comparing to an attenuation free calculation it is found that at time scales less than 2 h the lossy and nonlossy scintillations behave very similarly and the approach to saturation is from below 1, with no partially saturated or strong focusing regime. This result indicates that the supersaturation effect comes from the low frequency component of the scintillation behavior; this effect is not fully understood. Further, the result that the approach to saturation is from below 1 suggests that multipath amplitude variability is quite small.

ACKNOWLEDGMENTS

We acknowledge the able seamanship of the captain and crew of the *R/V Endeavor* who helped make these measurements possible. We also thank John Kemp and Keith von der Heydt for their deck-skills and engineering expertise. We have benefited from several fruitful conversations with Tim Duda. This is Woods Hole Oceanographic Institution Contribution No. 11065.

[1]In Eq. (10), there is no connection to ''stepwise'' coupled mode propagation. For coupled modes, the field ψ is a complex vector of mode amplitudes and F_j are mode coupling matrices. In the limit of multiple coupling the complex mode amplitudes and the complex field become uncorrelated Gaussian random variables (Dozier and Tappert, 1978; Colosi and Flatté, 1996).

[2]This definition of Q is not standard, as the reciprocal relation is most commonly used.

Apel, J. R., Ostrovsky, L. A., and Stepanyants, Y. A. (**1995**). ''Internal solitons in the ocean,'' Report MERCJRA0695, Applied Physics Laboratory, Johns Hopkins University.

Collins, M. D. (**1993**). ''A split step padé solution for the parabolic equation method,'' J. Acoust. Soc. Am. **93**, 1736–1842.

Collins, M. D. (**1997**). *A User's Guide for RAM* (Naval Research Laboratory, Washington, DC).

Colosi, J. A., and Baggeroer, A. B. (**2004**). ''On the kinematics of broadband multipath scintillation and the approach to saturation,'' J. Acoust. Soc. Am. **116**(6), 3515-3522.

Colosi, J. A., and Flatté, S. M. (**1996**). ''Mode coupling by internal waves for multimegameter acoustic propagation in the ocean,'' J. Acoust. Soc. Am. **100**, 3607–3620.

Colosi, J. A., Beardsley, R. C., Lynch, J. F., Gawarkiewicz, G., Chiu, C.-S., and Scotti, A. (**2001**). ''Observations of nonlinear internal waves on the outer New England continental shelf during the summer Shelfbreak PRIMER study,'' J. Geophys. Res. **106**, 9587–9601.

Creamer, D. (**1996**). ''Scintillating shallow water waveguides,'' J. Acoust. Soc. Am. **99**, 2825–2838.

Dozier, L. B., and Tappert, F. D. (**1978**). ''Statistics of normal mode amplitudes in a random ocean. I. Theory,'' J. Acoust. Soc. Am. **63**, 353–365.

Duda, T. F., and Farmer, D. M. (eds.) (**1999**). ''The 1998 WHOI/IOS/ONR Internal Solitary Wave Workshop: Contributed Papers,'' WHOI Technical Report, WHOI-99-07.

Duda, T. F., and Preisig, J. C. (**1999**). ''A modeling study of acoustic propagation through moving shallow-water solitary wave packets,'' IEEE J. Ocean. Eng. **24**, 16–32.

Duda, T. F., Lynch, J. F., Newhall, A. E., Liu, W., and Chiu, C. S. (**2004**). ''Fluctuations of 400 Hz sound intensity in the 2001 ASIAEX South China Sea experiment,'' J. Oceanic Eng. (in press).

Dyer, I. (**1970**). ''Statistics of sound propagation in the ocean,'' J. Acoust. Soc. Am. **48**, 337–345.

Dyson, F., Munk, W., and Zetler, B. (**1976**). ''Interpretation of multipath scintillations Eleuthera to Bermuda in terms of internal waves and tides,'' J. Acoust. Soc. Am. **59**, 1121–1133.

Ewart, T. (**1976**). ''Acoustic fluctuations in the opon ocean—A measurement using a fixed refracted path,'' J. Acoust. Soc. Am. **60**, 46–59.

Ewart, T., and Reynolds, S. (**1984**). ''The mid-ocean acoustic transmission experiment,'' J. Acoust. Soc. Am. **75**, 785–802.

Flatté, S., Dashen, R., Munk, W., Watson, K., and Zachariasen, F. (**1979**). *Sound Transmission through a Fluctuating Ocean* (Cambridge U.P., Cambridge).

Gawarkiewicz, G., Brink, K. H., Bahr, F., Beardsley, R. C., Caruso, M., Lynch, J. F., and Chiu, C. S. (**2004**). ''A large-amplitude meander of the shelfbreak front during summer south of New England: Observations from the Shelfbreak PRIMER experiment,'' J. Geophys. Res. **109**.

Headrick, R., Lynch, J., and the SWARM Group (**2000a**). ''Acoustic normal mode fluctuation statistics in the 1995 SWARM internal wave scattering experiment,'' J. Acoust. Soc. Am. **107**, 201–220.

Headrick, R., Lynch, J., and the SWARM Group (**2000b**). ''Modeling mode arrivals in the 1995 SWARM experiment acoustic transmissions,'' J. Acoust. Soc. Am. **107**, 221–236.

Liu, A. K. (**1988**). ''Analysis of nonlinear internal waves in the New York bight,'' J. Geophys. Res. **93**, 12312–12329.

Lynch, J., Newhall, A. E., Sperry, B., Gawarkiewicz, G., Tyack, P., Chiu, C.-S., and Abbot, P. (**2003**). ''Spatial and temporal variations in acoustic propagation characteristics at the New England shelfbreak front,'' IEEE J. Ocean. Eng. **28**, 129–150.

Macaskill, C., and Ewart, T. (**1996**). ''Numerical solution of the fourth moment equation for acoustic intensity correlations and comparison with the mid-ocean acoustic transmission experiment,'' J. Acoust. Soc. Am. **99**, 1419–1429.

Munk, W. H., and Zachariasen, F. (**1976**). ''Sound propagation through a fluctuating stratified ocean: Theory and observation,'' J. Acoust. Soc. Am. **59**, 818–838.

Pace, N. G., and Jensen, F. B. (eds.) (**2003**). *Impact of Littoral Environmental Variability on Acoustic Predictions and Sonar Performance* (Kluwer Academic, Dordrecht).

Pasewark, B., Wolf, S., Orr, M., and Lynch, J. (**2002**). ''Acoustic intensity variability in a shallow water environment,'' Acoustic Variability 2002, SACLANT Undersea Research Center, pp. 11–19.

Potty, G. R., Miller, J. H., and Lynch, J. F. (**2003**). ''Inversion for sediment geoacoustic properties at the New England Bight,'' J. Acoust. Soc. Am. **114**, 1874–1887.

Potty, G. R., Miller, J. H., Lynch, J. F., and Smith, K. B. (**2000**). ''Tomographic inversion for sediment parameters in shallow water,'' J. Acoust. Soc. Am. **108**, 973–986.

Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (**1992**). *Numerical Recipies in Fortran: The art of scientific computing (Second Edition)* (Cambridge U.P., Cambridge).

Shearman, R. K., and Lentz, S. J. (**2004**). ''Observations of tidal variability on the New England shelf,'' J. Geophys. Res. **109**(C6).

Sperry, B. (**1999**). ''Analysis of acoustic propagation characteristics at the New England continental shelfbreak,'' MIT/WHOI Joint Program PhD. thesis.

Sperry, B., Lynch, J., Gawarkiewicz, G., Chiu, C., and Newhall, A. (**2003**). ''Characteristics of acoustic propagation to the eastern vertical line receiver during the summer 1996 New England shelfbreak PRIMER experiment,'' IEEE J. Ocean. Eng. **28**, 729–749.

Teilburger, D., Finette, S., and Wolf, S. (**1997**). ''Acoustic propagation through an internal wave field in a shallow water waveguide,'' J. Acoust. Soc. Am. **101**, 789–808.

Zhou, J. X., Zhang, X. S., and Rogers, P. (**1991**). ''Resonant interaction of sound waves with internal solitons in the coastal zone,'' J. Acoust. Soc. Am. **90**, 2042–2054.

Numerical study of broadband sound pulse propagation in three-dimensional oceanic waveguides

Frédéric Sturm[a)]
Laboratoire de Mécanique des Fluides et d'Acoustique, UMR CNRS 5509, École Centrale de Lyon, 36 Avenue Guy de Collongue, FR-69134 Ecully Cedex, France

(Received 22 July 2004; revised 14 December 2004; accepted 15 December 2004)

In this paper, the propagation of a broadband sound pulse in three-dimensional (3D) shallow water waveguides is investigated numerically. Two cases are examined: (i) the 3D ASA benchmark wedge, and (ii) the 3D Gaussian canyon. The numerical method used to solve the four-dimensional acoustic problem is based on a Fourier synthesis technique. The frequency-domain calculations are carried out using the fully 3D parabolic equation based model 3DWAPE, recently modified to include a wide-angle paraxial approximation for the azimuthal component. A broadband sound pulse with a central frequency of 25 Hz and a bandwith of 40 Hz is considered. For both test cases, 3D results corresponding to a 25 Hz cw point source are first presented and compared with predictions from a 3D adiabatic modal model. Then, the acoustic problem is solved considering the broadband source pulse. The modal structure of the received signals is analyzed and exhibits multiple mode arrivals of the propagating signal. [DOI: 10.1121/1.1855791]

PACS numbers: 43.30.Bp, 43.30.Dr, 43.30.Gv [AIT] Pages: 1058–1079

I. INTRODUCTION

It has been demonstrated both experimentally[1–3] and numerically[4,5] that in some particular oceanic environments, the horizontal refraction of propagating sound waves cannot be neglected and leads, far from the source, to significant three-dimensional (3D) effects. Fully three-dimensional models are needed to predict such 3D effects. Note that in this work, a model is referred to as 3D if it allows horizontal refraction to be considered. Otherwise, the model is said to be two-dimensional (2D), $N\times 2$D, or pseudo-3D. Among existing 3D codes available in the underwater acoustics community,[6–10] parabolic equation (PE) based models are largely used since they are efficient for solving complex sound propagation problems in various oceanic environments. The reader is referred to Ref. 11 for an exhaustive review of the 3D codes based on modal theory, parabolic equation, rays, and hybrid models, and in particular to Ref. 12 for a specialized review of existing 3D PE models. The main drawback of 3D models in cylindrical coordinates (which is the case for most 3D PE models) is that they are computationally expensive. Indeed, a 3D code is at least two orders of magnitude slower than any $N\times 2$D code since (i) a very large number of points is required in the azimuthal direction to maintain the necessary arclength between adjacent bearing angles, and (ii) a differential operator with respect to the azimuthal coordinate must be incorporated into the solution. However, rising computer performance as well as the development of efficient numerical techniques[9,13] allows 3D PE models to be treated at a reasonable computational cost. A calculation in four-dimensional (4D) (i.e., three spatial dimensions and time) using Fourier synthesis methods is thus now possible (at least at low frequencies), allowing the important question of broadband signal dispersion in general 3D waveguides to be addressed.

The aim of this work is to study the propagation of broadband sound pulses in three-dimensional shallow water waveguides. The numerical method used to solve the relevant 4D acoustic problems is based on Fourier synthesis of frequency-domain solutions. The calculations in 3D are done using the 3DWAPE code based on a fully three-dimensional parabolic equation.[13] Two 3D acoustic problems are treated in this paper. They both consist of an isovelocity water layer overlying a lossy, homogeneous, half-space sedimental layer. They only differ in the description of the bottom geometry. The first acoustic problem considered is a 3D benchmark wedge based on a three-dimensional extension of the original two-dimensional ASA wedge configuration.[14] Preliminary results of broadband sound pulse propagation obtained using 3DWAPE for this specific 3D wedge problem can be found in Ref. 15. They showed that the modal structures of the propagating pulses calculated using a 3D PE model were qualitatively consistent with previous results obtained by Westwood[16] for a similar case using an analytical method. Though these results were satisfying and encouraging, no quantitative comparison with any exact solution was made. Note that in its original configuration, the 3DWAPE code had a very-wide-angle capability along the vertical direction but only a narrow-angle capability along the azimuthal direction. It has recently been modified to handle higher-order paraxial approximations along the azimuthal direction. The results of Ref. 15 were obtained using the azimuthal narrow-angle version of the code. In the present paper, the issue of using wide-angle approximation in azimuth is addressed. In particular, numerical solutions obtained using various paraxial approximations in azimuth are compared with a reference analytical solution based on the image method. Broadband calculations are now carried out using the azimuthal wide-angle version of the code. The second problem considered is the 3D Gaussian canyon test case. It corresponds to a variant of the original 3D Gaussian canyon test case created for the

a)Electronic mail: frederic.sturm@ec-lyon.fr

SWAM'99 Shallow Water Acoustic Modeling Workshop.[17] Results for the original SWAM'99 test case were obtained for a 25 Hz cw point source.[7,18] Modifications have been made such that only three propagating modes exist at the source position (as in the 3D benchmark wedge) which makes the modal structure analysis of the received signals easier. Note that in its original form, eleven propagating modes were present. The problem involving a broadband sound pulse is analyzed with the same methodology as for the 3D wedge problem.

The organization of this paper is as follows: In Sec. II, the Fourier synthesis technique is briefly summarized, and the 3DWAPE model is presented. Its wide-angle azimuthal capability is discussed; a review of the existing 3D PE code is also given. In Sec. III, the two acoustic shallow water problems are described. For both test cases, a broadband source pulse with a central frequency of 25 Hz and a bandwidth of 40 Hz is considered. The 3D ASA benchmark problem is studied in Sec. IV. First, the 3D results corresponding to a 25 Hz cw point source are presented, analyzed, and compared with other solutions. Then, the acoustic problem is solved in 4D considering the broadband source pulse. The signals received by a set of vertical arrays placed in the cross-slope direction are analyzed. Section V deals with the 3D Gaussian canyon test case. Again, results corresponding to both cw calculations and time series are presented and discussed. No reference solution is available for this test case. The received signals on two distinct vertical arrays placed along the canyon axis are then analyzed. In Sec. VI, a summary of the results of this work is provided and future improvements are suggested.

II. MATHEMATICAL MODELING

A. Fourier synthesis method

A multilayered waveguide composed of one water layer overlying one (or several) fluid sedimental layer(s) is considered. The model for each layer is three-dimensional. Cylindrical coordinates are used, where z is the depth (increasing downwards) below the ocean surface, θ is the azimuthal (bearing) angle, and r is the horizontal range, related to the Cartesian coordinates by $x=r\cos\theta$ and $y=r\sin\theta$. An isotropic, broadband point source, S, is located at $r=0$ and $z=z_S$. The acoustic wave equation,

$$\rho\nabla\cdot\left(\frac{1}{\rho}\nabla P\right)-\frac{1}{c^2}\frac{\partial^2 P}{\partial t^2}=-\mathcal{S}(t)\frac{2\,\delta(z-z_S)\,\delta(r)}{r}, \tag{1}$$

is solved, where $P=P(r,\theta,z;t)$ is the acoustic pressure as a function of the three spatial variables r, θ, z, and time t, and $\mathcal{S}(t)$ is the time-dependence of point source $\mathcal{S}$. In Eq. (1) c and ρ represent, respectively, the varying (in space) sound speed and the density (constant within each layer). Let $\hat{P}$ denote the Fourier transform of the time-domain acoustic pressure P, defined by

$$\hat{P}(r,\theta,z;\omega)=\int_{-\infty}^{+\infty}P(r,\theta,z;t)e^{i\omega t}dt, \tag{2}$$

where $\omega=2\pi f$ is the angular frequency (expressed in rad/s) and f is the frequency (expressed in Hz). Then, Fourier transform of Eq. (1) leads to the frequency-domain (or Helmholtz) equation

$$\rho\nabla\cdot\left(\frac{1}{\rho}\nabla\hat{P}\right)+k_\alpha^2\hat{P}=-\hat{\mathcal{S}}(\omega)\frac{2\,\delta(z-z_S)\,\delta(r)}{r}, \tag{3}$$

where $k_\alpha=k(1+\mathrm{i}\eta\alpha)$ is the complex (to account for lossy layers) wave number, with $k=\omega/c$, α is the attenuation expressed in decibels per wavelength, $\eta=1/(40\pi\log_{10}e)$ (with $\eta\alpha\ll1$), and $\hat{\mathcal{S}}(\omega)$ is the source spectrum defined by

$$\hat{\mathcal{S}}(\omega)=\int_{-\infty}^{+\infty}\mathcal{S}(t)e^{i\omega t}dt. \tag{4}$$

The angular frequency ω is treated as a parameter in Eq. (3). The complex-valued field $\hat{P}=\hat{P}(r,\theta,z;\omega)$ is sought as a function of the spatial variables r, θ, and z, for selected (nonnegative) discrete frequencies within the frequency band of interest. The acoustic pressure $\hat{P}$ is assumed to satisfy a pressure-release condition on the ocean surface, an outgoing radiation condition at infinity (in both range and depth), a 2π-periodicity condition in the azimuthal direction, and appropriate transmission conditions at each sedimental interface. The frequency-domain solution, $\hat{P}$, is then transformed to the time-domain using the following inverse Fourier transform:

$$P(r,\theta,z;t)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\hat{P}(r,\theta,z;\omega)e^{-i\omega t}d\omega, \tag{5}$$

where $\hat{P}(r,\theta,z;-\omega)=\overline{\hat{P}(r,\theta,z;\omega)}$ so that the real-valued time-domain acoustic pressure P satisfies the initial time-dependent wave equation (1). In summary, solving a pulse propagation problem with the Fourier synthesis approach[19,20] requires one to (i) decompose the source pulse using a Fourier transform, (ii) select a frequency spacing and solve the 3D propagation problem for each discrete frequency within a frequency-band of interest, and (iii) perform inverse Fourier transforms of the frequency-domain solutions to obtain the time signal at any given receiver. As in Ref. 21, step (ii) is achieved by using a three-dimensional parabolic equation based model. Note that an alternative to the Fourier synthesis approach would be to solve the pulse propagation problem directly in the time-domain.[19,22–26] In particular, time-domain methods related to various PE formulations can be found in Refs. 23–26.

B. Three-dimensional parabolic equation

The acoustic problem is solved in the frequency-domain using a parabolic equation (PE) approach. Dropping the source spectrum from Eq. (3) yields

$$\rho\nabla\cdot\left(\frac{1}{\rho}\nabla\hat{P}_{\mathrm{norm}}\right)+k_\alpha^2\hat{P}_{\mathrm{norm}}=-\frac{2\,\delta(z-z_S)\,\delta(r)}{r}, \tag{6}$$

where the unknown is now the normalized acoustic pressure $\hat{P}_{\mathrm{norm}}(r,\theta,z;\omega)$. Cylindrical spreading is handled by expressing $\hat{P}_{\mathrm{norm}}(r,\theta,z;\omega)$ as

$$\hat{P}_{\text{norm}}(r,\theta,z;\omega)=H_0^{(1)}(k_0 r)\,\psi(r,\theta,z;\omega),$$

where $H_0^{(1)}$ denotes the zeroth-order Hankel function of the first kind and $k_0=\omega/c_{\text{ref}}$ with c_{ref} a reference sound speed. Assuming that r^{-2} approximately commutes with $\partial/\partial r$ for $r\gg 0$, the left-hand side of Eq. (6) can be factorized, and assuming only outward propagation in range, the 3D problem based on (elliptic-type) Eq. (6) is reduced to an initial- and boundary-value problem. For any given value of ω, a complex function $\psi=\psi(r,\theta,z;\omega)$ is sought, which represents the acoustic field for $r_0\leqslant r\leqslant r_{\max}$, $0\leqslant\theta\leqslant 2\pi$ and $0\leqslant z\leqslant z_{\max}$, and which satisfies

$$\frac{\partial\psi}{\partial r}=ik_0(\sqrt{\mathcal{I}+\mathcal{X}+\mathcal{Y}}-\mathcal{I})\,\psi \tag{7}$$

and $\psi(r=r_0,\theta,z;\omega)=\psi^{(0)}(\theta,z;\omega)$. Here, $\psi^{(0)}$ denotes the initial outgoing field at $r=r_0$, $\mathcal{I}$ is the identity operator, $\mathcal{X}$ is the 2D depth operator in the rz plane, and $\mathcal{Y}$ is the azimuthal operator, defined as

$$\mathcal{X}=(n_\alpha^2(r,\theta,z)-1)\mathcal{I}+\frac{\rho}{k_0^2}\frac{\partial}{\partial z}\left(\frac{1}{\rho}\frac{\partial}{\partial z}\right),$$

$$\mathcal{Y}=\frac{1}{(k_0 r)^2}\frac{\partial^2}{\partial\theta^2},$$

where $n_\alpha(r,\theta,z)=(c_{\text{ref}}/c(r,\theta,z))(1+i\,\eta\alpha)$ is the complex index of refraction. In order to prevent spurious reflections from a pressure-release imposed boundary condition at $z_{\max}$, an increasing attenuation coefficient is introduced in the lower part of the domain. The operator $\mathcal{Y}$ handles the azimuthal diffraction term. Neglecting $\mathcal{Y}$ in Eq. (7) but retaining azimuthal dependence in $n_\alpha(r,\theta,z)$ would lead to an $N\times 2$D or pseudo-3D (i.e., azimuthally uncoupled) PE model which could not predict horizontal refraction. The square-root operator present in Eq. (7) is then approximated using a higher-order Padé approximation along z and a linear approximation along θ:

$$\sqrt{\mathcal{I}+\mathcal{X}+\mathcal{Y}}=\mathcal{I}+\sum_{k=1}^{n_p}\frac{a_{k,n_p}\mathcal{X}}{\mathcal{I}+b_{k,n_p}\mathcal{X}}+\frac{1}{2}\mathcal{Y}+O(\mathcal{X}^{2n_p+1},\mathcal{Y}^2,\mathcal{X}\mathcal{Y}), \tag{8}$$

where n_p is the number of terms, and a_{k,n_p}, b_{k,n_p}, $1\leqslant k\leqslant n_p$, are real coefficients given by[27]

$$a_{k,n_p}=\frac{2}{2n_p+1}\sin^2\left(\frac{k\pi}{2n_p+1}\right),\quad 1\leqslant k\leqslant n_p,$$
$$b_{k,n_p}=\cos^2\left(\frac{k\pi}{2n_p+1}\right),\quad 1\leqslant k\leqslant n_p. \tag{9}$$

Complex coefficients can be used to attenuate Gibb's oscillations.[28] The Padé series expansion is very convenient since it allows for a very-wide-angle propagation along z, the angular limitation depending on parameter n_p. It is thus able to model energy at vertical angles approaching $\pm 90°$ with respect to the horizontal. The linear approximation allows only for narrow-angle propagation along θ. Substitution of Eq. (8) into Eq. (7), and neglection of the term in $O(\mathcal{X}^{2n_p+1},\mathcal{Y}^2,\mathcal{X}\mathcal{Y})$ leads to

$$\frac{\partial\psi}{\partial r}=ik_0\left(\sum_{k=1}^{n_p}\frac{a_{k,n_p}\mathcal{X}}{\mathcal{I}+b_{k,n_p}\mathcal{X}}+\frac{1}{2}\mathcal{Y}\right)\psi. \tag{10}$$

This equation accounts for refraction effects which are greater along z than along θ. It has been implemented in the research code 3DWAPE.[13] This model has a very-wide-angle capability in depth, and a narrow-angle capability in azimuth. The original 3D PE proposed by Tappert[29] can be obtained by expanding the square-root in a Taylor series and retaining only the linear terms in $\mathcal{X}$ and $\mathcal{Y}$,

$$\sqrt{\mathcal{I}+\mathcal{X}+\mathcal{Y}}=\mathcal{I}+\tfrac{1}{2}(\mathcal{X}+\mathcal{Y})+O(\mathcal{X}^2,\mathcal{Y}^2,\mathcal{X}\mathcal{Y}). \tag{11}$$

The resulting parabolic equation (known as the standard 3-D PE) thus has a narrow-angle capability in both depth and azimuth:

$$\frac{\partial\psi}{\partial r}=\frac{ik_0}{2}(\mathcal{X}+\mathcal{Y})\,\psi. \tag{12}$$

Note that when $n_p=1$ (which leads to the Claerbout's coefficients $a_{1,1}=1/2$ and $b_{1,1}=1/4$), Eq. (10) reduces to the three-dimensional parabolic equation used by Collins *et al.*[10] and by Fawcett,[8]

$$\frac{\partial\psi}{\partial r}=ik_0\left(\frac{\frac{1}{2}\mathcal{X}}{\mathcal{I}+\frac{1}{4}\mathcal{X}}+\frac{1}{2}\mathcal{Y}\right)\psi. \tag{13}$$

Since the higher-order terms neglected are in $O(\mathcal{X}^3,\mathcal{Y}^2,\mathcal{X}\mathcal{Y})$, this 3D PE has a (Claerbout) wide-angle capability in depth and a narrow-angle capability in azimuth. Instead of using a higher-order Padé approximation along z, Lee–Saad–Schultz use a Taylor series expansion along z and a linear approximation along θ:

$$\sqrt{\mathcal{I}+\mathcal{X}+\mathcal{Y}}=\mathcal{I}+\tfrac{1}{2}\mathcal{X}-\tfrac{1}{8}\mathcal{X}^2+\tfrac{1}{2}\mathcal{Y}+O(\mathcal{X}^3,\mathcal{Y}^2,\mathcal{X}\mathcal{Y}). \tag{14}$$

Neglecting the higher-order terms in $O(\mathcal{X}^3,\mathcal{Y}^2,\mathcal{X}\mathcal{Y})$, the resulting equation (referred to in the literature as the LSS-3D wide angle wave equation) is

$$\frac{\partial\psi}{\partial r}=ik_0\left(\frac{1}{2}\mathcal{X}-\frac{1}{8}\mathcal{X}^2+\frac{1}{2}\mathcal{Y}\right)\psi. \tag{15}$$

This 3D PE has been implemented by Botseas *et al.* in the research computer code FOR3D[30] and applied to realistic three-dimensional environments with bottom topographic variations and sound-speed profiles.[6]

Parabolic equations can be solved numerically using various techniques. For example, Smith[9] uses a marching algorithm based on the split-step Fourier technique in both depth and azimuth. In 3DWAPE, instead of the SSF algorithm, an alternating direction method is used. This technique is used in many 3D PE codes.[6–8,10,18] The alternating direction method consists in splitting Eq. (10) into the following system of equations:

$$(\mathcal{I}+b_{k,n_p}\mathcal{X})\frac{\partial\psi}{\partial r}(r,\theta,z;\omega)=ik_0 a_{k,n_p}\mathcal{X}\psi(r,\theta,z;\omega),\quad 1\leqslant k\leqslant n_p, \tag{16}$$

$$\frac{\partial \psi}{\partial r}(r,\theta,z;\omega)=\frac{ik_0}{2}\mathcal{Y}\psi(r,\theta,z;\omega), \tag{17}$$

and then these n_p+1 equations are solved sequentially at any discrete range using an implicit Crank-Nicolson scheme. Let Δr be the increment in range. Given the 3D field ψ at discrete range value r_n, ψ is obtained at the next discrete range value $r_n+\Delta r$ in two steps. Following the notations used in Ref. 7, the first step consists in computing n_p intermediate fields denoted $u^{(1)}(\theta,z)$, $u^{(2)}(\theta,z),\ldots,u^{(n_p)}(\theta,z)$ solving

$$\left(\mathcal{I}+\left(b_{k,n_p}-\frac{ik_0\Delta r}{2}a_{k,n_p}\right)\mathcal{X}\right)u^{(k)}(\theta,z)$$
$$=\left(\mathcal{I}+\left(b_{k,n_p}+\frac{ik_0\Delta r}{2}a_{k,n_p}\right)\mathcal{X}\right)u^{(k-1)}(\theta,z), \tag{18}$$

for $1\leqslant k\leqslant n_p$, where $u^{(0)}(\theta,z)$ denotes the 3D field ψ at range r_n: $u^{(0)}(\theta,z)=\psi(r_n,\theta,z;\omega)$. The second step consists in computing $\psi(r_n+\Delta r,\theta,z;\omega)$ from the last intermediate field $u^{(n_p)}(\theta,z)$ obtained in step 1, by solving

$$\left(\mathcal{I}-\frac{ik_0\Delta r}{4}\mathcal{Y}\right)\psi(r_n+\Delta r,\theta,z;\omega)$$
$$=\left(\mathcal{I}+\frac{ik_0\Delta r}{4}\mathcal{Y}\right)u^{(n_p)}(\theta,z). \tag{19}$$

The discretization of Eq. (18) for any $1\leqslant k\leqslant n_p$ is achieved using a piecewise-linear finite-element/Galerkin scheme. Let N and M denote the numbers of mesh points along z and θ, respectively. Solving Eq. (18) for the $N\times M$ points of the θz grid requires the inversion of n_p large algebraic linear systems of order $M\times N$. The matrix for each of the n_p systems has a block-diagonal structure. Each inversion is hence equivalent to the inversion of M (auxiliary) linear systems of order N. Since each block is a square tridiagonal matrix of order N, these inversions are performed using a fast and robust Gaussian (direct) algorithm optimized for tridiagonal matrices. The discretization of Eq. (19) is achieved using an efficient higher-order accurate finite difference (FD) scheme. The solution of Eq. (19) involves in this case the inversion of N linear systems of order M with entries in the upper right and lower left corners of the banded matrices to account for the 2π-periodicity condition along θ. The bandwidth of each block depends on the order of the centered FD formula used. Again, a fast and robust Gaussian algorithm optimized for banded matrices is used. Using a higher-order accurate FD scheme along θ allows a significant reduction of the azimuthal sampling, and faster computations (see the discussion in Ref. 13). Alternatively, Eq. (19) can also be solved using any Fourier-based transformation techniques (e.g., FFTs). The 3DWAPE code offers the possibility to use both FD- and FFT-based techniques for solving Eq. (19).

Note that all three-dimensional parabolic equations reviewed in this section make use of the following approximation:

$$\sqrt{\mathcal{I}+\mathcal{X}+\mathcal{Y}}=\sqrt{\mathcal{I}+\mathcal{X}}+\tfrac{1}{2}\mathcal{Y}+O(\mathcal{Y}^2,\mathcal{X}\mathcal{Y}), \tag{20}$$

and thus assume that 3D effects are sufficiently gradual (this will be discussed in the next section). They only differ in the way they approximate the term $\sqrt{\mathcal{I}+\mathcal{X}}$. In Eq. (20), the operator $\mathcal{X}$ and the operator $\mathcal{Y}$ are separated. Any PE model based on this formulation is thus amenable to the alternating direction method. It is worth mentioning that this approach is very advantageous for 3D modeling. Assume, for instance, that an implicit Crank–Nicolson scheme is used instead of the alternating direction method, and is applied directly on Eq. (12). Given the 3D field ψ at the discrete range r_n, ψ is obtained at the next discrete range $r_n+\Delta r$ by solving

$$\left(\mathcal{I}-\frac{ik_0\Delta r}{4}(\mathcal{X}+\mathcal{Y})\right)\psi(r_n+\Delta r,\theta,z;\omega)$$
$$=\left(\mathcal{I}+\frac{ik_0\Delta r}{4}(\mathcal{X}+\mathcal{Y})\right)\psi(r_n,\theta,z;\omega). \tag{21}$$

Then, applying a FD discretization in z and θ, Eq. (21) leads to a large block-tridiagonal linear system of order $M\times N$. Unfortunately, because of its block-tridiagonal structure, this large linear system cannot be decomposed into smaller auxiliary linear systems. Furthermore, since realistic acoustic wave propagation problems generally require a large number of points in both the z and θ direction, it is not possible to use any direct algorithm (like Gaussian elimination) due to memory storage limitations. Instead, a preconditioned iterative algorithm must be used, the efficiency of the solver highly depending on the preconditioning procedure. In TRIPARADIM, the standard narrow-angle 3D PE model was rewritten in a new coordinate system in an effort to handle properly the varying bottom topography of the three-dimensional waveguide. The resulting mathematical formulation did not allow the coordinate decomposition of the operator as for other 3D PEs. The use of the alternating direction method was thus not possible and a numerical technique similar to Eq. (21) was chosen. The reader is referred to Ref. 31 for more details on the TRIPARADIM model. Though the resulting linear systems were sparse and could be solved using an efficient preconditioning technique at each range step, it has been shown that using such an iterative algorithm could lead to prohibitive computation times in comparison with other 3D PE models that are amenable to alternating direction methods, when solving practical problems in three-dimensional environments.[32,33] Computation time considerations are of major importance in 3D modeling, especially when broadband pulse propagation problems are addressed. In this latter case, any numerical technique using an alternating direction method should definitely be preferred to basic implicit Crank–Nicolson schemes.

C. Azimuthal wide-angle capability

The azimuthal narrow-angle capability of the 3D PE model 3DWAPE is now discussed. It is clear that the linear approximation with respect to the azimuth operator $\mathcal{Y}$ used in Eq. (8), or more generally in Eq. (20), is only valid when $|\mathcal{Y}\psi|\ll|\mathcal{X}\psi|$. This means that any horizontal deviation should be small compared to the vertical in-plane deviation of the propagating energy. This assumption is valid for slowly varying properties. The azimuthal narrow-angle approximation

may be inappropriate for problems with large out-of-plane deviations of the outward propagating field. For such problems, one should utilize a three-dimensional parabolic equation that handles wide-angle propagation in both the z and the θ directions. Siegmann–Kriegsmann–Lee[34,35] developed a three-dimensional parabolic equation with a wide-angle capability applying the Claerbout's coefficients to the operator $\mathcal{X}+\mathcal{Y}$:

$$\sqrt{\mathcal{I}+\mathcal{X}+\mathcal{Y}}=\mathcal{I}+\frac{\frac{1}{2}(\mathcal{X}+\mathcal{Y})}{\mathcal{I}+\frac{1}{4}(\mathcal{X}+\mathcal{Y})}+O((\mathcal{X}+\mathcal{Y})^3). \tag{22}$$

Neglecting the higher-order terms in $O((\mathcal{X}+\mathcal{Y})^3)$ then yields

$$\frac{\partial\psi}{\partial r}=ik_0\left(\frac{\frac{1}{2}(\mathcal{X}+\mathcal{Y})}{\mathcal{I}+\frac{1}{4}(\mathcal{X}+\mathcal{Y})}\right)\psi. \tag{23}$$

The wide-angle capability of this 3D parabolic equation has been demonstrated using an asymptotic multiscale analysis.[35] However, it is not amenable to the use of an alternating direction method. Instead, the following numerical scheme is obtained by applying an implicit Crank–Nicolson range-stepping procedure directly on Eq. (23):

$$\left(\mathcal{I}+\left(\frac{1}{4}-\frac{ik_0\Delta r}{4}\right)(\mathcal{X}+\mathcal{Y})\right)\psi(r_n+\Delta r,\theta,z;\omega)$$
$$=\left(\mathcal{I}+\left(\frac{1}{4}+\frac{ik_0\Delta r}{4}\right)(\mathcal{X}+\mathcal{Y})\right)\psi(r_n,\theta,z;\omega). \tag{24}$$

A FD technique in z and θ has been proposed and validated on several test examples.[36] Following our previous discussion, solving Eq. (24) may be costly. In order to take advantage of the alternating direction technique, higher-order terms may be incorporated while keeping the two operators $\mathcal{X}$ and $\mathcal{Y}$ separated. For instance, in PECAN,[7] the following [1/1] Padé azimuthal expansion is used:

$$\sqrt{\mathcal{I}+\mathcal{X}+\mathcal{Y}}=\sqrt{\mathcal{I}+\mathcal{X}}+\frac{\frac{1}{2}\mathcal{Y}}{\mathcal{I}+\frac{1}{4}\mathcal{Y}}+O(\mathcal{Y}^3,\mathcal{X}\mathcal{Y}). \tag{25}$$

By neglecting higher-order terms, $O(\mathcal{Y}^3,\mathcal{X}\mathcal{Y})$, in (25), Eq. (7) then yields

$$\frac{\partial\psi}{\partial r}=ik_0\left(\sqrt{\mathcal{I}+\mathcal{X}}+\frac{\frac{1}{2}\mathcal{Y}}{\mathcal{I}+\frac{1}{4}\mathcal{Y}}\right)\psi, \tag{26}$$

where the square-root operator present on the right-hand side is approximated using the split-step Padé algorithm.[37] Chen *et al.* use a quadratic Taylor series azimuthal expansion,[38]

$$\sqrt{\mathcal{I}+\mathcal{X}+\mathcal{Y}}=\mathcal{I}+\tfrac{1}{2}\mathcal{X}-\tfrac{1}{8}\mathcal{X}^2+\tfrac{1}{2}\mathcal{Y}-\tfrac{1}{8}\mathcal{Y}^2+O(\mathcal{X}^3,\mathcal{X}\mathcal{Y},\mathcal{Y}^3). \tag{27}$$

Neglecting terms in $O(\mathcal{X}^3,\mathcal{X}\mathcal{Y},\mathcal{Y}^3)$ in Eq. (27) leads to the following parabolic equation:

$$\frac{\partial\psi}{\partial r}=ik_0\left(\frac{1}{2}\mathcal{X}-\frac{1}{8}\mathcal{X}^2+\frac{1}{2}\mathcal{Y}-\frac{1}{8}\mathcal{Y}^2\right)\psi. \tag{28}$$

Equation (28) can be seen as an azimuthal quadratic correction of the *LSS-3-D Wide Angle Wave Equation* given in Eq. (15). It has been implemented in the FOR3D code. Notice that the azimuthal rational-function approximation used in Eq. (25) and the azimuthal polynomial-function approximation used in Eq. (27) are correct to quadratic terms in azimuth, but both neglect the term in $O(\mathcal{X}\mathcal{Y})$. Retaining this term would not allow the use of an alternating direction method.

In its original configuration, the 3DWAPE code had a narrow-angle capability in azimuth [see Eq. (10)]. It has been modified to handle higher-order approximations along θ while keeping the two operators $\mathcal{X}$ and $\mathcal{Y}$ separated. Instead of the azimuthal linear approximation in Eq. (8), a Padé series azimuthal expansion is used:

$$\sqrt{\mathcal{I}+\mathcal{X}+\mathcal{Y}}=\mathcal{I}+\sum_{k=1}^{n_p}\frac{a_{k,n_p}\mathcal{X}}{\mathcal{I}+b_{k,n_p}\mathcal{X}}+\sum_{k=1}^{m_p}\frac{a_{k,m_p}\mathcal{Y}}{\mathcal{I}+b_{k,m_p}\mathcal{Y}}$$
$$+O(\mathcal{X}^{2n_p+1},\mathcal{Y}^{2m_p+1},\mathcal{X}\mathcal{Y}), \tag{29}$$

where m_p is the number of Padé terms and a_{k,m_p}, b_{k,m_p}, $1\leqslant k\leqslant m_p$, are real coefficients given analytically by Eq. (9) wherein n_p is to be replaced by m_p. Then, by neglecting the last term in Eq. (29), Eq. (7) yields

$$\frac{\partial\psi}{\partial r}=ik_0\left(\sum_{k=1}^{n_p}\frac{a_{k,n_p}\mathcal{X}}{\mathcal{I}+b_{k,n_p}\mathcal{X}}+\sum_{k=1}^{m_p}\frac{a_{k,m_p}\mathcal{Y}}{\mathcal{I}+b_{k,m_p}\mathcal{Y}}\right)\psi. \tag{30}$$

The 3D PE has now a very-wide-angle capability in depth and a very-wide-angle capability in azimuth, but, due to the term in $O(\mathcal{X}\mathcal{Y})$, the 3DWAPE model does not, strictly speaking, have a wide-angle capability. Since the depth operator $\mathcal{X}$ and the azimuthal operator $\mathcal{Y}$ are well separated in Eq. (30), an alternating direction method is used. Given the 3D field ψ at the discrete range r_n, ψ is obtained at the next discrete range $r_n+\Delta r$ in two steps. The first step consists in computing n_p intermediate fields denoted $u^{(1)}(\theta,z)$, $u^{(2)}(\theta,z),\ldots,u^{(n_p)}(\theta,z)$ solving Eq. (18) for $1\leqslant k\leqslant n_p$. The second step consists in computing m_p intermediate fields $v^{(1)}(\theta,z)$, $v^{(2)}(\theta,z),\ldots,v^{(m_p)}(\theta,z)$ solving

$$\left(\mathcal{I}+\left(b_{k,m_p}-\frac{ik_0\Delta r}{2}a_{k,m_p}\right)\mathcal{Y}\right)v^{(k)}(\theta,z)$$
$$=\left(\mathcal{I}+\left(b_{k,m_p}+\frac{ik_0\Delta r}{2}a_{k,m_p}\right)\mathcal{Y}\right)v^{(k-1)}(\theta,z), \tag{31}$$

for $1\leqslant k\leqslant m_p$, where $v^{(0)}(\theta,z)$ denotes the last intermediate field $u^{(n_p)}(\theta,z)$ computed in the first step, and $v^{(m_p)}(\theta,z)$ corresponds to the 3D field ψ at the discrete range $r_n+\Delta r$: $\psi(r_n+\Delta r,\theta,z;\omega)=v^{(m_p)}(\theta,z)$. Note that for the particular value $m_p=1$ (for which we have $a_{1,1}=1/2$ and $b_{1,1}=1/4$), Eq. (31) reduces to

$$\left(\mathcal{I}+\left(\frac{1}{4}-\frac{ik_0\Delta r}{4}\right)\mathcal{Y}\right)\psi(r_n+\Delta r,\theta,z;\omega)$$
$$=\left(\mathcal{I}+\left(\frac{1}{4}+\frac{ik_0\Delta r}{4}\right)\mathcal{Y}\right)u^{(n_p)}(\theta,z), \tag{32}$$

which corresponds exactly to the equation used in PECAN to account for 3D coupling [see Eq. (2.38) of Ref. 7].

III. DESCRIPTION OF THE THREE-DIMENSIONAL TEST CASES

Two three-dimensional test cases are investigated, referred to as test cases A and B. Test case A is based on a three-dimensional extension of the original two-dimensional ASA wedge configuration and is similar to the 3D test case considered by other modelers in more recent papers.[7–9] Test case A has been proposed as a 3D benchmark problem at the 141st ASA meeting held in Chicago in June 2001. It consists of an isovelocity water layer of sound speed $c_w=1500$ m/s and density $\rho_w=1$ g/cm^3, overlying a lossy homogeneous half-space sedimental layer of sound speed $c_{\text{sed}}=1700$ m/s, density $\rho_{\text{sed}}=1.5$ g/cm^3 and attenuation $\alpha_{\text{sed}}=0.5$ dB/λ, which leads to a critical grazing angle value of approximately 28°. There is no attenuation in the water layer. The parametrization of the water-sediment interface is given by $z_{\text{sed}}(r,\theta)=h_{\text{sed}}(x)$ where $x=r\cos\theta$ and

$$h_{\text{sed}}(x)=\begin{cases}200(1-x/4000) & \text{if } |x|\leqslant 3600\\ 20 & \text{if } x\geqslant 3600\\ 380 & \text{if } x\leqslant -3600.\end{cases} \tag{33}$$

The water depth decreases linearly from 200 m at $r=0$ to 20 m at $r=3.6$ km, and is range-independent for $r\geqslant 3.6$ km along the $\theta=0°$ azimuth (up-slope direction). It increases linearly from 200 m at $r=0$ to 380 m at $r=3.6$ km, and is range-independent for $r\geqslant 3.6$ km along the $\theta=180°$ azimuth (down-slope direction). It is invariant along the $\theta=90°$ and $\theta=270°$ azimuths (cross-slope directions). It thus makes an angle with a constant value of 2.86° with respect to the ocean surface at both $\theta=0°$ and $\theta=180°$, and leads to a zero-slope at both $\theta=90°$ and $\theta=270°$ (see Fig. 1).

Test case B is based on a modification of the three-dimensional gaussian canyon test case devised for the SWAM'99 Workshop[17] held in Monterey CA in September 1999, and only differs from test case A in the description of the bottom topography. Test case B consists of an isovelocity water layer overlying a lossy homogeneous half-space sedimental layer (the geoacoustic parameters corresponding to each layer are identical to the ones used in test case A) with a parametrization of the water-sediment interface given by

$$h_{\text{sed}}(x)=h_0+h_1\exp(-x^2/\sigma^2), \tag{34}$$

where the parameters h_0, h_1, and σ are expressed in meters. Like test case A, the water depth only depends on the x direction. Assuming $\sigma=4h_1$, the maximum slope of the bathymetry in the x direction (which we will refer to as the cross-canyon direction) is approximately 12.1° at $x=\sigma/\sqrt{2}$. The values $h_0=200$ m, $h_1=500$ m, and $\sigma=4\times h_1=2000$ m correspond to the ones used during the SWAM'99 workshop.

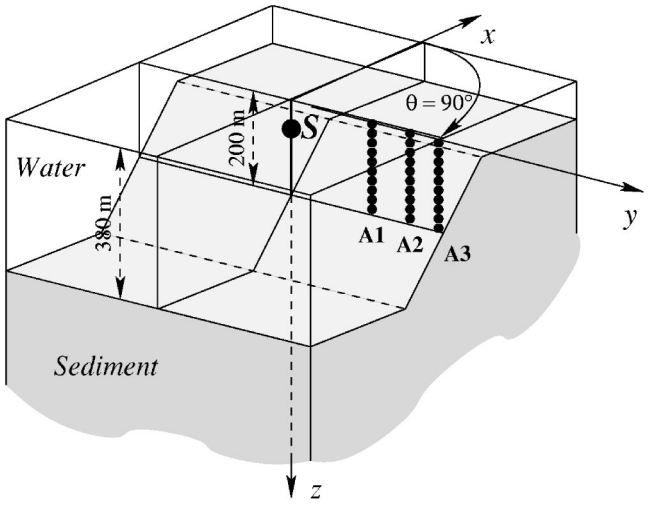


FIG. 1. Geometry of the 3D (truncated) wedge shaped waveguide considered in test case A.

In this paper, $h_0=20$ m, $h_1=180$ m, and $\sigma=4h_1=720$ m. The configuration used in test case B is depicted in Fig. 2. The water depth decreases from 200 m at $r=0$ to 20 m at $r\to\infty$ along the $\theta=0°$ and $\theta=180°$ azimuths (cross-canyon directions). It is invariant along the $\theta=90°$ and $\theta=270°$ azimuths (along-canyon directions). It makes a varying angle with respect to the ocean surface (with a maximum value of approximately 12.1° at $r=509.12$ m) at $\theta=0°$ and $\theta=180°$, and leads to a zero-slope at $\theta=90°$ and $\theta=270°$.

For both test cases, an isotropic point source is located at point $S=(x_S=0,y_S=0,z_S=40\text{ m})$. Its time dependence is a Hanning-weighted four-period sine wave (see Fig. 3) given by

$$\mathcal{S}(t)=\begin{cases}\frac{1}{2}(1-\cos(\omega_c t/4))\sin(\omega_c t) & \text{if } 0\leqslant t\leqslant 4/f_c\\ 0 & \text{if } t>4/f_c,\end{cases} \tag{35}$$

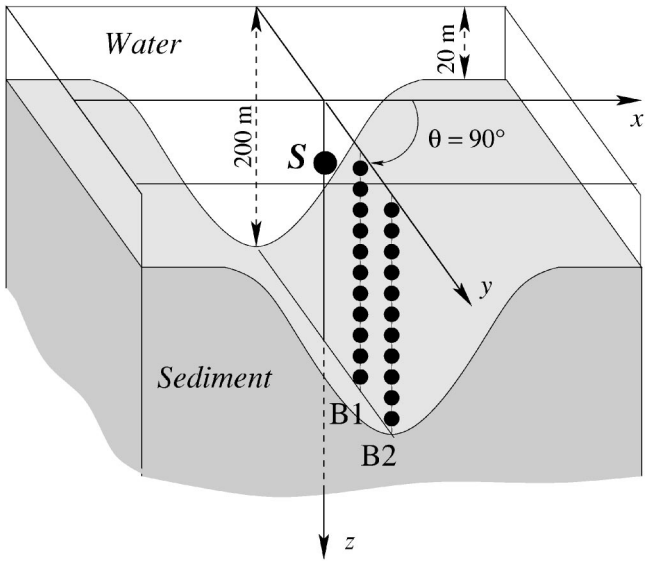


FIG. 2. Geometry of the 3D Gaussian canyon waveguide considered in test case B.

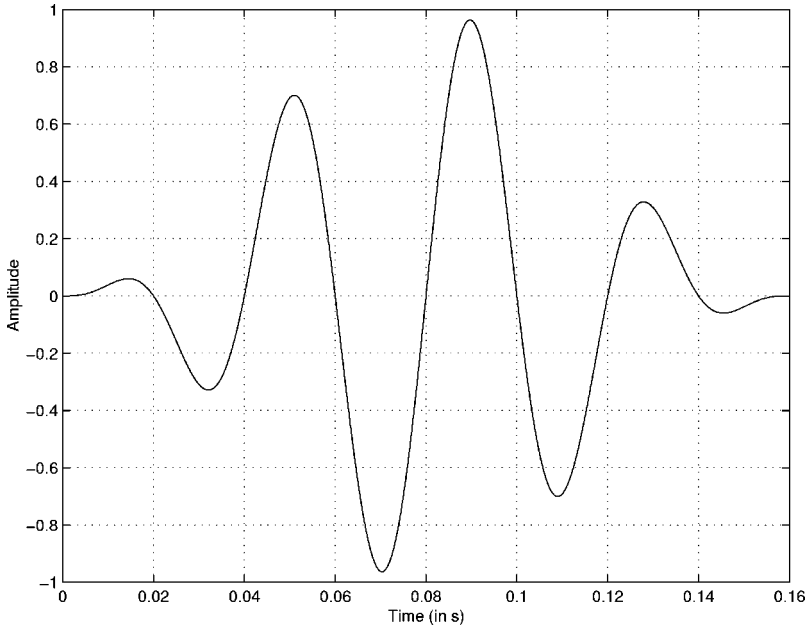


FIG. 3. Time-dependence of the source pulse, for test cases A and B.

where $\omega_c = 2\pi f_c$ and $f_c = 25$ Hz. The source pulse is centered at frequency $f_c = 25$ Hz with a 40 Hz bandwidth covering the band 5–45 Hz. The amplitudes of both the real and imaginary parts of the source spectrum are very small for frequencies below 5 Hz and above 45 H (see Fig. 4). The pulse length is 0.16 s. For both test cases, the geometry of the waveguide at $\theta = 90°$ (which corresponds to the cross-slope direction for test case A and to the canyon axis for test case B) is characterized by a zero-slope (the water depth is constant and equal to 200 m). Due to the geometry of the two waveguides, large 3D effects are expected in this direction, along which several vertical arrays are placed. In test case A, three vertical arrays labeled A1, A2, and A3 are placed across slope at ranges $r_{A1} = 16$ km, $r_{A2} = 22$ km, and $r_{A3} = 25$ km from the source *S*. In test case B, two vertical arrays labeled B1 and B2 are placed along the canyon axis at ranges $r_{B1} = 16$ km and $r_{B2} = 20$ km from the source *S*. Each of the vertical arrays is composed of 19 elements evenly spaced in depth between 10 and 190 m. Note that, for both test cases, the vertical arrays and the source lie in the same 200 m isobath vertical plane. All the numerical simulations shown in the next sections were performed on a 2.2 GHz mono-processor Dell-workstation. Neither vectorization nor parallel computing was used. Unless specified otherwise, all the following numerical results were obtained using the 3D PE model 3DWAPE.

IV. RESULTS FOR THE 3D ASA WEDGE

A. cw point source results

Since the source pulse is centered at $f_c = 25$ Hz, the acoustic problem at that specific frequency is treated first.

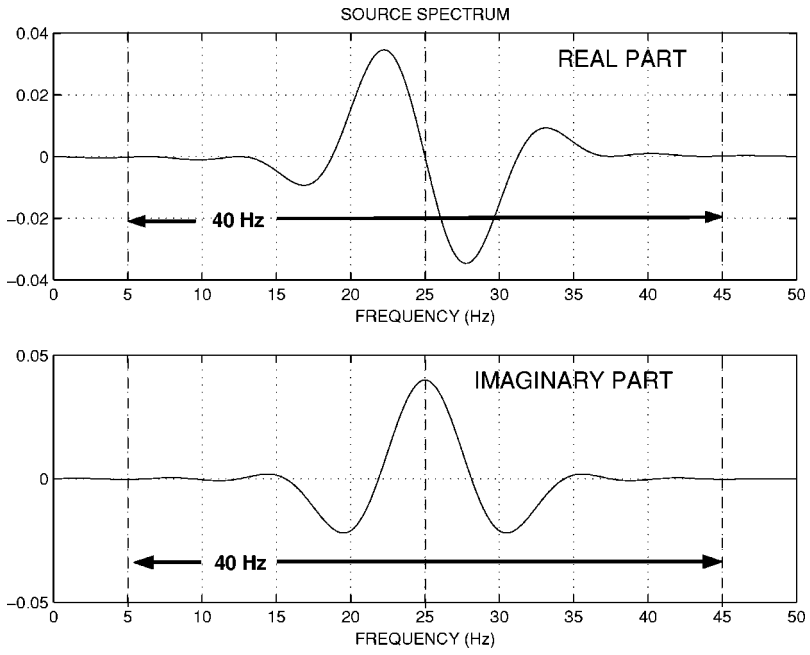


FIG. 4. Spectrum of the source pulse, for test cases A and B.

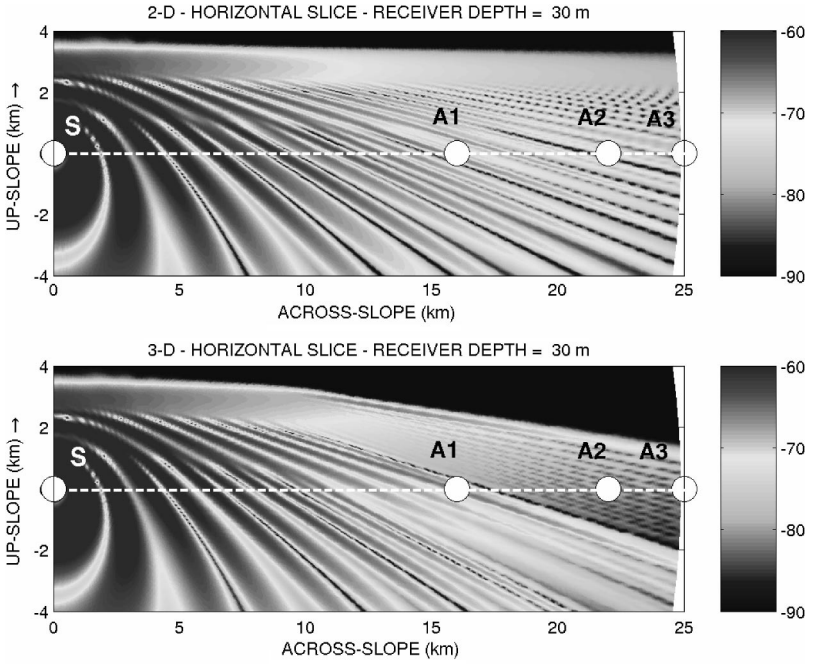


FIG. 5. Transmission loss (in dB re 1 m) at 25 Hz at a depth of 30 m for test case A corresponding to $N\times 2$D (upper subplot) and 3D (lower subplot) PE calculations. For both calculations, a modal sum was used as a starting field. The 200 m isobath is indicated by a dashed white line.

The $N\times 2$D and 3D computations were carried out using $\Delta r=10$ m, $\Delta z=1$ m (i.e., $\Delta r=\lambda/6$ and $\Delta z=\lambda/60$ where λ denotes the acoustic wavelength) and $n_p=2$. The maximum computation range is $r_{\max}=25$ km and the reference sound speed is $c_{\text{ref}}=1500$ m/s. The $N\times 2$D and 3D PE algorithms were initialized at $r=0$ using a modal source. Since long-range propagation is considered, the modal sum was limited to the discrete modal spectrum. Only three propagating modes are present at a frequency of 25 Hz at the source location. The maximum depth of the computational grid is $z_{\max}=600$ m. An increasing attenuation coefficient $\alpha_{\text{abs}}(z)$ was introduced in the lower part of the domain (corresponding to depths between $z_{\text{abs}}=450$ m and $z_{\max}=600$ m) to prevent spurious reflections from the pressure-release imposed boundary condition at $z_{\max}$. For the 3D calculations, a Padé 1 approximation in azimuth (i.e., $m_p=1$) and an eighth-order FD azimuthal scheme with $M=3240$ points (i.e., an azimuthal step size of 1/9th of a degree) were used. This azimuthal increment corresponds to an arclength increment ΔS of the order of $3\lambda/4$ at the maximum computation range $r_{\max}$. As shown in Ref. 32, using a second-order FD azimuthal scheme would require in this case $M=23\,040$ discrete points in azimuth, i.e., an azimuthal step size of 1/64th of a degree (this azimuthal increment corresponding to an arclength increment ΔS of the order of $\lambda/10$ at the maximum computation range $r_{\max}$).

Gray-scale images of the transmission loss (TL $=-20\log_{10}(|\psi(r,\theta,z;\omega_c)|/\sqrt{r}$ with $\omega_c=2\pi f_c$) at a receiver depth of 30 m corresponding to $N\times 2$D and 3D calculations are displayed in Fig. 5. The $\theta=90°$ direction corresponding to the 200 m isobath is indicated by a dashed line. The positions of the source S and the three vertical arrays A1, A2, A3 are also indicated on each subplot. Due to the geometrical symmetry of the problem about the up-slope direction, both $N\times 2$D and 3D solutions are displayed as a function of range and azimuth in a limited azimuthal sector. Figure 6 shows transmission loss-vs-range curves at $z=30$ m and $\theta=90°$ (across-slope). The thin dashed curve is a $N\times 2$D PE calculation and the bold solid curve is a 3D PE calculation. Comparing the two subplots of Fig. 5, the effects of azimuthal coupling are evident, mainly in the vicinity of the cross-slope direction and at long ranges. These effects are well known and have been explained in detail by several authors. They correspond to intramodal interference effects. Recall that three propagating modes are excited at the source. Since the vertical geometry in the cross-slope direction (characterized by a zero-slope) is a classical 200-m-deep Pekeris waveguide, any 2D or $N\times 2$D model can predict the presence of the three initial modes at $\theta=90°$ for $0\leqslant r$

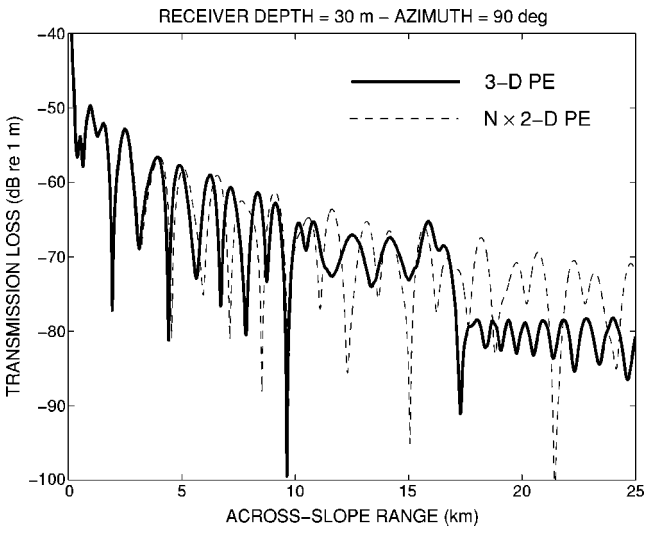


FIG. 6. $N\times 2$D and 3D transmission loss (in dB re 1 m) comparisons at 25 Hz at a receiver depth of 30 m along a 200 m isobath in the cross-slope direction for test case A. The thin dashed curve is an $N\times 2$D PE calculation and the bold solid curve is a 3D PE calculation. For both calculations the Greene's source was used as a starting field.

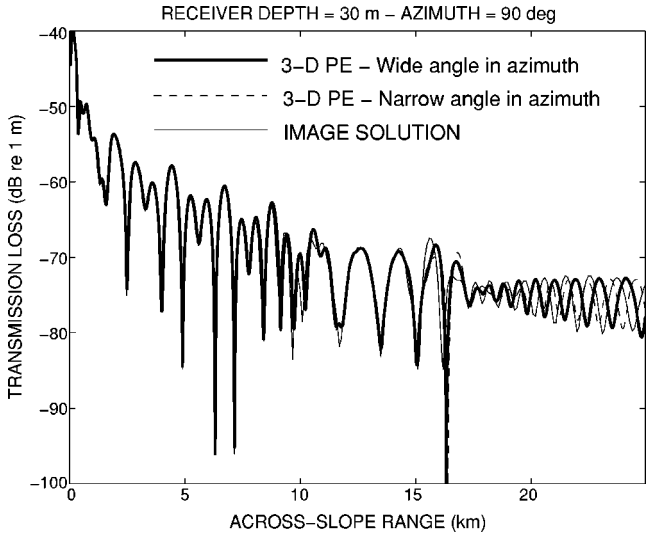


FIG. 7. 3D transmission loss (in dB re 1 m) comparisons at 25 Hz at a receiver depth of 30 m along a 200 m isobath in the cross-slope direction for test case A. The thin solid curve corresponds to the image solution. The two other curves are 3D PE calculations using a wide-angle (bold solid curve) or a narrow-angle (thin dashed curve) approximation in azimuth. The source depth is 100 m. Both 3D PE calculations were initialized using the Greene's source.

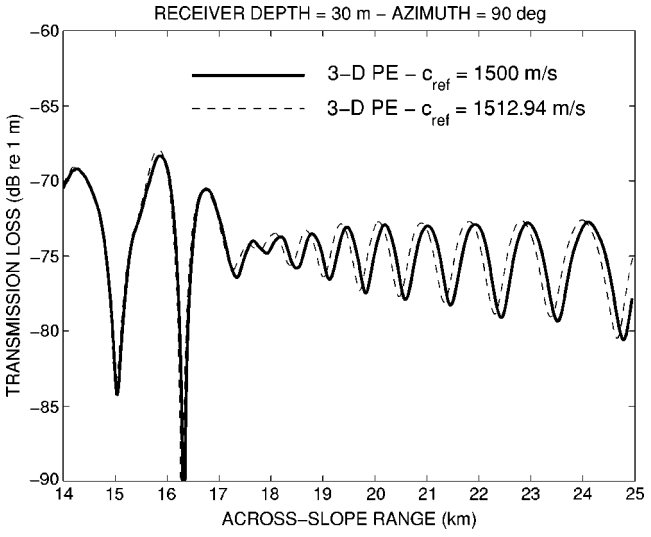


FIG. 8. 3D transmission loss (in dB re 1 m) comparisons at 25 Hz at a receiver depth of 30 m along a 200 m isobath in the cross-slope direction ($\theta=90°$, 14 km$\leqslant r\leqslant$25 km) for test case A. The bold solid curve is a 3D PE calculation with $c_{\text{ref}}=1500$ m/s (same solution as shown in Fig. 7) and thin dashed curve is a 3D PE calculation with $c_{\text{ref}}=1512.94$ m/s. The source depth is 100 m. The Greene's source was used as a starting field.

$\leqslant$25 km. The use of a full 3-D PE (i.e., azimuthally coupled model) causes these three propagating modes to be horizontally refracted down the slope, which leads to a succession of three distinct zones in the cross-slope direction (see Fig. 6). For ranges less than approximately 11 km, the three initial propagating modes are present. Only two modes are present for $r\geqslant 11$ km up to $r\approx 16$ km, due to the 3D shadowing effect of mode 3. After $r\approx 16$ km where the mode shadowing effet of mode 2 occurs, only mode 1 is present. The interference pattern in the 3D solution starting at $r\approx 17.5$ km corresponds to the 3D mode self-interference effect of mode 1.

To show that no significant 3D effects are omitted in the 3D PE calculations, the results were compared with a reference solution based on the image source method and originally provided by Westwood.[39] This analytical solution is expressed as a sum of ray fields, each of which take the form of a double integral over plane waves. Details of the method are given in Ref. 16. Note that the source depth was 100 m to match the original ASA benchmark problem. The 3D PE and the image solutions are plotted in Fig. 7. The bold solid curve is the 3D PE solution obtained using two Padé terms in depth ($n_p=2$) and one Padé term in azimuth ($m_p=1$). Also plotted on the same figure as a thin dashed curve is the 3D PE solution obtained using two Padé terms in depth and a narrow-angle azimuthal approximation. The 3D PE calculations were initialized using the Greene's source.[40] We observe an overall good agreement between the two 3D PE solutions and the image solution in the cross-slope direction, showing that both models predict the same 3D effects. However, there are some differences, notably a shift in the phasing at long ranges mainly where mode 1 interferences occur. Note that the 3D PE solution obtained using a Padé 1 approximation in azimuth is closer to the reference solution than the azimuthal narrow-angle 3D PE solution, although a phase shift is still present. The use of higher-order approximations in azimuth (i.e., increasing the number m_p of Padé terms) does not improve phase predictions. Recall that the PE solutions depend on the value of the reference sound speed c_{ref}. For the solutions shown in Fig. 7, $c_{\text{ref}}=1500$ m/s. Figure 8 shows 3D PE solutions obtained using two distinct values of c_{ref}. The bold solid curve corresponds to $c_{\text{ref}}=1500$ m/s (the value of the sound speed in the homogeneous water layer) and the thin dashed curve to $c_{\text{ref}}=1512.94$ m/s (the value of the horizontal phase speed of mode 1 at the source location). As already pointed out by Smith,[9] changing the reference sound speed value induces a shift in the PE solutions. This shift is yet less pronounced than that between PE solutions shown in Fig. 7. This suggests that the disagreement between the image and the PE results is mainly due to the fact that the 3D PE model does not have a wide-angle capability since the term in $O(\mathcal{XY})$ is not handled (see the discussion in Sec. II C). However, despite the mismatch in the mode 1 interference pattern, all the physical 3D effects are reproduced by the 3D PE model.

B. Modal initialization results

Azimuthal coupling effects may be easily observed by exciting individual modes at the source location and propagating them outward in range. The 3D PE marching algorithm was initialized by each of the three propagating modes separately. Figure 9 shows TL plots (horizontal slices at constant depth $z=30$ m) obtained by initializing the 3D PE model using mode 1, mode 2, and mode 3. In order to reduce the phase error in PE calculations, the horizontal modal phase speed was used for the reference sound speed, i.e., $c_{\text{ref}}=1512.94$ m/s for mode 1, $c_{\text{ref}}=1554.44$ m/s for mode 2, and $c_{\text{ref}}=1632.42$ m/s for mode 3. For each mode, the source field was assumed to be omnidirectional. Vertical cross sections of the 3D PE solutions for mode 1, mode 2, mode 3

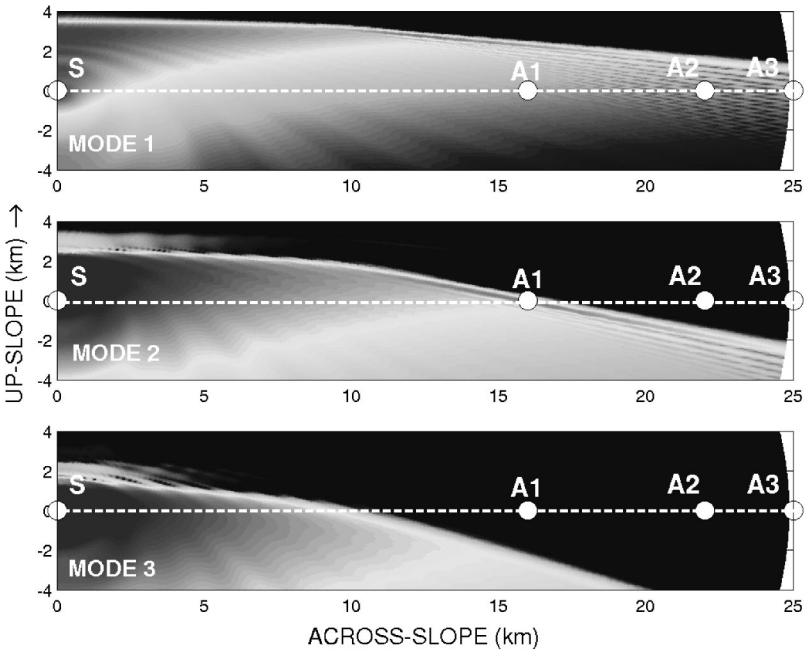


FIG. 9. Transmission loss (in dB re 1 m) at 25 Hz at a depth of 30 m for test case A corresponding to 3D PE calculations and different omnidirectional mode excitations. From top to bottom: mode 1, mode 2, and mode 3. On each subplot, the 200 m isobath is indicated by a dashed line.

along the $\theta=90°$ direction are shown in Fig. 10. Also shown in Fig. 10 is the 3D PE solution obtained initializing by the Greene's source.

The 3D PE results were compared with the predictions from adiabatic modal theory.[41] Figure 11 shows the modal ray diagrams in the (horizontal) yx plane for each omnidirectional modal initialization. The modal ray paths were calculated using the method given in Ref. 8. Hereafter, a modal-ray path that corresponds to the mth mode and makes initially an angle ϕ_0 with the y axis (i.e., $\phi_0=0°$ points across slope) is denoted Γ_{m,ϕ_0} (see Fig. 12). According to the Cartesian coordinate system chosen, for any value of $m \in\{1,2,3\}$ and any value of $\phi_0 \in]-90°,90°[$, the modal-ray path Γ_{m,ϕ_0} is a representative function $y \mapsto x_{m,\phi_0}(y)$ satisfying the following Cauchy problem

$$\frac{dx}{dy}=\frac{\sqrt{1-(c_m(x)\times\cos(\phi_0)/c_{r,m})^2}}{c_m(x)\times\cos(\phi_0)/c_{r,m}}, \quad y\geqslant 0,$$
$$x(y=0)=0. \tag{36}$$

Here, $\phi_0=90°$ ($\phi_0=-90°$) corresponds to an initial launch in the up-slope direction (in the down-slope direction) and c_m is the x-dependent phase velocity of the mth mode satisfying $c_w<c_m(x)<c_{\text{sed}}$ and $c_m(x=0)=c_{r,m}$, where $c_{r,m}$ is

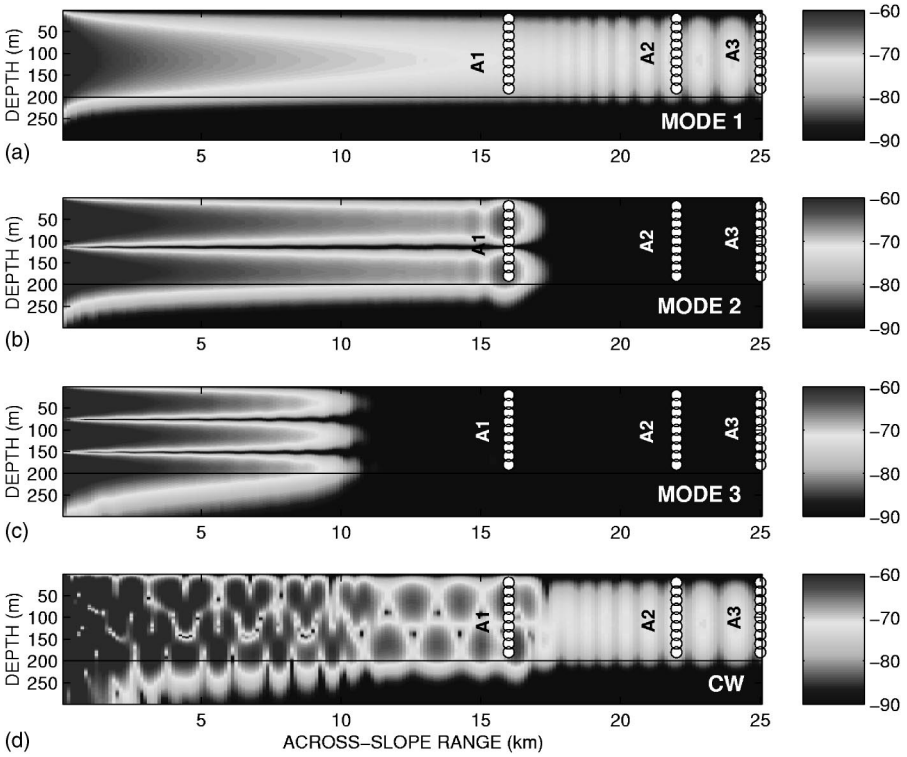


FIG. 10. 3D transmission loss (in dB re 1 m) at 25 Hz for test case A (vertical slices in the cross-slope direction) corresponding to 3D PE calculations and (a) mode 1, (b) mode 2, (c) mode 3, and (d) Greene's source excitations.

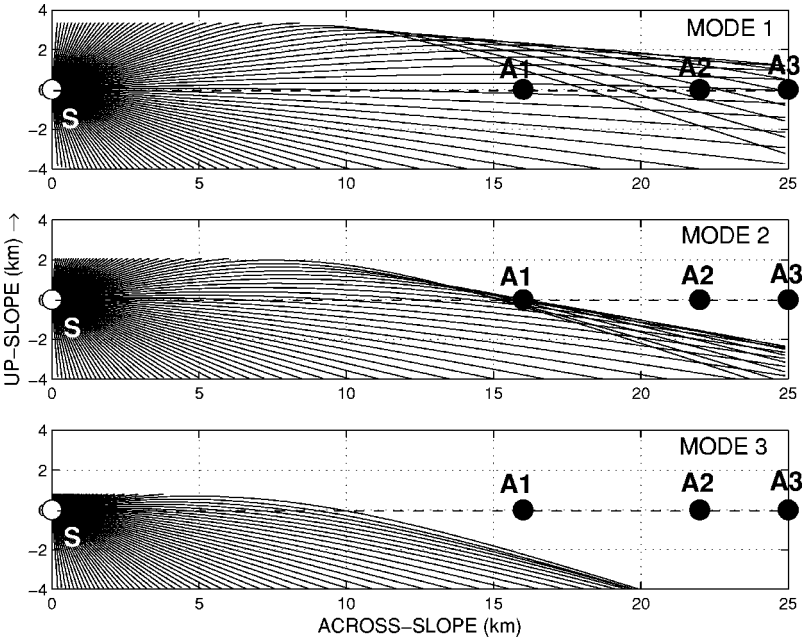


FIG. 11. Modal ray diagrams (top view) for test case A, obtained solving the differential equation given in (36) for $-90° < \phi_0 < 90°$. From top to bottom: mode 1, mode 2, and mode 3.

the horizontal phase speed of the mth mode at the source location. The value of $c_m(x)$ for any given x is numerically evaluated by solving

$$\tan\left(\sqrt{(\omega_c/c_w)^2-(\omega_c/c_m(x))^2}\times h_{\text{sed}}(x)\right) = -\frac{\rho_{\text{sed}}}{\rho_w}\times\frac{\sqrt{(\omega_c/c_w)^2-(\omega_c/c_m(x))^2}}{\sqrt{(\omega_c/c_m(x))^2-(\omega_c/c_{\text{sed}})^2}},$$

where $\omega_c = 2\pi f_c$. Note that a modal ray is stopped if it reaches an x coordinate such that $c_m(x)$ does not belong to $]c_w, c_{\text{sed}}[$ (i.e., when mode m becomes leaky). Characteristics of the modal rays that connect the source S with one of the three arrays A1, A2, and A3, are listed in Table I.

Comparison of the results displayed in Figs. 9 and 11 shows that the 3D PE solutions are in good agreement with the predictions from adiabatic modal theory. A modal ray which has an initial launch direction ϕ_0 toward the wedge apex turns around, as long as its grazing angle does not exceed the critical angle. A region of multiple arrivals is created in the vicinity of the cross-slope direction for sufficiently large ranges from the source, followed by a shadow region. Due to the increasing grazing angle with respect to the mode number, 3D effects are stronger for higher modes than for lower modes. No shadow zone is observed for mode 1 for $r \leqslant 25$ km in the cross-slope direction. Computations for ranges greater than 25 km for mode 1 would also show a shadow zone in the cross-slope direction. Following Glegg *et al.*,[2] three regions exist in the wedge-shaped waveguide for each modal initialization: (1) an inner region (corresponding to ϕ_0 greater than a critical launch angle ϕ_{crit}) where the modal rays propagating upslope exceed the critical grazing angle and are not turned around; (2) an outer region (corresponding to ϕ_0 less than ϕ_{crit}) where the modal-ray-turn-around occurs; and (3) a shadow region where there is no propagation. The value of the critical launch angle ϕ_{crit} depends on the mode number: $\phi_{\text{crit}} \approx 27°$ for mode 1, $\phi_{\text{crit}} \approx 23.7°$ for mode 2, and $\phi_{\text{crit}} \approx 16°$ for mode 3.

Results obtained with both models (PE and adiabatic mode) confirm the presence of mode 1 at the three vertical arrays A1, A2, and A3. Note that arrays A2 and A3 (unlike array A1) both lie in the multiple mode arrival area of mode 1, which means that multiple modal ray path arrivals exist for mode 1 on array A2 and array A3 (see Table I). Indeed, there are two modal ray path arrivals for mode 1 on array A2 (on array A3), the first one with a shallow angle ϕ_0

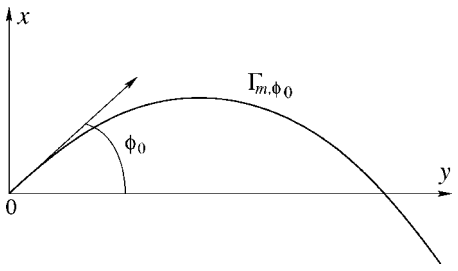


FIG. 12. Schematic of a modal-ray path Γ_{m,ϕ_0} in the (horizontal) yx plane corresponding to mode m, with initial launch angle ϕ_0. The y axis (the x axis) corresponds to the cross-slope (up-slope) direction in the wedge shaped waveguide considered in test case A. The source is placed at $x = y = 0$.

TABLE I. Characteristics (for a 25 Hz cw point source, test case A) of the modal rays launched from the source S and connected with one of the three arrays A1, A2, A3. The angle ϕ_0 denotes the initial launch angle ($\phi_0 = 0°$ points across-slope) of the modal eigenray, and L_{m,ϕ_0} denotes the length of the modal eigenray path Γ_{m,ϕ_0}.

Array	Mode m	ϕ_0 (deg)	L_{m,ϕ_0} (m)
A1	1	+1.8	16 002
A1	2	+13.5	16 186
A1	2	+16.5	16 300
A2	1	+2.7	22 008
A2	1	+19.7	22 845
A3	1	+3.2	25 013
A3	1	+16.7	25 640

$=+2.7°$ ($\phi_0=+3.2°$) with respect to the cross-slope direction, and the second one with a steeper angle $\phi_0=+19.7°$ ($\phi_0=+16.7°$). Two distinct time arrivals of the signal carried by mode 1 on both arrays A2 and A3 are thus expected, but only one single time arrival of the signal carried by mode 1 on array A1. The array A1 lies in the limit of the insonified region of mode 2. The signal carried by mode 2 should be present at array A1 only. In addition, the array A1 lies in the multiple mode arrival area of mode 2. Two distinct modal ray path arrivals for mode 2 with initial launch angles $\phi_0=+13.5°$ and $\phi_0=+16.5°$ are detected on array A1. Hence, two distinct arrival times of the signal carried by mode 2 on A1 are expected. Obviously, the three arrays A1, A2, and A3 lie in the shadow zone region of mode 3. No arrival of the signal carried by mode 3 is thus detected by any of the three arrays.

C. Broadband results

The case of the broadband source signal given in Eq. (35) is now discussed. The pulse response at a specific receiver R of range r_R, azimuth θ_R, and depth z_R, is obtained via a Fourier transform of the frequency-domain solution using

$$P(r_R,\theta_R,z_R;t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\hat{\mathcal{S}}(\omega)\hat{P}_{\text{norm}}(r_R,\theta_R,z_R;\omega)e^{-i\omega t}d\omega, \qquad (37)$$

where $\hat{\mathcal{S}}(\omega)$ is the source spectrum given by Eq. (4), and $\hat{P}_{\text{norm}}$ is the solution of the normalized frequency-domain wave equation (6). In Eq. (37), $\hat{P}_{\text{norm}}$ is set to zero for any (non-negative) frequency f outside the band 5–45 Hz. The frequency integral in Eq. (37) is evaluated numerically using a discrete Fourier transform (DFT). A time window of length $T=7$ s, with 4096 points, is used in the DFT algorithm. This yields values of the received signal at R in the time window $[t_{\min},t_{\min}+T]$ with a very fine time resolution Δt of 0.0017 s ($\Delta t=T/4096$). The length of the time window corresponds to a frequency sampling Δf of 0.1429 Hz ($\Delta f=1/T$), which leads to 281 discrete values within the frequency-band 5–45 Hz. Producing a pulse response at receiver R requires first computing values of $\hat{P}_{\text{norm}}$ at 281 discrete frequencies. This is achieved through repeated solution of the 3D PE model for uniformly distributed discrete frequencies ranging from 5 to 45 Hz and for $0\leqslant r\leqslant r_R$, $0\leqslant\theta\leqslant 2\pi$, $0\leqslant z\leqslant z_{\max}$.

For each solution at a single frequency f, several parameters need to be changed, their values depending on the value of the frequency f, or equivalently, of the acoustic wavelength λ. Suitable selections of the range, azimuthal, and depth increments Δr, $\Delta\theta$, Δz are crucial. For instance, undersampling (oversampling) the azimuthal direction may appear inappropriate to accurately compute the 3D effects (may lead to untimely computations). It is well established that the range and depth increments Δr and Δz should be sufficiently small in comparison to the acoustic wavelength λ. Our calculations were carried out using $\Delta r=\lambda/6$ and $\Delta z=\lambda/60$. Since an eighth-order FD scheme was used in azimuth, selection of the azimuthal increment $\Delta\theta=2\pi/M$ was achieved using a less restricting criterion with respect to the acoustic wavelength: $\Delta\theta$ (or equivalently M) was selected such that $\Delta S\approx 3\lambda/4$, where ΔS denotes the arclength increment at the maximum computation range $r_{\max}$. Note that using a second-order FD scheme in azimuth would require that $\Delta S\approx\lambda/10$. Our calculations were carried out using $M=648$ at $f=5$ Hz and $M=5832$ at $f=45$ Hz. For intermediate values of the frequency within the band 5–45 Hz, the number of discrete points in azimuth is obtained by linearly interpolating between 648 and 5832. For each frequency-domain calculation, the 3D PE marching algorithm was initialized at $r=0$ using the Greene's source. Two Padé terms in depth (i.e., $n_p=2$) and one Padé terms in azimuth (i.e., $m_p=1$) were used. The maximum depth $z_{\max}$ of the computation grid was placed several wave-lengths below the maximum depth of the seafloor and a layer of increasing absorption was placed in the lower part of the domain (just above $z_{\max}$) to attenuate the reflected energy, its width also depending on the acoustic wavelength.

The signal arrivals on the three vertical arrays A1, A2, and A3 were calculated. Comparisons of the signals received on vertical array A1 at a depth of 20 m corresponding to 2D and 3D computations are shown in Fig. 13. Both 2D and 3D solutions were multiplied by a factor $r_{\text{A1}}=16\,000$ m to compensate for spherical spreading. The pulse responses on vertical arrays A1, A2, and A3 were calculated using $t_{\min}=10.6$ s, $t_{\min}=14.6$ s, $t_{\min}=16.6$ s, respectively. They are displayed in Figs. 14, 15, and 16. The received signals are plotted intentionally as stacked time series versus depth which is helpful when analyzing the modal structure of the signal arrivals. Snapshots of the propagating pulse in the cross-slope direction at two distinct times, $t=10.7$ s and $t=16.7$ s, are shown in Figs. 17 and 18, respectively. For comparison, the signals obtained using 2D calculations are also shown.

Let us analyze first the signal arrivals on receiver arrays A1, A2, and A3, computed using a 2D PE model. The geometry of the waveguide in the cross-slope direction (characterized by a zero-slope) is seen by the 2D PE model as a classical 200-m-deep waveguide, which leads to the existence of three propagating modes when a 25 Hz cw source signal is considered. As expected, considering a broadband source pulse with a central frequency of 25 Hz, the propagating signal splits up in three distinct wave packets, which correspond to the signal carried by the three propagating modes of the waveguide. Now, using a 3D PE model, the signal arrivals exhibit only two distinct mode arrivals (instead of three as predicted by 2D calculations) at each of the three vertical arrays A1, A2, and A3. For each of them, it is clear that the first wave packet corresponds to the signal carried by mode 1. Indeed, its amplitude is low near the ocean surface, increases with depth toward mid-depth, then decreases toward $z=200$ m. This is obviously the depth dependence of mode 1. Following the same analysis, the second wave packet received on A1 (respectively, on A2 and A3) corresponds to the signal carried by mode 2 (respectively, by mode 1). These observations are coherent with the predictions of the previous section. Indeed, as expected, the signal carried by mode 1 is present for each of the three vertical arrays A1, A2, and

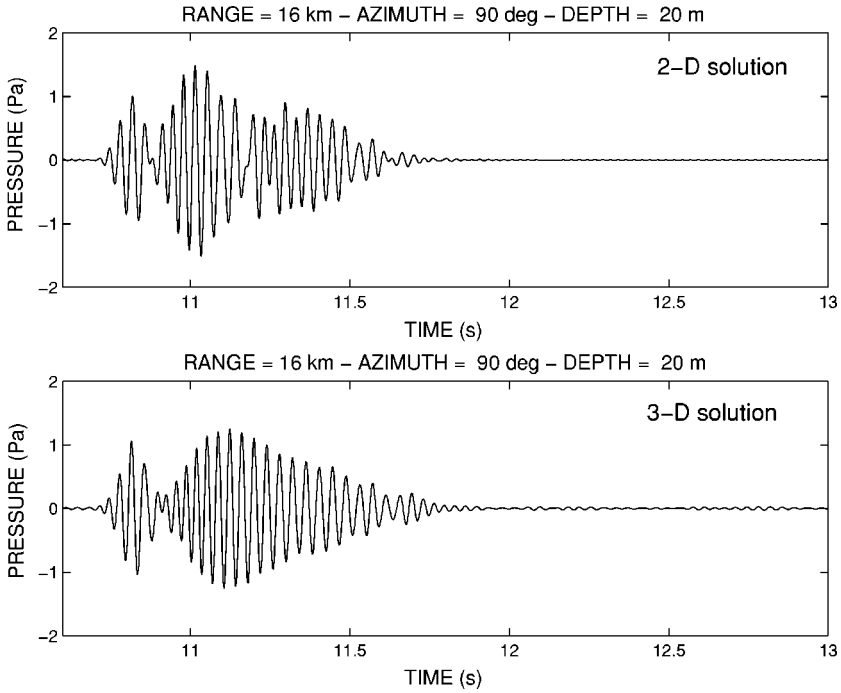


FIG. 13. Comparison of broadband pulse solutions for test case A and array A1 at a depth of 20 m: (upper) 2D calculation and (lower) 3D calculation.

A3. Since only A2 and A3 lie in the multiple mode arrival area of mode 1, two distinct mode 1 arrivals are observed on A2 and A3 (and only one single mode 1 arrival on A1), the first one corresponding to a modal ray path with a shallow angle with respect to the cross-slope direction, and the second one (weaker) to a modal ray path with a steeper angle. The time delay between the first and second time arrival of mode 1 is shorter at A3 than at A2. As expected, the signal carried by mode 2 is only present at A1 (recall that A1 lies in the limit of the insonified region of mode 2). Note that the second wave packet received on A1 corresponds to the merger of the two distinct (but close) time arrivals of mode 2. It thus appears more dispersed in time than the single mode 2 arrival predicted by a 2D calculation. Since the three arrays A1, A2, and A3 all lie in the shadow zone region of mode 3, no mode 3 signal arrival is observed at A1, A2, and A3. Note that all the modal eigenrays have been constantly refracted in the horizontal direction during their upslope and downslope propagation. Thus, the corresponding wave packets are different from the wave packets predicted using a 2D model.

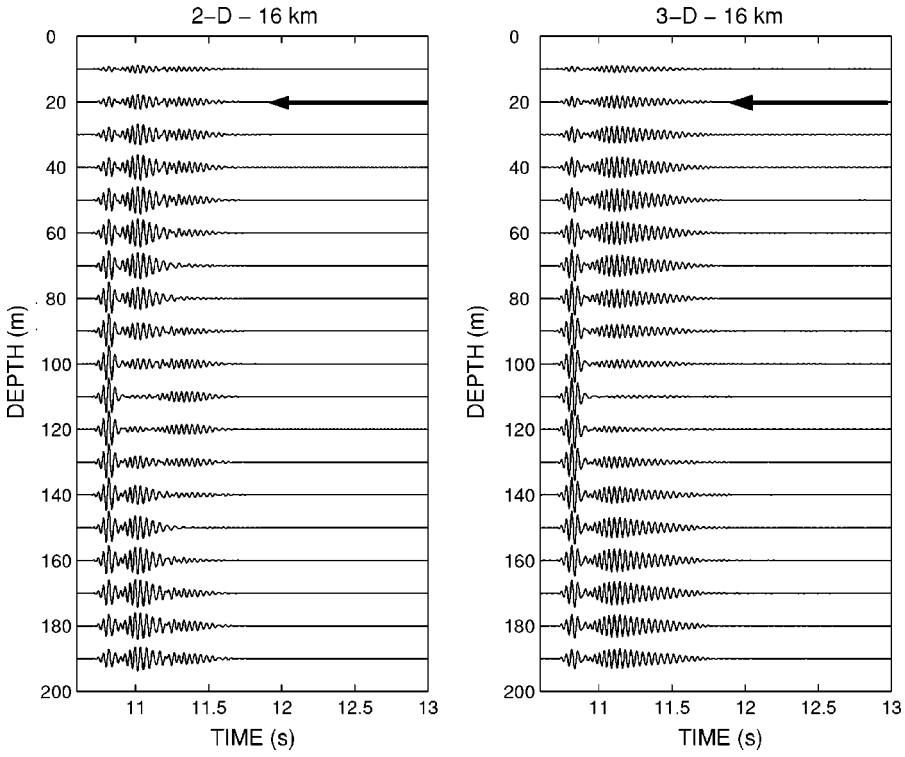


FIG. 14. Stacked time series vs depth for test case A corresponding to vertical array A1: (left) 2D calculation and (right) 3D calculation.

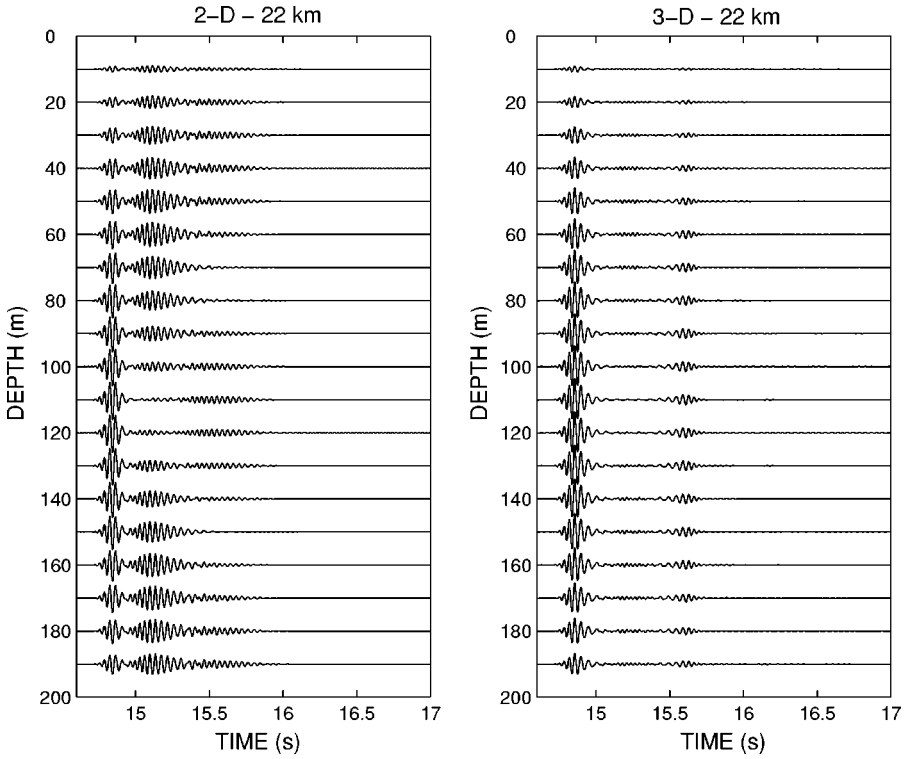


FIG. 15. Stacked time series vs depth for test case A corresponding to vertical array A2: (left) 2D calculation and (right) 3D calculation.

V. RESULTS FOR THE 3D GAUSSIAN CANYON

A. cw point source results

The acoustic problem involving a 25 Hz cw point source is analyzed first. The $N\times 2$D and 3D results presented hereafter were obtained using the same range and depth increments as in test case A ($\Delta r=10$ m, $\Delta z=1$ m). Both $N\times 2$D and 3D calculations were generated using a fifth-order Padé 3 approximation in depth (i.e., $n_p=3$). For the 3D calculations, a third-order Padé 1 approximation in azimuth (i.e., $m_p=1$) and $M=2880$ points were used. The number of discrete points used in azimuth is less than in test case A simply because the value of the maximum computation range $r_{\max}$ has been reduced from 25 to 20 km. As in test case A, the azimuthal increment $\Delta\theta=2\pi/2880$ corresponds to an arclength increment ΔS of the order of $3\lambda/4$ at the maximum computation range $r_{\max}=20$ km.

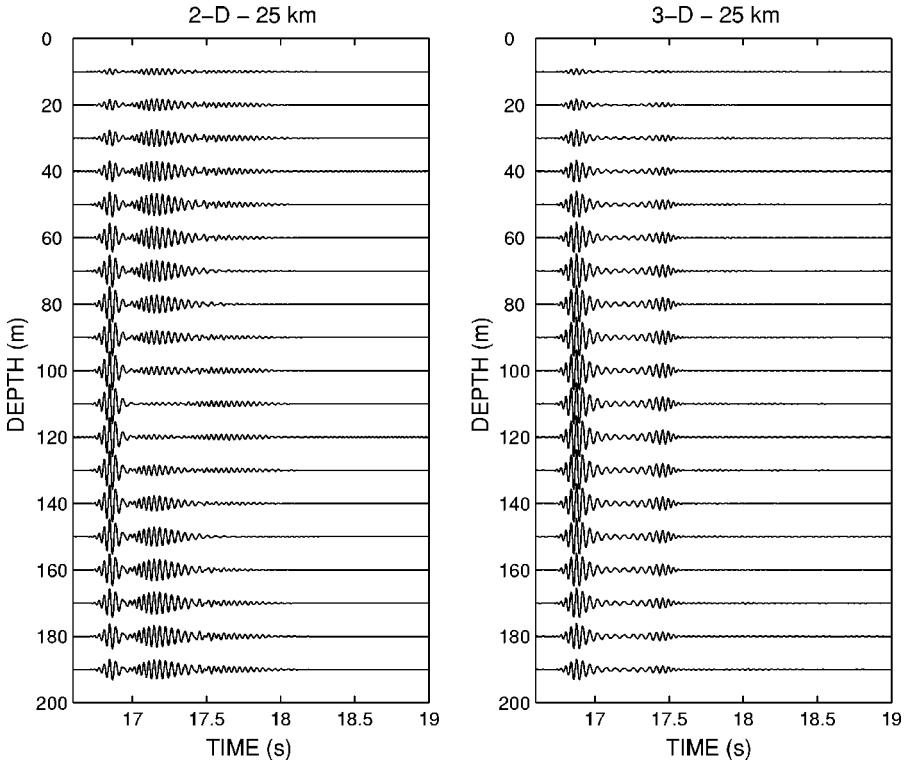


FIG. 16. Stacked time series vs depth for test case A corresponding to vertical array A3: (left) 2D calculation and (right) 3D calculation.

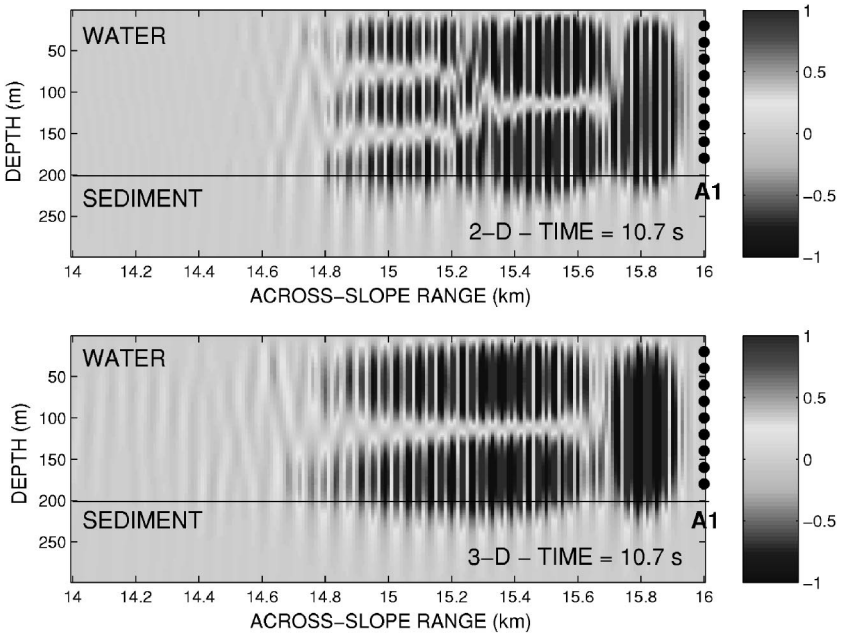


FIG. 17. Snapshots of the propagating pulse for test case A at $t=10.7$ s: (upper) 2D calculation and (lower) 3D calculation.

Gray-scale images of the transmission loss (horizontal slices at a receiver depth of 30 m) corresponding to $N\times 2$D and 3D calculations are shown in Fig. 19. For both of them, the modal sum was used as a starting field. A maximum computation depth of 600 m and a reference sound speed of 1500 m/s were used. The $\theta=90°$ direction corresponding to the canyon axis is indicated by a dashed line. The positions of the source S and the two vertical arrays B1, B2 are also indicated on each subplot. Like test case A, due to the geometrical symmetry of the problem about the x axis, both solutions are displayed as a function of range and azimuth in a limited azimuthal sector. By comparing the two subplots of Fig. 19, noticeable differences in both fields can be observed. The effects of azimuthal coupling are evident. Indeed, when azimuthal coupling is taken into account in the calculation, the acoustic energy is horizontally refracted by the sidewalls of the canyon and gets channeled in the y direction along the canyon axis. Note that both $N\times 2$D and 3D solutions are symmetric about the along-canyon direction.

Figure 20 shows transmission loss-versus-range curves at $z=30$ m and $\theta=90°$ (along the canyon axis) corresponding to 3D solutions obtained with various paraxial approximations in azimuth. Also shown on the same plot is the 2D solution. The comparison of the 3D solutions with the 2D solution confirms an enhancement of the acoustic level when azimuthal coupling is handled, due to the focusing of the acoustic energy along the canyon axis. There are also some differences between the two 3D PE solutions. The validity of the narrow-angle approximation in azimuth is subject to discussion for the present test case since, as shown in Fig. 20, using a Padé 1 paraxial approximation in azimuth changes the solution. It is worth mentioning that no change in the solution was observed for increasing values of m_p. How-

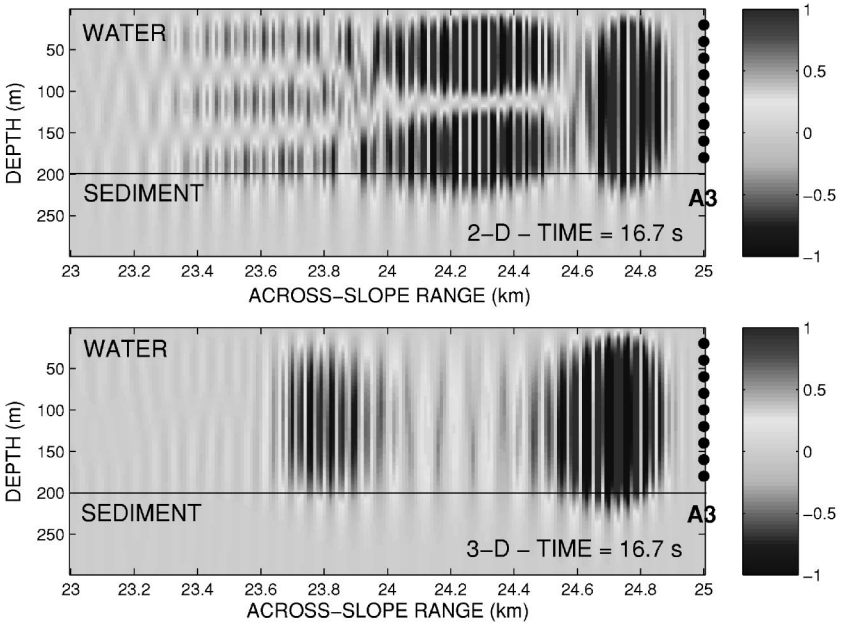


FIG. 18. Snapshots of the propagating pulse for test case A at $t=16.7$ s: (upper) 2D calculation and (lower) 3D calculation.

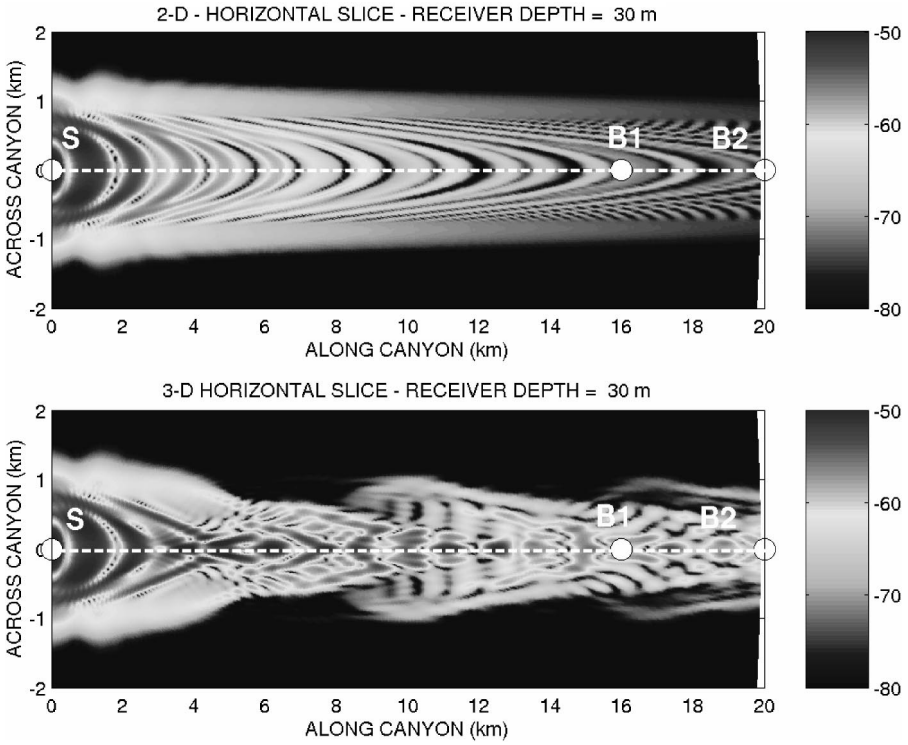


FIG. 19. Transmission loss (in dB re 1 m) at 25 Hz at a depth of 30 m for test case B corresponding to $N\times 2$D (upper subplot) and 3D (lower subplot) PE calculations. For each calculation, the modal sum was used as a starting field. The canyon axis is indicated by a dashed white line.

ever, the error between the 3D PE solution and the exact solution can not be quantified since there is no reference solution available for this test case.

B. Modal initialization results

In order to characterize the 3D effects present in the canyon, the three propagating modes were excited individually at the source location, and propagated outward in range. For each modal initialization, the reference sound speed was selected equal to the phase speed of the *m*th mode. Figure 21 shows the gray-scale TL plots (horizontal slices at constant depth $z=30$ m) obtained initializing the 3D PE model using mode 1, mode 2, and mode 3. Vertical cross sections of the 3D PE solutions are shown in Fig. 22. Figure 23 shows the modal ray diagrams for each omnidirectional modal initialization. Characteristics of the modal rays that connect the source *S* with one of the two arrays B1 and B2, are listed in Table II.

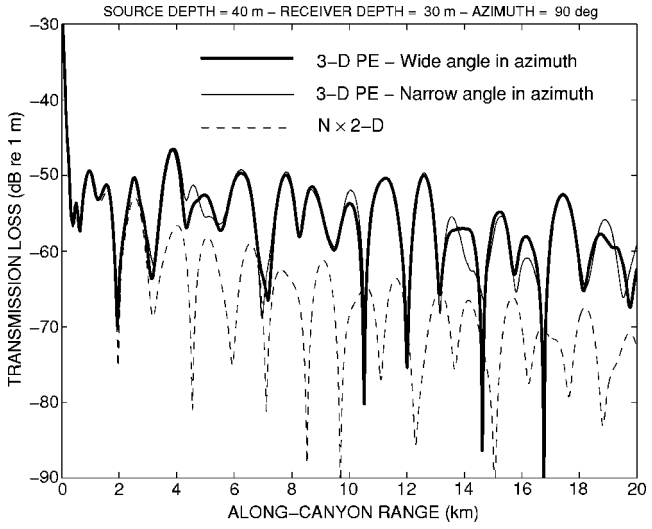


FIG. 20. $N\times 2$D and 3D transmission loss (in dB re 1 m) comparisons at 25 Hz at a receiver depth of 30 m along a 200 m isobath along the canyon axis for test case B. The thin dashed curve is a 2D PE calculation. The two solid curves are 3D PE calculations with three Padé terms in depth and with a narrow angle (thin solid curve) or a wide-angle (bold solid curve, $m_p=1$) approximation in azimuth. For each of the three calculations, the Greene's source was used as a starting field.

By comparing the results shown in Figs. 21 and 23, the solutions obtained using the PE approach and the adiabatic modal ray theory are satisfactorily in good agreement. The effects of the 3D varying bathymetry on the different modal propagations are now evident. For each propagating mode, the focusing of the energy along the canyon axis is repetitive in range and, as expected, is more pronounced for higher modes than for lower modes. Indeed, three distinct focusing zones are detected for mode 1, four distinct focusing zones for mode 2, and five distinct focusing zones for mode 3. For each propagating mode, the focusing-zone width (along the canyon axis) increases with zone number. As a result, two consecutive focusing zones may overlap at sufficiently large ranges (whose values depend on mode number) from the source along the canyon axis, and, thus, become indistinguishable. For example, for mode 1, the third focusing zone overlaps with the second one near $r\approx 19.5$ km along the canyon axis, and, for mode 2, the fourth focusing zone overlaps with the third one near $r\approx 17.5$ km along the canyon axis. Unlike test case A, no shadow zone is observed along the canyon axis. Besides, the 3D effects are more pronounced for test case B than for test case A. For example, the onset of mode 1 interference pattern along the canyon axis appears sooner in range for test case B (at $r\approx 6.5$ km) than for test case A (at $r\approx 17.5$ km).

Modes 1, 2, and 3 are present at receiver arrays B1 and

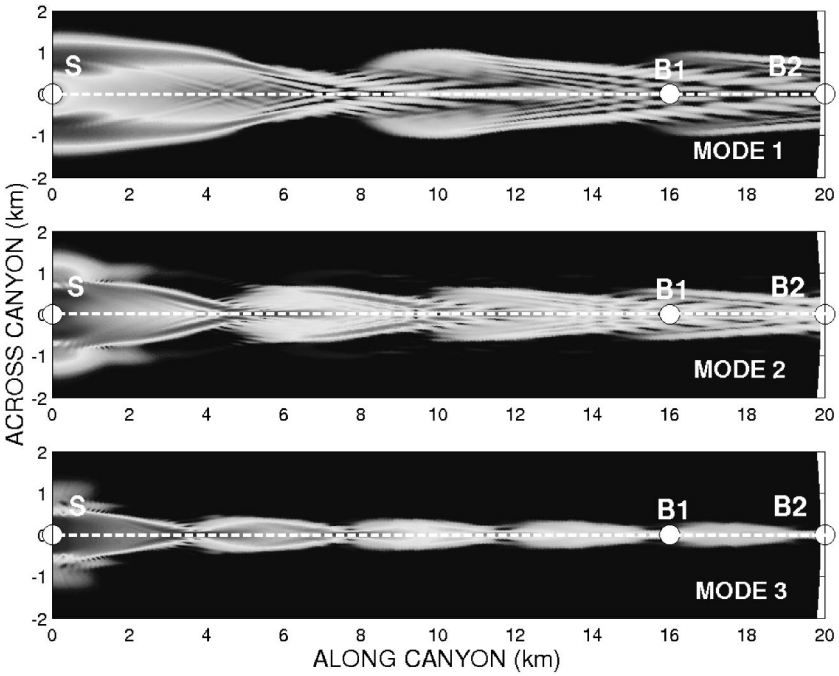


FIG. 21. Transmission loss (in dB re 1 m) at 25 Hz at a depth of 30 m for test case B corresponding to 3D PE calculations and different omnidirectional mode excitations. From top to bottom: mode 1, mode 2, and mode 3.

B2. This observation is coherent with the fact that there is no shadow zone along the canyon axis. Both receiver arrays B1 and B2 lie in the multiple mode arrival area of each of the three propagating modes, the number of eigenrays depending on both the mode number and the receiver array (see Table II). There are three modal ray paths connecting S and B1 for mode 1, the first one corresponding to $\phi_0 = 0°$ (the direct path), and the two other ones corresponding to a pair of eigenrays with initial launch angles $\phi_0 = \pm 15.4°$. Similarly, there are three modal ray paths connecting *S* and B1 for mode 2, the direct path and a pair of eigenrays with initial launch angles $\phi_0 = \pm 13.8°$. On the other hand, there are five modal ray paths connecting *S* and B1 for mode 3, the first one corresponding to the direct path, and the other ones to two distinct pairs of eigenrays with initial launch angles $\phi_0 = \pm 9.5°$ and $\phi_0 = \pm 16°$. For receiver array B2, there are one direct path and two pairs of eigenrays for each of the three propagating modes. Anticipating the analysis presented in the next section, multiple arrivals (not necessarily distinct in time) for each of the three propagating modes on both

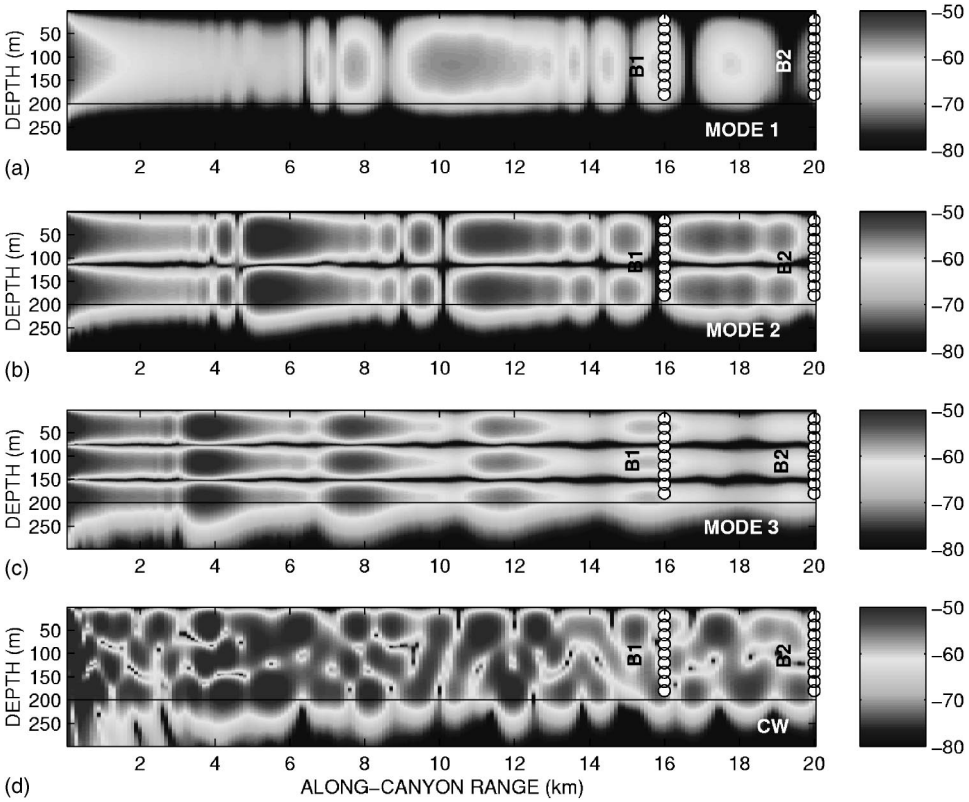


FIG. 22. 3D transmission loss (in dB re 1 m) at 25 Hz for test case B (vertical slices along the canyon axis) corresponding to 3D PE calculations and (a) mode 1, (b) mode 2, (c) mode 3, and (d) Greene's source excitations.

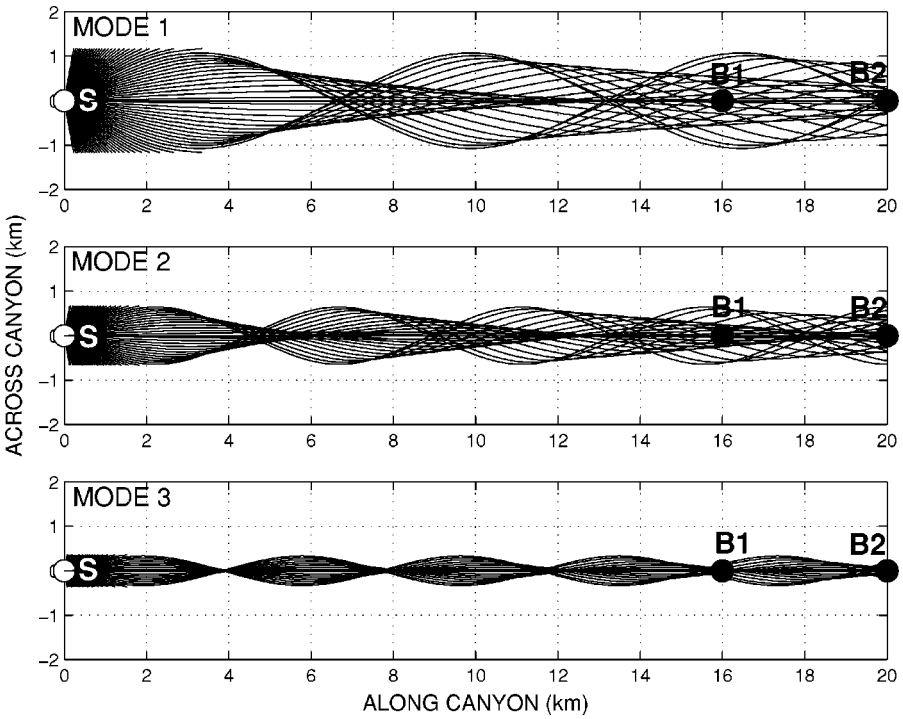


FIG. 23. Modal ray diagrams (top view) for test case B, obtained solving the differential equation given in (36) for $-90° < \phi_0 < 90°$. From top to bottom: mode 1, mode 2, and mode 3.

receiver arrays B1 and B2 can be predicted. More precisely, two mode 1 arrivals, two mode 2 arrivals, and three mode 3 arrivals are expected on B1. Three arrivals for each of the propagating modes are expected on B2. On the contrary, any azimuthally uncoupled model should predict only one single arrival (the direct path) for each of the three propagating modes.

C. Broadband results

Let us turn now to the analysis of the acoustic problem involving the broadband source signal given by Eq. (35). The signal arrivals on the two vertical arrays B1 and B2 were calculated. The calculations were carried out using $t_{\min} = 10.6$ s and $t_{\min} = 13$ s for receiver arrays B1 and B2 respectively, and, as in test case A, a time window of length $T = 7$ s with 4096 points in the DFT algorithm. This required running the 3D PE model at 281 discrete frequencies within the frequency band 5–45 Hz. For each frequency-domain calculation, $\Delta r = \lambda/6$, $\Delta z = \lambda/60$. The number M of discrete points in the azimuthal direction was obtained by linearly interpolating between $M = 576$ at $f = 5$ Hz and $M = 5184$ at $f = 45$ Hz, which gave $M = 2880$ at $f = 25$ Hz. The 3D PE marching algorithm was initialized at $r = 0$ using the Greene's source. Three Padé terms in depth (i.e., $n_p = 3$) and one Padé term in azimuth (i.e., $m_p = 1$) were used.

TABLE II. Characteristics (for a 25 Hz cw point source, test case B) of the modal rays launched from the source S and connected with one of the two arrays B1 and B2. The angle ϕ_0 denotes the initial launch angle ($\phi_0 = 0°$ points along the canyon axis) of the modal eigenray, and L_{m,ϕ_0} denotes the length of the modal eigenray path Γ_{m,ϕ_0}.

Array	Mode m	ϕ_0 (deg)	L_{m,ϕ_0} (m)
B1	1	0	16 000
B1	1	±15.4	16 360
B1	2	0	16 000
B1	2	±13.8	16 258
B1	3	0	16 000
B1	3	±9.5	16 112
B1	3	±16	16 315
B2	1	0	20 000
B2	1	±9	20 138
B2	1	±23	21 050
B2	2	0	20 000
B2	2	±2	20 005
B2	2	±16.5	20 476
B2	3	0	20 000
B2	3	±9.5	20 140
B2	3	±16	20 394

The signals received on vertical array B1 at a specific depth of 20 m corresponding to 2D and 3D computations are shown in Fig. 24. Both solutions were multiplied by the same factor $r_{B1} = 16000$ m to compensate for spherical spreading. As expected, due to the focusing of the energy along the canyon axis, the amplitude of the 3D solution is significantly higher than the amplitude of the 2D solution. The signal arrivals on vertical arrays B1 and B2 obtained using 3D calculation are displayed in Figs. 25 and 26, respectively. For comparison, the signals obtained using 2D calculation are also displayed. The two signals marked with arrows in Fig. 25 correspond to the ones shown in Fig. 24. For both vertical arrays B1 and B2, the 2D results (see first columns of Figs. 25 and 26) clearly show that the propagating signal splits up in three distinct wave packets which correspond to the signals carried by the three propagating modes of the waveguide. The 2D results are similar to that obtained in test case A. On the other hand, the modal structure of the 3D results is much more complicated (see second columns of Figs. 25 and 26). Indeed, though the first wave packet can clearly be attributed to the first arrival of mode 1, the rest of the received signals does not clearly show any modal structure. One way to discriminate the multiple arrivals of one

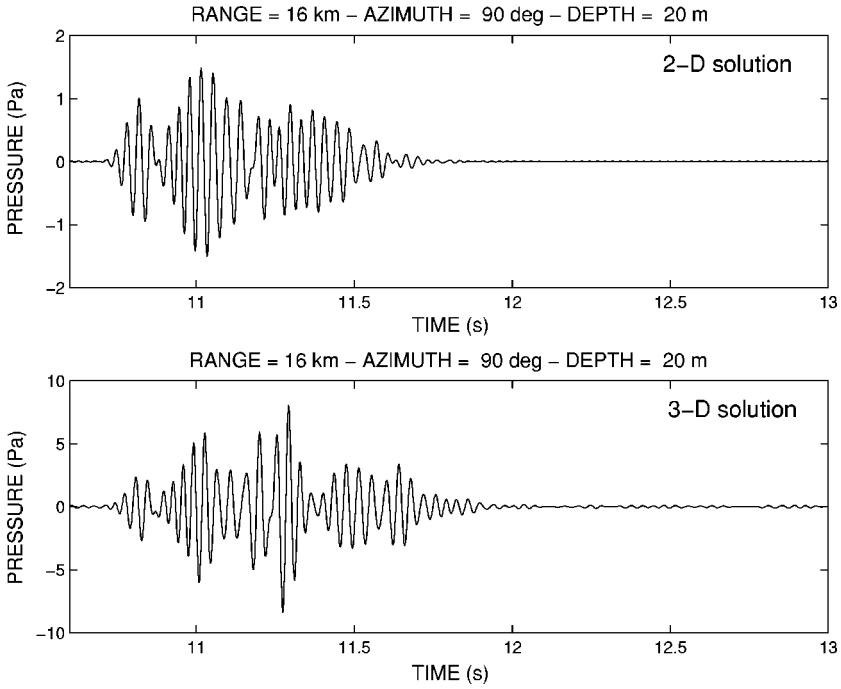


FIG. 24. Comparison of broadband pulse solutions for test case B and array B1 at a depth of 20 m: (upper) 2D calculation and (lower) 3D calculation.

mode from the multiple arrivals of the two other modes consists in initializing (in the Fourier domain) the PE model by only one single mode (if present) at each single discrete frequency, instead of initializing by the Greene's source (for which all the propagating modes are excited simultaneously). The results obtained by initializing separately by mode 1, by mode 2, and by mode 3, are displayed, respectively, in the first, second, and third columns of Figs. 27 and 28. They confirm multiple arrivals for each of the three propagating modes on receiver arrays B1 and B2. Note that multiple arrivals of the same mode can be well separated in time. For example, this is the case for the two distinct arrivals of mode 1 on B1 (see first column of Fig. 27) and for the two distinct (though very close in time but still distinguishable) arrivals of mode 2 on B1 (see second column of Fig. 27). This is also the case for the three distinct arrivals of mode 1 on B2 (see first column of Fig. 28). On the contrary, the multiple arrivals of mode 3 on B1 and B2 (see third columns of Figs. 27 and 28) are very close in time. The wave packets observed correspond to the merger of the three distinct arrivals of the same mode. They thus appear more dispersed in time than the single mode 3 arrival obtained using a 2D model. Again,

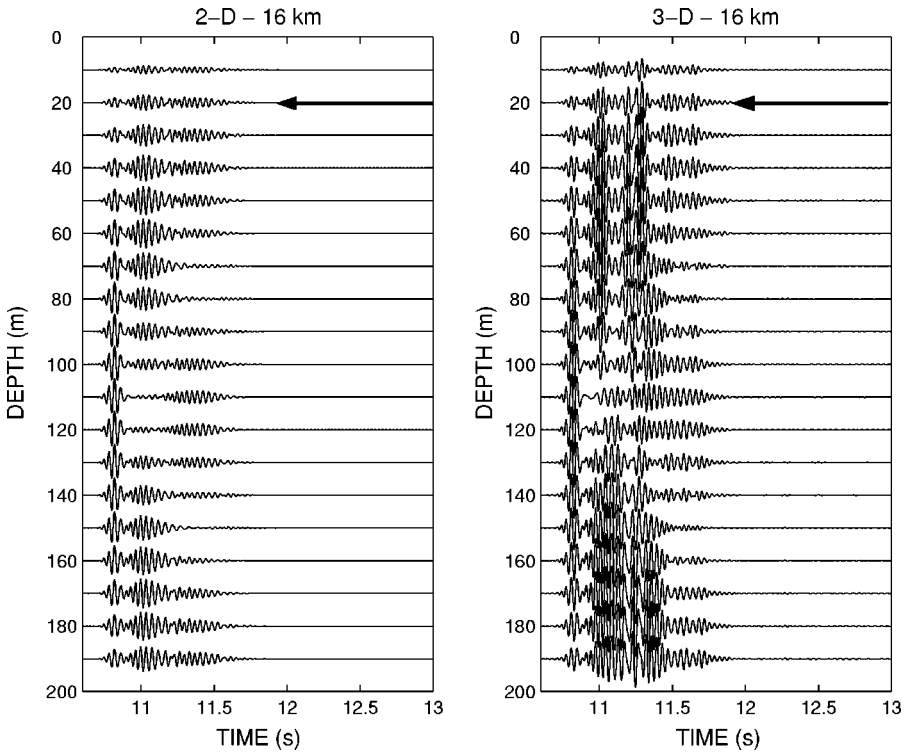


FIG. 25. Stacked time series vs depth corresponding to vertical array B1 (placed along the canyon axis at a distance of 16 km) obtained using 2D computation (left column) and 3D computation (right column).

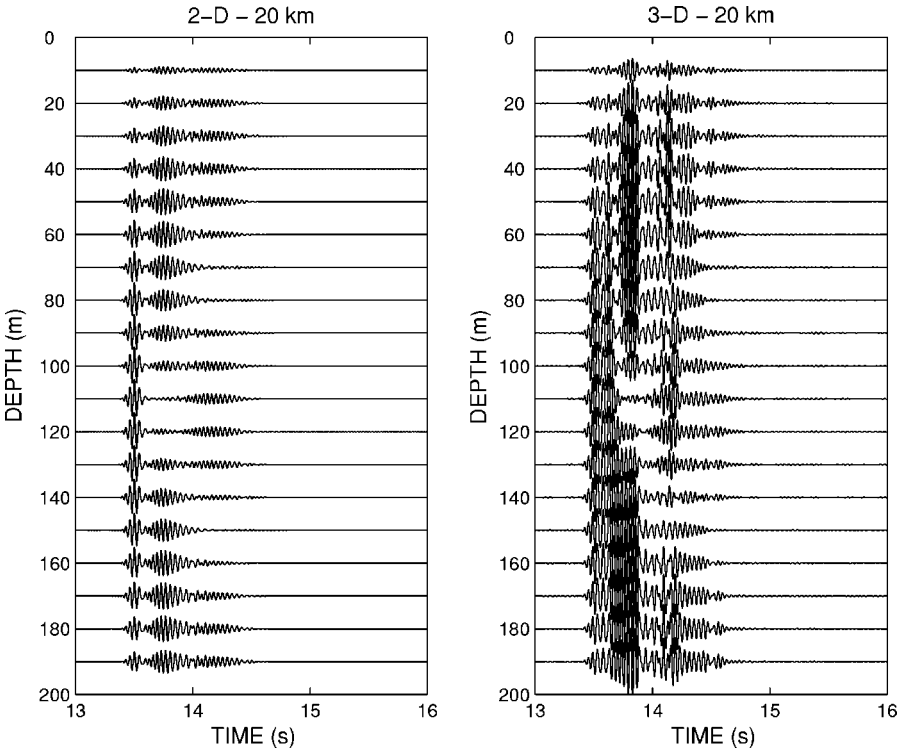


FIG. 26. Stacked time series vs depth corresponding to vertical array B2 (placed along the canyon axis at a distance of 20 km) obtained using 2D computation (left column) and 3D computation (right column).

all these observations are consistent with the predictions of the previous section.

VI. CONCLUSION

In this paper, the propagation of a broadband acoustic pulse with a central frequency of 25 Hz and a bandwidth of 40 Hz in three-dimensional shallow water waveguides was studied. The 3D ASA benchmark problem, and a variant of the original SWAM'99 Gaussian canyon test case were investigated. Both test cases were treated following the same approach. First, the acoustic problem was simplified to an harmonic point source emitting at 25 Hz. Results for both point source and modal initializations were obtained. Solutions from the 3D PE model were compared with the predictions from a 3D adiabatic modal theory. Good agreement was obtained between the two models for both test cases. For the

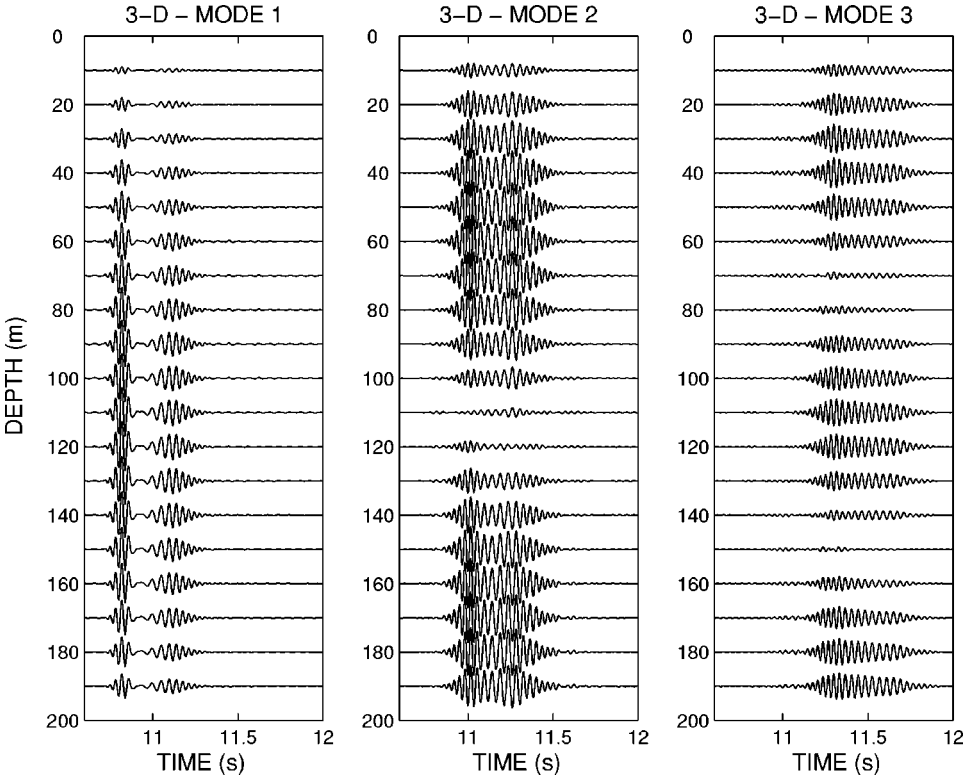


FIG. 27. Stacked time series vs depth corresponding to vertical array B1 (placed along the canyon axis at a distance of 16 km) obtained using 3D computations. The signals have been obtained initializing the PE model by mode 1 (left), by mode 2 (middle), and by mode 3 (right).

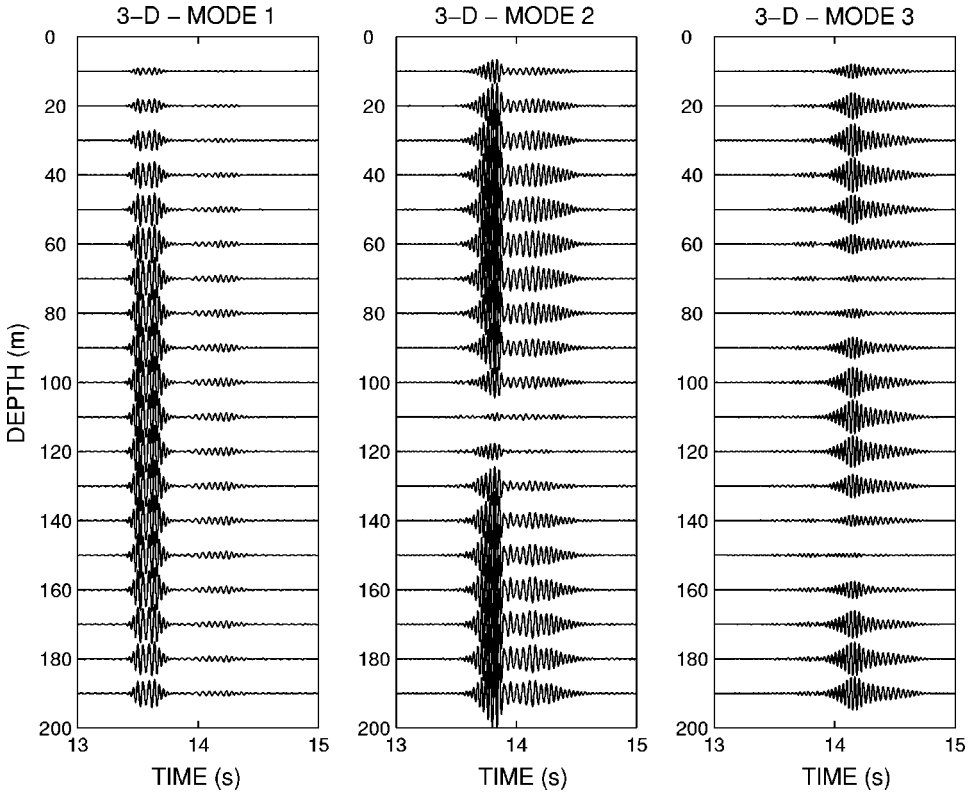


FIG. 28. Stacked time series vs depth corresponding to vertical array B2 (placed along the canyon axis at a distance of 20 km) obtained using 3D computations. The signals have been obtained initializing the PE model by mode 1 (left), by mode 2 (middle), and by mode 3 (right).

3D ASA benchmark problem, comparisons of 3D PE solutions with a reference analytical solution (based on the image method) were made, hence validating the 3D PE model for this test case. Unfortunately, no reference solution was available for the Gaussian canyon test case. The signals received by vertical arrays located far from the source were calculated. For the 3D ASA wedge problem, multiple mode arrival times (e.g., two distinct mode 1 arrivals on A2 and A3, and only one single mode 1 arrival on A1) or modal shadow zones (e.g., no mode 3 arrival on each of the vertical arrays A1, A2, and A3) were observed. For the Gaussian canyon test case, the analysis of the signals received by two vertical arrays B1 and B2 located along the canyon confirmed an increase of the acoustic level due to energy focusing along the canyon axis. Multiple mode arrivals on each of the vertical arrays for each of the three propagating modes were observed. No modal shadow zone was observed along the canyon axis. For both test cases, geometrical dispersion of the propagating modes was consistent with results obtained for a 25 Hz cw point source.

These phenomena are typical of shallow water oceanic environments in the presence of a sloping bottom. They have already been described numerically for a 3D wedge-shaped waveguide using a code based on the image source method.[16] Note that the latter code was limited to specific oceanic environments. The main advantage of using a 3D PE model is that it can be applied quite generally and is thus not restricted to a specific acoustic problem. *A priori*, the methodology that has been used to analyze the two acoustic problems, the 3D wedge and the 3D Gaussian canyon, could be applied in future for investigating any other shallow water acoustic problem. Nevertheless, the modeler should be wary when performing numerical simulations with a 3D PE model. For instance, it is important to keep in mind that any existing and available 3D PE based model is only an approximate model (though general) and is limited in its ability to handle horizontal refraction. Indeed, though the 3D PE model used in this paper, 3DWAPE, has a wide-angle capability along the azimuthal direction (comparisons for the 3D wedge seem to favor the use of the azimuthal wide-angle approximation), it does not strictly speaking have a wide-angle capability since the cross-derivative terms present in $O(\mathcal{XY})$ are not taken into account in the modeling. This point was discussed in detail in this paper. The influence of handling cross-derivative terms is not known. Efforts to incorporate these terms in 3DWAPE as well as in a 3D version of the Monterey-Miami Parabolic Equation model developed by Smith[9] are currently under way. Another issue that can be addressed when one is interested with pulse propagation in shallow water environments is the following: if shear waves are present in the bottom, do they modify the three-dimensional effects described in this paper, and if so, by how much? This question has not been addressed in this present work since the 3D model used could not handle shear waves. It is left for future works.

ACKNOWLEDGMENTS

The author wishes to thank Evan Westwood (Applied Research Laboratories, University of Austin, Texas, USA) who provided numerical data. Part of this work has been presented at the 6th International Conference on Theoretical and Computational Acoustics held in August 2003 in Hawaii. The author is very grateful to Michael Taroudakis (Department of Mathematics, University of Crete, Greece) for his

invitation to the ICTCA'03 conference and to Finn Jensen (Saclant Undersea Research Center, Italy) for his encouragement to write the present paper.

[1] R. Doolittle, A. Tolstoy, and M. Buckingham, "Experimental confirmation of horizontal refraction of CW acoustic radiation from a point source in a wedge-shaped ocean," J. Acoust. Soc. Am. **83**, 2117–2125 (1988).
[2] S. Glegg, G. Deane, and I. House, "Comparison between theory and model scale measurements of the three-dimensional sound propagation in a shear suporting penetrable wedge," J. Acoust. Soc. Am. **94**, 2334–2342 (1993).
[3] C. Chen, J.-T. Lin, and D. Lee, "Acoustic three-dimensional effects around Taiwan strait: Computational results," J. Comput. Acoust. **7**, 15–26 (1999).
[4] M. Buckingham, "Theory of three-dimensional acoustic propagation in a wedgelike ocean with a penetrable bottom," J. Acoust. Soc. Am. **82**, 198–210 (1987).
[5] G. Deane and M. 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).
[6] D. Lee, G. Botseas, and W. Siegmann, "Examination of three-dimensional effects using a propagation model with azimuth-coupling capability (FOR3D)," J. Acoust. Soc. Am. **91**, 3192–3202 (1992).
[7] G. Brooke, D. Thomson, and G. Ebbeson, "PECan: A Canadian parabolic equation model for underwater sound propagation," J. Comput. Acoust. **9**, 69–100 (2001).
[8] J. Fawcett, "Modeling three-dimensional propagation in an oceanic wedge using parabolic equation methods," J. Acoust. Soc. Am. **93**, 2627–2632 (1993).
[9] K. Smith, "A three-dimensional propagation algorithm using finite azimuthal aperture," J. Acoust. Soc. Am. **106**, 3231–3239 (1999).
[10] M. Collins and S. Ching-Bing, "A three-dimensional parabolic equation model that includes the effects of rough boundaries," J. Acoust. Soc. Am. **87**, 1104–1109 (1990).
[11] A. Tolstoy, "3-D propagation issues and models," J. Comput. Acoust. **4**, 243–271 (1996).
[12] D. Lee, A. Pierce, and E.-C. Shang, "Parabolic equation development in the twentieth century," J. Comput. Acoust. **8**, 527–637 (2000).
[13] F. Sturm and J. Fawcett, "On the use of higher-order azimuthal schemes in 3-D PE modelling," J. Acoust. Soc. Am. **113**, 3134–3145 (2003).
[14] F. Jensen and C. Ferla, "Numerical solutions of range-dependent benchmark problems in ocean acoustics," J. Acoust. Soc. Am. **87**, 1499–1510 (1990).
[15] F. Sturm, "Examination of signal dispersion in a 3-D wedge-shaped waveguide using 3DWAPE," Acta Acustica united with Acustica **88**, 714–717 (2002).
[16] E. Westwood, "Broadband modeling of the three-dimensional penetrable wedge," J. Acoust. Soc. Am. **92**, 2212–2222 (1992).
[17] A. Tolstoy, K. Smith, and N. Maltsev, "The SWAM'99 workshop—An overview," J. Comput. Acoust. **9**, 1–16 (2001).
[18] J. Arvelo and A. Rosenberg, "Three-dimensional effects on sound propagation and matched-field processor," J. Comput. Acoust. **9**, 17–39 (2001).
[19] F. Jensen, W. Kuperman, M. Porter, and H. Schmidt, *Computational Ocean acoustics*, AIP Series in Modern Acoustics and Signal Processing (AIP, New York, 1994).
[20] F. Jensen, C. Ferla, P. Nielsen, and G. Martinelli, "Broadband signal simulation in shallow water," J. Comput. Acoust. **11**, 577–591 (2003).
[21] F. Tappert, "Parabolic equation modeling with the split-step Fourier algorithm in four dimensions," J. Acoust. Soc. Am. **103**, 2090:4aUW6 (1998).
[22] M. Porter, "The time-marched fast-field program (FFP) for modeling acoustic pulse propagation," J. Acoust. Soc. Am. **87**, 2013–2023 (1990).
[23] J. Murphy, "Finite-difference treatment of a time-domain parabolic equation: Theory," J. Acoust. Soc. Am. **77**, 1958–1960 (1985).
[24] B. McDonald and W. Kuperman, "Time domain formulation for pulse propagation including nonlinear behavior at a caustic," J. Acoust. Soc. Am. **81**, 1406–1417 (1987).
[25] M. Collins, "The time-domain solution of the wide-angle parabolic equation including the effects of sediment dispersion," J. Acoust. Soc. Am. **84**, 2114–2125 (1988).
[26] M. Collins, "Applications and time-domain solution of higher-order parabolic equations in underwater acoustics," J. Acoust. Soc. Am. **86**, 1097–1102 (1989).
[27] A. Bamberger, B. Engquist, L. Halpern, and P. Joly, "Higher order paraxial wave equation approximations in heterogeneous media," SIAM (Soc. Ind. Appl. Math.) J. Appl. Math. **48**, 129–154 (1988).
[28] M. Collins, "Higher order Padé approximations for accurate and stable elastic parabolic equations with application to interface wave propagation," J. Acoust. Soc. Am. **89**, 1050–1057 (1991).
[29] F. Tappert, "The parabolic approximation method," in *Wave propagation and underwater acoustics*, Lecture notes in Physics Vol. 70, edited by J. B. Keller and J. S. Papadakis (Springer, Berlin, 1977), pp. 224–287.
[30] G. Botseas, D. Lee, and D. King, "FOR3D: A computer model for solving the LSS three-dimensional wide angle wave equation," Technical report, Naval Underwater Systems Center, Technical Document 7934, New London, CT.
[31] F. Sturm, "Modélisation mathématique et numérique d'un problème de propagation en acoustique sous-marine: prise en compte d'un environnement variable tridimensionnel," Ph.D. thesis, Université de Toulon et du Var, 1997.
[32] F. Sturm, J. Fawcett, and F. Jensen, "Benchmarking two three-dimensional parabolic equation methods," J. Acoust. Soc. Am. **103**, 2090:4aUW5 (1998).
[33] F. Sturm and J. Fawcett, "Numerical simulation of the effects of the bathymetry on underwater sound propagation using three-dimensional parabolic equation models," Technical report, SM-342, SACLANT Undersea Research Centre, La Spezia, Italy.
[34] W. Siegmann and D. Lee, "Aspects of three-dimensional parabolic equation computations," Comput. Math. Appl. **11**, 853–862 (1985).
[35] W. Siegmann, G. Kriegsmann, and D. Lee, "A wide-angle three-dimensional parabolic wave equation," J. Acoust. Soc. Am. **78**, 659–664 (1985).
[36] D. Lee and W. Siegmann, "A mathematical model for the 3-dimensional ocean sound propagation," Math. Modelling **7**, 143–162 (1986).
[37] M. Collins, "The split-step Padé solution for the parabolic equation method," J. Acoust. Soc. Am. **93**, 1736–1742 (1993).
[38] C. F. Chen, Y.-T. Lin, and D. Lee, "A three-dimensional azimuthal wide-angle model," J. Comput. Acoust. **7**, 269–288 (1999).
[39] E. Westwood, "Complex ray solutions to the 3-D wedge ASA benchmark problems," J. Acoust. Soc. Am. **109**, 2333:2aAOa6 (2001).
[40] R. Greene, "The rational approximation to the acoustic wave equation with bottom interaction," J. Acoust. Soc. Am. **76**, 1764–1773 (1984).
[41] H. Weinberg and R. Burridge, "Horizontal ray-theory for ocean acoustics," J. Acoust. Soc. Am. **55**, 63–79 (1974).

Effect of realistic grounds and atmospheric conditions on single-channel active control of outdoor sound propagation

Ann Nakashima[a] and Murray Hodgson
Department of Mechanical Engineering and School of Occupational and Environmental Hygiene, The University of British Columbia, 3rd Floor, 2206 East Mall, Vancouver, British Columbia, V6T 1Z3, Canada

(Received 17 March 2004; revised 21 December 2004; accepted 25 December 2004)

Engine run-up tests are a part of routine aircraft maintenance at the Vancouver International Airport. A source of noise complaints is the Dash-8 aircraft, which emits low-frequency, tonal noise. Active noise control is a potentially cost-effective alternative to passive noise-control methods, which are ineffective at controlling low-frequency noise. Since the run-up tests are performed outdoors, the effects of outdoor conditions on the performance of an active control system must be considered. In this paper, the results of a preliminary investigation of the effects of realistic meteorological conditions and ground impedance on the performance of a single-channel active-control system are presented. Computer simulations of single-channel active control of a monopole source were performed using the Green's-function parabolic-equation method. Different realistic atmospheric conditions, and reflective or soft ground, were used in the simulations. The results show that atmospheric refraction causes fluctuations in the noise attenuation achieved by a single-channel control system, and has the overall effect of decreasing its performance, making the system ineffective in some cases. [DOI: 10.1121/1.1859231]

PACS numbers: 43.30.Nb, 43.50.Ki [KAC] Pages: 1080–1087

I. INTRODUCTION

At the Vancouver International Airport (YVR), aircraft engine run-up tests are a source of noise complaints from residents living in neighboring communities. The engine run-up tests, which are a part of normal aircraft maintenance, sometimes involve revving the engines to full power. The resulting noise includes low-frequency components that are heard in residential communities located several kilometers away from the airport. One source of run-up noise complaints is the Dash-8 aircraft.[1] Measurements from a Dash-8 aircraft indicate that the full-power engine noise is dominated by tonal, low-frequency components starting at 80 Hz.[2] Low-frequency noise cannot be controlled effectively with traditional run-up noise-control methods such as hush-houses or ground run-up enclosures. It is thus of interest to investigate alternate methods of noise control.

Active noise control (ANC) has been the focus of a number of studies in recent years because of its effectiveness in controlling low-frequency noise. Most practical applications of ANC have been for use in headphones or small enclosures; the concept of ANC in open space is still at the research stage. The most desirable result of ANC in open space would be the attenuation of noise at all points in space; this is known as global control. However, significant global control can only be achieved when the secondary sources are placed less than one half-wavelength away from the primary noise source at the frequency of interest.[3] In practical applications, such as the control of aircraft propeller noise, it is not possible to place the secondary sources at such a short distance from the primary source. For these cases, it seems that it is only possible to achieve sound attenuation in localized areas (quiet zones). This is known as the local-control method.[3] For local control, it is possible to achieve quiet zones with the secondary sources placed at several wavelengths away from the primary source at the frequency of interest, as long as the spacing of the control channels is optimized.[3,4]

The effectiveness of the local-control method using multiple control channels has been investigated in a free-field and a half-space by simulations and experiments. The control of monopole[3,4] and dipole[5] primary sources has been studied. A preliminary investigation into the feasibility and effectiveness of using local control to attenuate propeller-aircraft noise was performed by Hodgson *et al.*[6] The study involved computer simulations of multi-channel, local active-control systems in the free-field or over a half-space (i.e., in the presence of a totally absorptive or reflective ground) with a nonrefractive atmosphere. It was found that a significant amount of attenuation could be achieved over a wide range of angles, for monopole and dipole primary sources, when a large number of control channels were used. However, because the simulations did not account for atmospheric refraction or finite ground impedance, they did not represent realistic outdoor conditions. The work presented here is an extension of the Hodgson *et al.* study to investigate the effects of outdoor conditions on the performance of an active local control system. This was done with the intention of estimating the amount of noise attenuation that could be achieved for Dash-8 run-up noise at a receiver position in a community located 3 km north of the YVR run-up site.

A number of prediction methods for outdoor sound

[a]Current address: Human Factors Research and Engineering Section, Defense R&D Canada—Toronto, 1133 Sheppard Avenue West, P.O. Box 2000, Toronto, ON M3M 3B9, Canada.

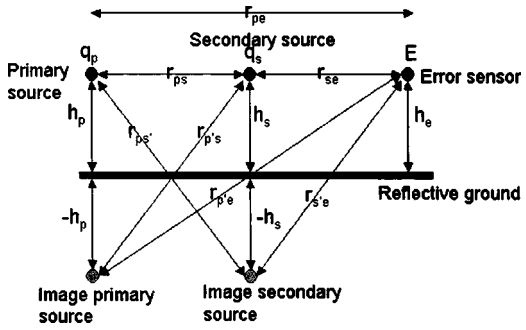


FIG. 1. Single-channel ANC system in a half-space.

propagation have been developed over the years. It is generally difficult to validate the results of an outdoor prediction model experimentally, due to the unpredictable behavior of turbulence and atmospheric inhomogeneities that exist in a real outdoor environment. The models have historically been validated by comparison of the predicted results to those of other models, as was done by Attenborough *et al.* for several benchmark cases.[7] In order to incorporate the prediction of active local control into an outdoor model, the model must be capable of predicting the sound field in both amplitude and phase. One model with this capability, which was validated in Ref. 7, is the Green's function parabolic equation method (GFPE). The GFPE is a computationally efficient algorithm for calculating one-way wave propagation in a refractive atmosphere above a locally-reactive ground surface. The theoretical formulation of the GFPE model is given in Ref. 8.

For the purpose of studying the effect of atmospheric and ground conditions on the performance of a local control system, it was of interest to modify the GFPE code to account for control sources. While multi-source and multi-pole outdoor sound-propagation models have been developed,[9,10] and multi-channel active-control simulations have been performed for wind in a horizontal plane,[11] to the authors' knowledge there have been no attempts to predict the active control of long-range sound propagation in a refractive atmosphere. In this paper we discuss some preliminary results of active control under realistic meteorological conditions, using the GFPE with a single-channel local-control system. Although is it clear from the work done by Hodgson *et al.*[6] that a single-channel active control system is insufficient to provide significant noise attenuation over a large area, the specific objective here is to investigate whether realistic meteorological conditions and grounds affect the performance of a local control system. As the algorithm complexity and processing time increases significantly when using a 3-dimensional model useful in studying multi-channel control, it is of interest, as a first step, to investigate atmospheric and ground effects on single-channel control in one-dimension.

II. THEORY

A. Local control system in a half-space

A single-channel, local active-control system above a reflective plane is shown in Fig. 1. The strengths of the primary and secondary sources are given by q_p and q_s, E is the position of the error sensor and r_{ps}, r_{pe}, and r_{se} are the distances between the primary and secondary sources, the primary source and the error sensor, and the secondary source and the error sensor, respectively. The contribution of the reflective plane to the sound field is accounted for by image primary and secondary sources. The distances from the primary image source to the secondary source, the secondary image source to the primary source, the primary image source to the error sensor and the secondary image source to the error sensor are given, respectively, by $r_{p's}$, $r_{s'p}$, $r_{p'e}$, and $r_{s'e}$. The heights of the primary source, secondary source and error sensor above the plane are equal, such that $h_p = h_s = h_e$.

The acoustic transfer impedances between the primary source and error sensor, and between the secondary source and error sensor are

$$Z_{pe} = iZ_0\left(\frac{\exp(-ikr_{pe})}{kr_{pe}} + C_r\frac{\exp(-ikr_{p'e})}{kr_{p'e}}\right), \tag{1}$$

$$Z_{se} = iZ_0\left(\frac{\exp(-ikr_{se})}{kr_{se}} + C_r\frac{\exp(-ikr_{s'e})}{kr_{s'e}}\right), \tag{2}$$

where k is the wavenumber and C_r is the reflection coefficient of the plane.[4] To minimize the sound pressure at the error microphone, the required strength of the control source is

$$q_s = Aq_p, \tag{3}$$

where $A = -Z_{pe}/Z_{se}$. The total sound pressure of the system at any point above the reflective plane is

$$P = Z_{pl}q_p + Z_{sl}q_s, \tag{4}$$

where

$$Z_{pl} = iZ_0\left(\frac{\exp(-ikr_{pl})}{kr_{pl}} + C_r\frac{\exp(-ikr_{p'l})}{kr_{p'l}}\right), \tag{5}$$

$$Z_{sl} = iZ_0\left(\frac{\exp(-ikr_{sl})}{kr_{sl}} + C_r\frac{\exp(-ikr_{s'l})}{kr_{s'l}}\right). \tag{6}$$

In Eqs. (4) to (6), the subscript l denotes the position of the observation point in the half-space. The distances from the observation point to the primary source, the primary image source, the secondary source and the secondary image source, are given by r_{pl}, $r_{p'l}$, r_{sl}, and $r_{s'l}$, respectively.[4] The controlled sound-pressure field at a given point in space [Eq. (4)] is a function of the transfer impedances, Z_{pe} and Z_{se}, given by Eqs. (1) and (2). The performance of the control system in a given environment thus depends on the distance of the primary source to the error sensor, r_{pe}, and the distance of the secondary source to the error sensor, r_{se}.

B. Green's function parabolic equation model

The GFPE model for outdoor sound propagation was developed by Gilbert and Di.[8] The theoretical formulation of the model is presented in Ref. 8; only the final equations will be shown here. The sound source is placed at a height z_s, and the starting field, or the sound field at a distance $r=0$, is calculated by

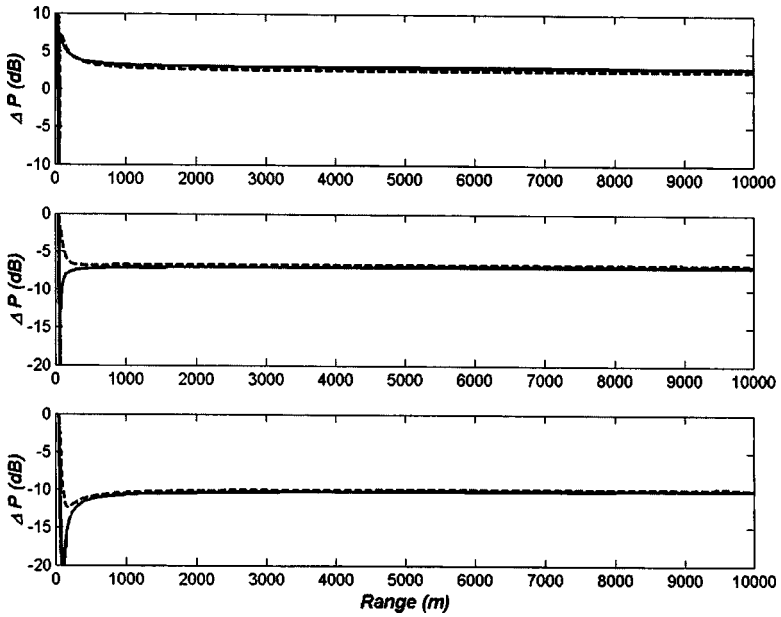


FIG. 2. Single-channel ANC results for the Guo and Pan image-source model (solid line) and the GFPE model (dashed line) for a monopole primary source at 80 Hz. Top: $r_{ps}=30$ m, $r_{se}=20$. Middle: $r_{ps}=r_{se}=30$ m. Bottom: $r_{ps}=30$ m, $r_{se}=60$ m.

$$\Psi(r=0,z)=\left\{\exp\left[-\left(\frac{z-z_s}{\sigma}\right)^2\right]+\frac{Z_g-1}{Z_g+1}\times\exp\left[-\left(\frac{z+z_s}{\sigma}\right)^2\right]\right\}(\sqrt{\sigma})^{-1}, \tag{7}$$

where $\sigma=2/k_r$, k_r is the wavenumber with respect to the sound speed at the ground, and Z_g is the complex ground impedance. The starting field is calculated at equally-spaced values of z, dividing the atmosphere into horizontal layers. The sound speed is constant within each layer. An artificial sound-absorptive layer is placed at an arbitrarily defined height, $z=z_{\text{top}}$, from which the sound waves are not reflected downwards. The horizontal distance r is then advanced by a range step Δr, and the field is calculated as

$$\Psi(r+\Delta r)=\exp\left[i\Delta r\left(\sqrt{\frac{\partial^2}{\partial z^2}+k_r^2+\frac{\delta k^2(z)}{2k_r}}\right)\right]\Psi(r), \tag{8}$$

where δk is the variation of k with z (small compared to k). The layering of the atmosphere into horizontal strips of equal height Δz allows for a nonconstant sound-speed profile, thus simulating atmospheric refraction.

III. VALIDATION OF THE GFPE MODEL

The GFPE model developed by Gilbert and Di[8] was used in this study. This implementation of the model allows for a one-dimensional prediction of the sound field of a monopole source, calculated at some receiver height, or a two-dimensional prediction in a vertical plane. Because the GFPE uses a parabolic-equation approximation of the wave equation, the results are only accurate for sound waves leaving the source at propagation angles of less than about 30°. However, as shown by the validation cases of Gilbert and Di[8] and Attenborough *et al.*,[7] the results are as accurate as other prediction models when the source and receiver are close to the ground (the results shown in Refs. 7 and 8 were for a source height of 5 m and a receiver height of 1 m). The GFPE results can be inaccurate close to the source because of the Gaussian starting field that is used [Eq. (7)], but again, as shown in Refs. 7 and 8, the results converge to those of other prediction models at distances more than a few wavelengths away from the source at the frequency of interest.

In the GFPE model, atmospheric refraction is calculated from a vertical sound-speed gradient that is defined by the user. The choice of layer thickness, Δz, range step, Δr, and height of the absorptive layer, z_{top}, are arbitrary, and must be adjusted for a given frequency to obtain accurate results. In the present work, using the benchmark cases described by Attenborough *et al.*,[7] values of the parameters were used which gave results which matched those shown in Ref. 7. The frequency of interest was 80 Hz, which is the blade-passing frequency of the Dash-8 aircraft;[2] however, Ref. 7 only shows predictions for 10, 100, and 1000 Hz. The Gilbert and Di GFPE model was thus validated at 100 Hz, and the parameters were adjusted slightly for use at 80 Hz. The parameter values that were chosen were $\Delta z=0.4$ m, $\Delta r=10$ m and $z_{\text{top}}=200$ m.

The GFPE code was modified to include a single active-control channel, using the transfer impedances given by Eqs. (1) and (2), and the secondary source strength given by Eq. (3). The model was validated by comparing the results obtained using the GFPE model with the results calculated using Eq. (4) for the total sound pressure above a reflective plane. Single-channel ANC for a monopole primary source was predicted to a horizontal distance of 10 km, using source and receiver heights of 4.5 m (the height of the center of the Dash-8 propellers[2]). For all of the predictions, the control source and error sensor were positioned in a line at the height of the primary source.

It was found that the ANC predictions using the GFPE method were sensitive to the reference sound speed (the sound speed at the ground). At 80 Hz, the GFPE results agreed with the results obtained using Eq. (4) when a reference sound speed of 343 m/s was used. When a reference speed of 340 m/s was used, the control results differed by as

TABLE I. Sound-speed profiles used in the GFPE model relative to the ground sound speed for temperature lapse and weak, medium, and strong inversion conditions.

	Relative sound speed (m/s)			
Height (m)	Weak inversion	Medium inversion	Strong inversion	Lapse
1				−2.97
10	+0.1	+0.1	+0.2	
100	+0.5	+1.1	+2.1	−6.0
200	+1.1	+2.1	+4.2	−6.0
300	+1.6	+3.2	+6.3	−6.0
350	+1.9	+3.7	+7.4	−6.0

much as 10 dB, although the amplitude predictions for the primary source and the secondary source separately did not change. The phase predictions of the GFPE thus appear to be quite sensitive at low frequencies.

In the simulations, r_{ps} was the distance from the primary source to the secondary (control) source, and r_{se} was the distance from the secondary source to the error sensor. Sample validation results for 80 Hz are shown in Fig. 2 for $r_{ps}=30$ m, $r_{se}=20$ m, $r_{ps}=r_{se}=30$ m and $r_{ps}=30$ m, $r_{se}=60$ m. The GFPE results do not match the result given by Eq. (4) in the near field, an error that was expected, due to the Gaussian starting field used by the GFPE. When r_{ps} is held constant, the performance of the control system seems to improve as r_{se} is increased. The optimum value for r_{se} when $r_{ps}=30$ m was $r_{se}=60$ m, at which approximately 10 dB of attenuation was achieved. The results support the conclusion of Hodgson *et al.* that a quiet zone (>10 dB attenuation) can be achieved when $r_{se}\geqslant 2r_{ps}$.[6]

IV. SINGLE-CHANNEL-CONTROL RESULTS

The effects of ground surface, temperature gradients and wind-speed gradients on the effectiveness of a single-channel control system were studied separately and in combination. The following atmospheric conditions were modeled:

(i) stationary atmosphere;
(ii) weak, medium, and strong temperature inversions;
(iii) temperature lapse;
(iv) weak, medium, and strong temperature inversions with downwind;
(v) temperature lapse with upwind.

For each atmospheric condition, predictions were made for reflective ground and for the finite-impedance ground used in the benchmark cases reported in Ref. 7. The temperature-inversion conditions were defined as an increase of 0.9 °C/100 m for weak inversion, 1.8 °C/100 m for medium inversion and 3.6 °C/100 m for strong inversion.[12] The GFPE model uses an input file, in which the sound speed is given at different heights, to calculate the sound-speed gradient. The sound speeds that were input for the various temperature gradients are listed in Table I.

The wind-speed gradient at height z was implemented using an effective sound-speed profile defined by

$$c_{\text{eff}}(z)=c_0+b\ln\left(\frac{z}{z_0}+1\right), \tag{9}$$

where c_0 is the sound speed at the ground, z_0 is the roughness length of the ground surface in meters, and b is a parameter describing refraction, in meters per second.[13] For the results shown here, $c_{\text{eff}}(z)$ was calculated using $b=1$ m/s for downwind, $b=-1$ m/s for upwind, $z_0=10^{-4}$ m for hard ground and $z_0=10^{-2}$ m for soft ground. The downwind sound-speed profile was combined with the temperature-inversion conditions to simulate maximal downward refraction. Similarly, the upwind sound-speed profile was combined with the temperature-lapse conditions to simulate maximal upward refraction. The resulting sound speeds input to the GFPE model are summarized in Table II.

The single-channel active-control simulations were run using $r_{ps}=30$ m, $r_{se}=60$ m for 80 Hz. Simulations using the same configuration were also performed for 160 Hz, but the results are not shown here for the sake of brevity (see Ref. 14). As in the validation simulations, results were calculated to a horizontal distance of 10 km, using source and receiver heights of 4.5 m.

A. Effect of temperature gradients

The results for temperature-inversion and temperature-lapse conditions above reflective ground are shown in Fig. 3. In the case of no refraction, approximately 10 dB of attenuation was achieved in the far field. The introduction of a slight downward refraction, caused by a weak temperature inversion, did not significantly affect the average amount of attenuation created by the control system, but introduced fluctuations in the attenuation beyond about 7 km as shown in Fig. 3(a). The spatial variations are caused by the interfer-

TABLE II. Sound-speed profiles used in the GFPE model relative to ground sound speed for temperature lapse with upwind and weak, medium, and strong inversion conditions with downwind, for hard and soft ground.

	Relative sound speed (m/s)							
Height	Weak inversion		Medium inversion		Strong inversion		Lapse	
(m)	Hard	Soft	Hard	Soft	Hard	Soft	Hard	Soft
1							−12.2	−7.6
10	+11.6	+7.0	+11.6	+7.0	+11.7	+7.1		
100	+14.4	+9.7	+14.9	+10.3	+15.9	+11.3	−19.8	−15.2
200	+15.6	+11.0	+16.6	+12.0	+18.7	+14.1	−20.5	−16.9
300	+16.6	+11.9	+18.1	+13.5	+21.2	+16.6	−20.9	−17.3
350	+16.9	+12.3	+18.8	+14.2	+22.4	+17.8	−21.0	−17.4

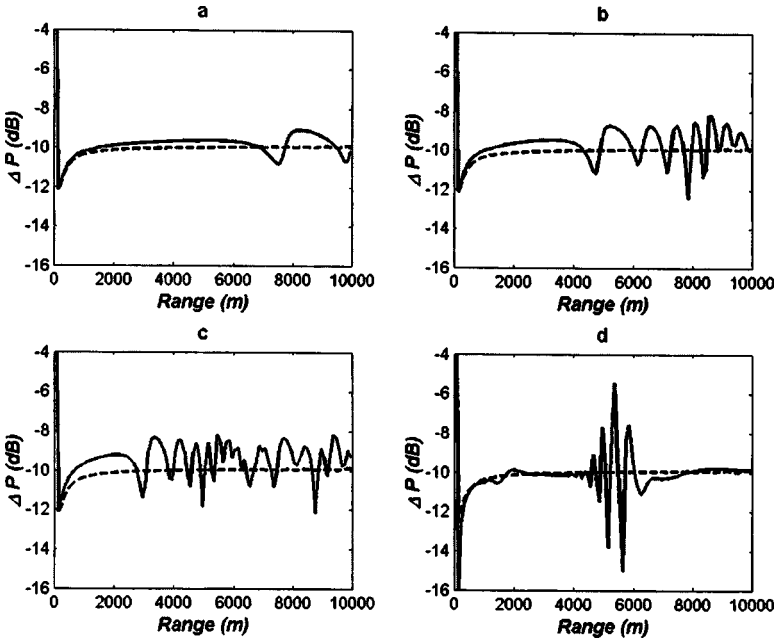


FIG. 3. Single-channel ANC results for a monopole primary source at 80 Hz in the presence of temperature gradients above reflective ground. The nonrefractive case is shown by the dashed line. (a) Weak inversion, (b) medium inversion, (c) strong inversion, (d) lapse.

ence patterns between the direct and reflected waves. For stronger inversion conditions, the sound waves were refracted downwards more forcefully. This caused the fluctuations to occur closer to the noise source, and to increase in amplitude (± 2 to 3 dB). The temperature lapse caused large fluctuations in the middle range (4 to 6 km).

B. Effect of temperature gradients over soft ground

Figure 4 shows the results for temperature inversion and lapse conditions over soft ground. In the nonrefractive case, the control system achieved about 10 dB of attenuation, the same amount as for reflective ground. The effect of the soft ground is only seen in the far field, where small fluctuations occurred. The introduction of temperature-inversion conditions again caused fluctuations in the attenuation but, unlike the cases with reflective ground, the average attenuation decreased with increasing distance from the source. In the case of downward refraction, soft ground appears to decrease the performance of the control system in the far field. For the temperature-lapse condition, the result for the temperature lapse is similar to that with reflective ground, with large fluctuations in the middle of the range.

C. Effect of temperature and wind-speed gradients

Figure 5 shows the results for temperature inversion with downwind conditions, as well as for temperature lapse with upwind conditions, in the presence of reflective ground. The results for the different degrees of temperature inversion are almost identical, showing about 3 dB of attenuation along the entire range. This indicates that the effects of a strong downwind dominate the temperature-inversion ef-

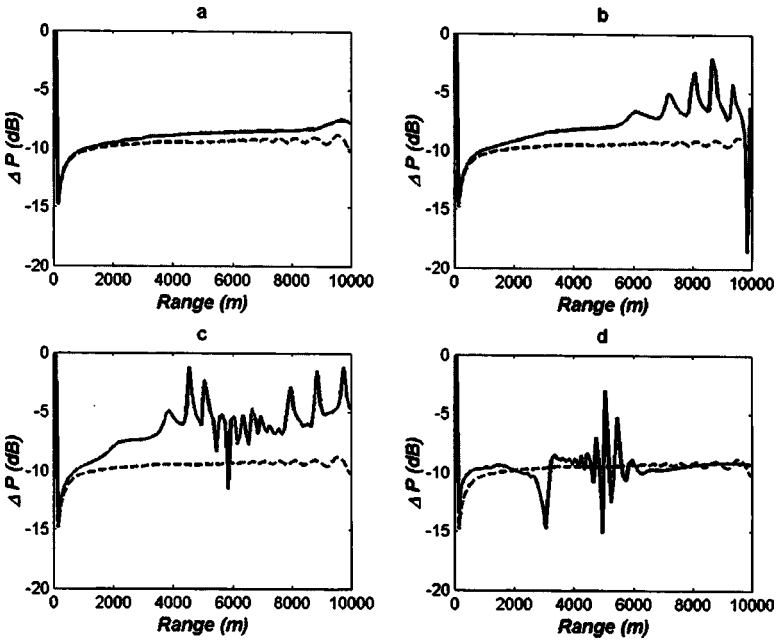


FIG. 4. Single-channel ANC results for a monopole primary source at 80 Hz in the presence of temperature gradients above soft ground. The nonrefractive case is shown by the dashed line. (a) Weak inversion, (b) medium inversion, (c) strong inversion, (d) lapse.

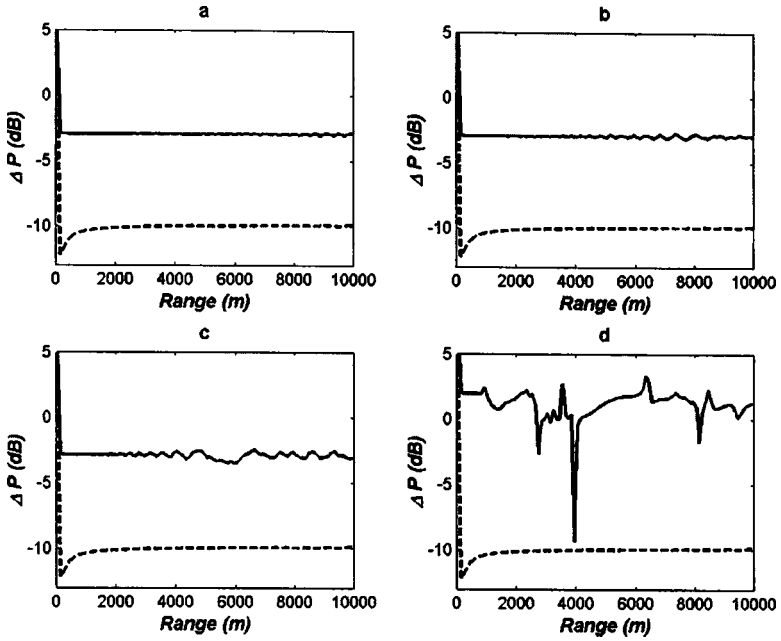


FIG. 5. Single-channel ANC results for a monopole primary source at 80 Hz in the presence of temperature and wind-speed gradients above reflective ground. The nonrefractive case is shown by the dashed line. (a) Weak inversion, (b) medium inversion, (c) strong inversion, (d) lapse.

fects. The control system is ineffective under temperature-lapse conditions in the presence of an upwind; the sound field increased by about 1 dB on average.

D. Effect of temperature and wind-speed gradients over soft ground

Figure 6 shows the results for the most complex conditions, combining temperature and wind-speed gradients with soft ground. For the downwardly refractive cases, about 5 dB of attenuation is achieved close to the source but, beyond about 1 km, the control system is ineffective. A small amount of attenuation (2–3 dB) was produced in the upwardly refractive case, except at a few points along the range where fluctuations occurred.

V. DISCUSSION

Clearly, atmospheric refraction and complex-impedance ground have significant effects on the performance of a single-channel control system. It is most difficult to achieve attenuation under temperature-lapse conditions, especially in the presence of an upwind. However, under such conditions, the noise levels near the ground have already been attenuated significantly due to upward refraction of the sound waves. It is thus more important to concentrate on attenuating the noise under temperature-inversion and downwind conditions. Overall, it seemed that atmospheric refraction had the effect of causing fluctuations in the attenuation, while the presence of soft ground decreased the overall performance of the control system in the far field.

The approximate total sound-power output increase of

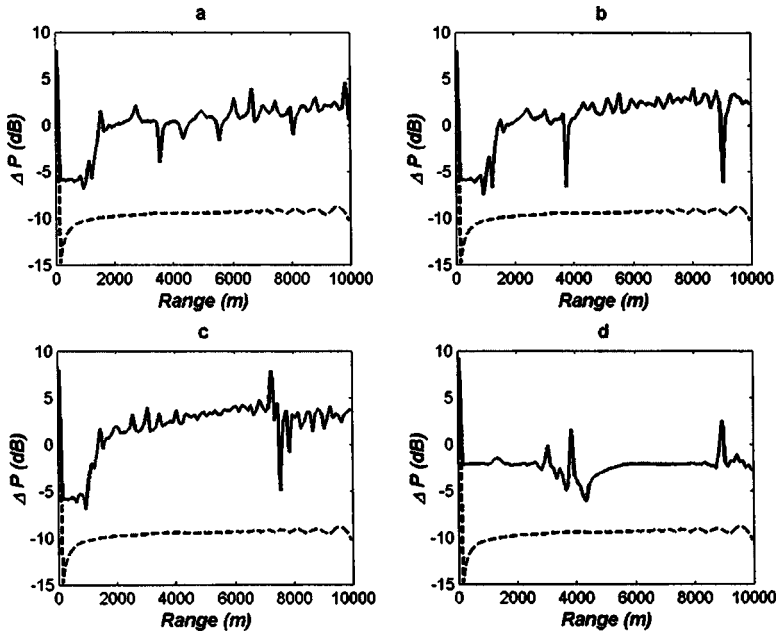


FIG. 6. Single-channel ANC results for a monopole primary source at 80 Hz in the presence of temperature and wind-speed gradients above soft ground. The nonrefractive case is shown by the dashed line. (a) Weak inversion, (b) medium inversion, (c) strong inversion, (d) lapse.

TABLE III. Approximate total sound power output increase of a single-channel control system, and total transmission loss including ANC, at a distance of 3 km for different grounds and atmospheric conditions.

Atmospheric conditions	Ground type	ΔP (dB)		Transmission loss (dB)	
		80 Hz	160 Hz	80 Hz	160 Hz
No refraction	Hard	−10.5	−7.0	−72.5	−71.0
	Soft	−10.0	−9.0	−93.5	−108.7
Weak inversion	Hard	−10.0	−7.0	−68.1	−62.8
	Soft	−10.0	−10.5	−92.2	−110.4
Medium inversion	Hard	−9.5	−7.0	−63.0	−60.2
	Soft	−8.5	−12.5	−85.7	−91.0
Strong inversion	Hard	−11.0	−7.5	−69.6	−59.7
	Soft	−7.5	−12.0	−77.6	−87.8
Lapse	Hard	−10.5	+2.0	−107.4	−106.8
	Soft	−13.0	0.0	−127.0	−124.2
Weak inversion+ downwind	Hard	−3.0	+6.0	−52.1	−41.5
	Soft	0.0	+5.0	−99.5	−63.6
Medium inversion+ downwind	Hard	−3.0	+6.0	−52.1	−41.5
	Soft	0.0	+6.0	−109.0	−60.4
Strong inversion+ downwind	Hard	−3.0	+7.0	−52.2	−39.4
	Soft	+1.0	+7.0	−98.5	−58.4
Lapse+upwind	Hard	0.0	+7.0	−138.8	−153.4
	Soft	−4.0	+3.0	−151.0	−137.1

the single-channel control system at the distance of 3 km, at 80 and 160 Hz, are listed in Table III. This table summarizes the amount of attenuation that can be expected at a community located 3 km away from the airport for the various environmental conditions. Transmission-loss values (i.e., the decibel level at a field point with respect to the pressure 1 m from the source), including the attenuation achieved by the ANC system, are also listed. A significant amount of attenuation (7.0 to 13.0 dB) can be expected at 80 and 160 Hz when atmospheric refraction is relatively weak—that is, when there is only refraction due to temperature gradients. For these cases, the transmission loss was already significant before the attenuation achieved by ANC was taken into account, so ANC may not be necessary.

In the presence of a downwind or upwind, the system achieves little or no attenuation at 80 Hz, and increases the noise by several dB at 160 Hz. In the case of an upwind, the transmission loss is already significant, so ANC is not required. Under downwind conditions, however, the 80 Hz signal only decays by about 50 dB in the case of hard ground. It has been shown that the tonal noise at 80 Hz from the Dash-8 engines when operating at full power can be as high as 120 dB[2]. Thus, a 70 dB signal would likely be heard at a community 3 km away. The 160 Hz noise would also likely be audible, as the transmission loss is only about 40 dB for hard ground. For soft ground, the transmission loss is significant at 80 Hz, indicating that strong destructive interference occurs at 3 km. The losses are less significant at 160 Hz (about 60 dB). It is clear that noise reduction is needed under downwind conditions, particularly in the case of hard ground.

VI. CONCLUSION AND FUTURE WORK

Preliminary simulations were performed to predict the effect of meteorological conditions and ground impedance on the performance of a single-channel active-noise-control system. The simulations were done in the interest of controlling tonal noise from a Dash-8 aircraft during engine run-up tests, with the targeted quiet zone being a community located 3 km to the north of the Vancouver International Airport.

A Green's-function parabolic-equation (GFPE) model was used. The code was modified to include a single active-control channel, and validation work was performed by comparing the results to those of a local ANC model in a half-space.[4] Predictions were done over a horizontal range of 10 km using the GFPE for a number of realistic atmospheric conditions, using reflective or soft ground. Atmospheric refraction tended to cause fluctuations in the attenuation, and decrease the overall performance of the control system. Little or no attenuation was achieved under temperature-lapse conditions, and soft ground decreased the average amount of attenuation in the far field.

The single-channel control results at a distance of 3 km were summarized to show the expected performance of the control system at a residential community near YVR. Significant attenuation can be achieved in the case of weak atmospheric refraction (temperature gradients only, with no wind) for both reflective and soft grounds. However, the system is ineffective in the presence of wind, particularly at 160 Hz. It is evident that a single-channel control system is insufficient to achieve significant attenuation of Dash-8 run-up noise under realistic meteorological conditions.

These simulation results show that a quiet zone (>10 dB of attenuation) cannot be achieved by a single-channel control system under realistic outdoor sound-propagation conditions. Results of previous work showed that greater attenuation can be obtained using a multi-channel active control system.[6] Thus, it is of interest to investigate the effect of realistic grounds and atmospheric conditions on multi-channel active control. This would involve including multiple control channels in the simulations. However, implementing an array of control channels into the

GFPE model might require extending the model to three dimensions.

The GFPE method that was used in this study is only valid for sound waves leaving the source at propagation angles of less than about 30°. Thus, predictions can only be made for a monopole source, as dipole and quadrupole radiation cannot be accurately modeled. Furthermore, although it is possible to predict the sound field due to multiple monopole noise sources, all of the sources must lie in the vertical (x-z) plane. Since it is not possible to model the contribution of sources lying in the horizontal plane, only single-channel control can be implemented. Salomons[8] described equations for a three-dimensional GFPE model, but there was no indication that the model has been successfully implemented numerically.

A Fast-Field Program (FFP) model for the outdoor sound propagation of dipole and quadrupole sources has been developed.[9,10] However, contributions from off-axis sources (i.e., from an array of control channels) cannot be included in the existing model—as was the case with the GFPE model, only single-channel control can be implemented. In addition, the FFP model is limited in that it cannot model range-dependent factors such as irregular terrain and turbulence.[13]

In order to predict the effectiveness of a multi-channel active-control system on a directional noise source under complex meteorological conditions, a fully three-dimensional model is desirable. In the interest of predicting the performance of an active-control system on the Dash-8 run-up noise, it would be best to extend the FFP model to predict a three-dimensional solution, so that multi-channel control could be implemented. While the model may have some limitations, it has the advantage over the GFPE model that directional noise sources can be modeled.

ACKNOWLEDGMENTS

This research was funded by the Vancouver International Airport Authority and the Natural Sciences and Engineering Research Council of Canada. The authors also wish to acknowledge Dr. Xiao Di of Pennsylvania State University for providing the GFPE code and Dr. Mike Stinson of the National Research Council, Dr. Keith Attenborough of Hull University and Dr. Sharam Taherzadeh of the Open University for their technical support.

[1] Vancouver International Airport Authority, ''YVR Aeronautical Noise Management—Annual Report,'' 1997.
[2] A. Nakashima and M. Hodgson, ''Measurement of Dash-8 run-up noise for local active noise control,'' on the CD-ROM *Proceedings of the 32nd International Congress and Exposition on Noise Control Engineering*, Internoise 2003, Paper No. N445, 2002.
[3] 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).
[4] J. Guo and J. Pan, ''Effects of reflective ground on the actively created quiet zones,'' J. Acoust. Soc. Am. **103**, 944–952 (1998).
[5] A. Nakashima, J. Guo, and M. Hodgson, ''Preliminary investigation of active control for dipole noise sources,'' on the CD-ROM *Proceedings of the 2002 International Symposium on Active Control of Sound and Vibration*, ACTIVE 2002, 2002.
[6] M. Hodgson, J. Guo, and P. Germain, ''Active local control of propeller-aircraft run-up noise,'' J. Acoust. Soc. Am. **114**, 3201–3210 (2003).
[7] K. Attenborough, S. Taherzadeh, H. E. Bass, X. Di, R. Raspet, G. R. Becker, A. Güdesen, A. Chrestman, G. A. Daigle, A. L'Espérance, Y. Gabillet, K. E. Gilbert, Y. L. Li, M. J. White, P. Naz, J. M. Noble, and H. A. J. M. van Hoot, ''Benchmark cases for outdoor sound propagation,'' J. Acoust. Soc. Am. **97**, 173–191 (1995).
[8] K. E. Gilbert and X. Di, ''A fast Green's function method for one-way sound propagation in the atmosphere,'' J. Acoust. Soc. Am. **94**, 2343–2352 (1993).
[9] K. M. Li, S. Taherzadeh, and K. Attenborough, ''Sound propagation from a dipole source near an impedance plane.'' J. Acoust. Soc. Am. **101**, 3343–3352 (1997).
[10] K. M. Li and S. Taherzadeh, ''The sound field of an arbitrarily oriented quadrupole near ground surfaces,'' J. Acoust. Soc. Am. **102**, 2050–2057 (1997).
[11] S. E. Wright and B. Vuksanovic, ''Active control of environmental noise, IV: Practical extensions to ECAS theory,'' J. Sound Vib. **222**, 635–668 (1999).
[12] T. F. W. Embleton, ''Tutorial on sound propagation outdoors,'' J. Acoust. Soc. Am. **100**, 31–48 (1996).
[13] E. M. Salomons, *Computational Atmospheric Acoustics* (Kluwer Academic, New York, 2001), Chap. 4, pp. 41–42.
[14] A. M. Nakashima, ''Active control of propeller aircraft run-up noise,'' Master of Applied Science dissertation, The University of British Columbia, 2003.

Reflection and transmission by randomly spaced elastic cylinders in a fluid slab-like region

Pierre-Yves Le Bas, Francine Luppé,[a] and Jean-Marc Conoir
LAUE, UMR CNRS 6068, Université du Havre, place R. Schuman, 76610 Le Havre, France

(Received 4 May 2004; revised 10 October 2004; accepted 16 November 2004)

An extension of Fikioris and Waterman's formalism is developed in order to describe both the reflection and transmission from a slab-like fluid region in which elastic cylindrical scatterers are randomly placed. The dispersion equation of the coherent wave inside the slab must be solved numerically. For solid cylinders, there is only one solution corresponding to a mean free path of the coherent wave larger than one wavelength. In that case, the slab region may be described as an effective dissipative fluid medium, and its reflection and transmission coefficients may be formally written as those of a fluid plate. For thin hollow shells, a second solution of the dispersion equation is found, at concentrations large enough for the shells to be coupled via the radiation of a circumferential Scholte–Stoneley *A* wave on each shell. This occurs at a few resonance frequencies of the shells. At those frequencies, then, two different coherent waves propagate in the slab, and it can no longer be considered a dissipating fluid slab. © *2005 Acoustical Society of America.*
[DOI: 10.1121/1.1848174]

PACS numbers: 43.35.Bf, 43.20.Bi, 43.20.Hq [JJM] Pages: 1088–1097

I. INTRODUCTION

Multiple scattering of waves by random distributions of scatterers is a problem that has received much attention since Foldy's early work on isotropic point scatterers.[1] Two different approaches are usually found in the litterature. In the first one, which is the subject of this paper, the study of the coherent wave leads to the description of the multiple scattering medium as an effective dissipative medium, in which incoherent scattering is considered only through the energy loss of the coherent wave. In the second approach, when the concentration of scatterers gets too high for a coherent wave to propagate, the propagation of the incoherent intensity is studied.[2,3] Many models have been derived in order to get the coherent wave properties for nonisotropic scatterers. The most well-known may be, once again, divided into two groups. In the first one,[1,4–9] it is the effective wave number (of the coherent wave) that is sought, usually in order to derive the mechanic constants (bulk and shear moduli) of the effective medium. In the models of the second group,[10–13] both the effective wave number and the amplitude of the coherent wave are sought. In this paper, the scatterers are infinitely long cylindrical elastic objects randomly placed in a slab-like region of a perfect fluid (water), and we are interested in both the coherent wave properties and the reflection and transmission coefficients of the slab. The model used is derived from Fikioris and Waterman, and the same dispersion relation as theirs is found. The reflection and transmission coefficients of the slab are then determined the same way as they did for the interface.

Throughout the paper, we shall often refer to the works of Foldy,[1] Waterman and Truell,[8] and Fikioris and Waterman.[11] They may be briefly summarized as follows.

For isotropic scatterers, namely point scatterers, and assuming double interactions between scatterers may be neglected,[14] Foldy[1] found an explicit dispersion equation for the coherent wave, relating the effective wave number to the concentration of scatterers and to the far-field scattered amplitude of each scatterer. However, scatterers of nonzero finite size are not isotropic scatterers, and this is the reason why Waterman and Truell[8] later extended Foldy's work to nonisotropic scatterers. In their work, multiple interactions between scatterers are taken into account, but it is assumed that the mean exciting field on a scatterer at a given position is the same as that when another scatterer is at a fixed position. This assumption is known as the quasicrystalline approximation.[9] In order to derive an explicit dispersion equation for the coherent wave, they did take into account the nonisotropic character of the field scattered by each object, but they let the size of each tend towards zero. The dispersion equation they obtained relates the effective wave number to the concentration of scatterers and to both the forward and backward far-field scattered amplitudes of each scatterer. Whenever the backward amplitude may be neglected, the expression for the effective wave number reduces to that of Foldy, with Foldy's far-field amplitude replaced by the forward far-field scattered amplitude. Taking into account both the nonisotropic characteristics of the scattered field of an object and its finite nonzero size leads to an implicit dispersion equation, as was shown by Fikioris and Waterman.[11]

The problem under study is presented in Sec. II, the general theory in Sec. III, and the numerical results, in both cases of aluminum cylinders and thin cylindrical shells, are discussed in Secs. IV and V.

II. PRESENTATION OF THE PROBLEM

Let us suppose N elastic scatterers, all identical, cylindrical, of radius a, and infinitely long in the z direction of a Cartesian system of coordinates (O,x,y,z). All cylinders are

[a]Electronic mail: francine.luppe@univ-lehavre.fr

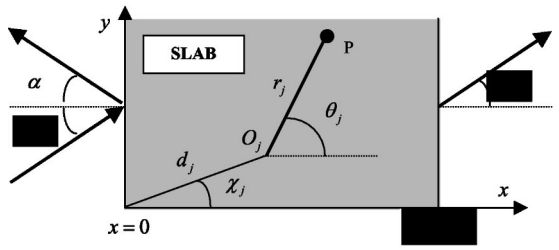


FIG. 1. Geometry of the problem. Local and global polar coordinates of O_j, center of cylinder j.

randomly distributed in a slab-like region, depth d, inside a perfect fluid that we shall suppose to be water. Figure 1 shows the different polar coordinates used. Cylinder j is centered at O_j. P is the observation point; its coordinates, in the (O,x,y) system, are $x=r\cos\theta$ and $y=r\sin\theta$.

The incident harmonic plane wave propagates in the fluid towards the slab, under angle α; it gives rise to a reflected wave and a transmitted wave outside the slab, as well as to a multiple scattering process inside the slab.

Let $p(\mathbf{d}_\ell)$ be the probability density of finding cylinder ℓ at $\mathbf{d}_\ell=(x_\ell=d_\ell\cos\chi_\ell, y_\ell=d_\ell\sin\chi_\ell)$, and $p(\mathbf{d}_\ell|\mathbf{d}_j)$ the conditional probability density of finding cylinder ℓ at $\mathbf{d}_\ell$ provided cylinder j is at $\mathbf{d}_j$.

The cylinders are supposed to be uniformly distributed in the slab, with concentration $c=(N/S)\pi a^2\simeq[(N-1)/S]\pi a^2$. S is the surface of the slab, and

$$p(\mathbf{d}_\ell)=\begin{cases}\dfrac{1}{S} & \text{if } \mathbf{d}_\ell\in S\\ 0 & \text{if } \mathbf{d}_\ell\notin S\end{cases}. \tag{1}$$

Exclusion of interpenetration of the cylinders of radius a is ensured by the following conditional probability density:

$$p(\mathbf{d}_\ell|\mathbf{d}_j)=\begin{cases}\dfrac{1}{S}(1-f(d_{\ell j})) & \text{if } \mathbf{d}_\ell\in S\\ 0 & \text{if } \mathbf{d}_\ell\notin S\end{cases}, \quad d_{\ell j}=\|\mathbf{d}_\ell-\mathbf{d}_j\|. \tag{2}$$

As in Refs. 6 and 11, the simplest ''pair correlation function'',[5] corresponding to the ''hole correction,'' is chosen

$$f(d_{\ell j})=\begin{cases}1 & \text{if } d_{\ell j}\leqslant\beta a\\ 0 & \text{if } d_{\ell j}>\beta a\end{cases}, \quad \beta\geqslant 2. \tag{3}$$

In the local coordinates system of cylinder j, the incident harmonic pressure field on the slab is

$$\varphi_{inc}^{(j)}=e^{i\mathbf{k}\cdot\mathbf{d}_j}e^{i\mathbf{k}\cdot\mathbf{r}_j}=e^{i\mathbf{k}\cdot\mathbf{d}_j}\sum_n i^n J_n(kr_j)e^{in\theta_j}e^{-in\alpha}, \tag{4}$$

with an $e^{-i\omega t}$ time dependence that will be omitted throughout the paper, Σ_n, standing for $\Sigma_{n=-\infty}^{+\infty}$, and $\mathbf{k}=(k_x=k\cos\alpha,\ k_y=k\sin\alpha)$.

III. THEORY

Varadan *et al.*[6] have derived the implicit dispersion equation of the coherent wave propagating in a fluid half-space with a random distribution of elastic cylinders inside. Their work is very similar to that of Fikioris and Waterman, with two main differences. First, they introduce the transition matrix of each scatterer, according to the T-matrix approach,[6] which allows them to deal with cylinders of arbitrary cross sections. Second, they are not interested in the reflection and transmission coefficients of the half-space. In this paper, the transition matrix is used, and the notations follow those of Ref. 6, but the reflection and transmission coefficients of the fluid slab will also be sought.

In a first step, then, the transition matrix formulation is used to express the different acoustic fields in the region containing the scatterers. These fields are the total pressure field exciting a given scatterer and the total field scattered by the same scatterer. They are related via the transition matrix. Each one of them is expanded as a modal series of cylindrical functions. In a second step, a configurational average[1] of the modal coefficients of the exciting field is performed, under the assumption of a uniform distribution of identical scatterers, and using Fikioris and Waterman's ''hole correction''[11] as well as Lax's quasicrystalline approximation.[9] The averaged modal coefficients of the total exciting field on a given scatterer are then each decomposed into plane waves, with the amplitudes of the plane waves depending on the mode number, and with a complex wave number independent of the mode number. Hence, the equations describing the multiple scattering process lead to two different kinds of systems, with the amplitudes of the plane waves as unknowns, and the complex effective wave number as a parameter. The first sets of equations are homogeneous; equating to zero their common determinant gives the effective wave number. This process is often referred to as the ''Lorentz–Lorenz law.'' So far, our model is exactly the same as in Ref. 6. The second set of equations can then be solved and gives the unknown amplitudes. This process is called the ''extinction Ewald–Oseen's theorem;'' this model differs from Varadan's in the way the extinction theorem is used.

A. Multiple scattering equations and configurational averaging

Generally speaking, the total exciting field on cylinder j can be expressed as a modal series

$$\varphi^{(j)}=\sum_n e_n^{(j)}J_n(kr_j)e^{in\theta_j}, \tag{5}$$

and the total scattered field by cylinder j as

$$\phi^{(j)}=\sum_n C_n^{(j)}H_n^{(1)}(kr_j)e^{in\theta_j}. \tag{6}$$

First, the diagonal transition T matrix of each cylinder, assumed to be known (cf. Appendix B), is used to link these two fields

$$C_n^{(j)}=T_{nn}e_n^{(j)}. \tag{7}$$

Next, the total exciting field $\varphi^{(j)}$ is composed of the incident plane wave and of the waves scattered by all cylinders $\ell\neq j$

$$\varphi^{(j)}=\varphi_{\text{inc}}^{(j)}+\sum_{\ell=1,\ell\neq j}^{N}\phi^{(\ell)}. \tag{8}$$

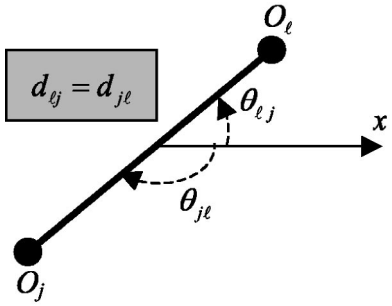


FIG. 2. Notations used in the Graf addition theorem.

In Eq. (8), $\varphi_{\text{inc}}^{(j)}$ is defined by Eq. (4), and $\phi^{(\ell)}$ by Eq. (6). The Graf addition theorem[15] is now introduced in order to express the waves scattered by cylinder ℓ, $\phi^{(\ell)}$, as incident waves on the surface of cylinder j

$$H_n^{(1)}(kr_\ell)e^{in\theta_\ell} = \sum_m G_{nm}^{(j\ell)} J_m(kr_j)e^{im\theta_j}, \quad r_j < d_{\ell j}, \quad (9)$$

with (see Fig. 2) $G^{(j\ell)}$ the Graf operator defined by

$$G_{mn}^{(j\ell)} = e^{i(m-n)\theta_{j\ell}} H_{m-n}^{(1)}(kd_{\ell j}). \quad (10)$$

It follows that

$$\varphi^{(j)} = e^{i\mathbf{k}\cdot\mathbf{d}_j}\sum_n i^n J_n(kr_j)e^{in\theta_j}e^{-in\alpha} + \sum_{\ell=1,\ell\neq j}^{N} \sum_m \sum_n G_{nm}^{(j\ell)} T_{nn}^{(\ell)} e_n^{(\ell)} J_m(kr_j)e^{im\theta_j}. \quad (11)$$

Comparison of relations (5) and (11) gives the linear system of equations

$$e_n^{(j)} = i^n e^{i\mathbf{k}\cdot\mathbf{d}_j} e^{-in\alpha} + \sum_{\ell=1,\ell\neq j}^{N} \sum_m G_{mn}^{(j\ell)} T_{mm} e_m^{(\ell)}. \quad (12)$$

Equation (12) is exact: no assumption has been made, except that the cylinders do not either touch or interpenetrate each other (otherwise the Graf theorem would not be valid).

For a given position of cylinder j, relation (12) is averaged over all possible positions of cylinders ℓ

$$\langle e_n^{(j)}\rangle_j = e^{i\mathbf{k}\cdot\mathbf{d}_j} i^n e^{-in\alpha} + \frac{1}{S}\sum_{\ell=1,\ell\neq j}^{N} \sum_m T_{mm}\int_{S'} d\mathbf{d}_\ell G_{mn}^{(j\ell)} \times \langle e_m^{(\ell)}\rangle_{j\ell}. \quad (13)$$

Relation (13) is equivalent to relation (23) of Ref. 6. S' is equal to the slab surface S, with exclusion of the surface of a cylinder of radius βa centered at $\mathbf{d}_j$. $\langle e_m^{(\ell)}\rangle_{j\ell}$ stands for the average of quantity $e_m^{(\ell)}$, over all possible positions of all cylinders, except cylinders j and ℓ, which are kept fixed.

In order to break down the averaging process,[9] the quasicrystalline approximation

$$\langle e_m^{(\ell)}\rangle_{j\ell} = \langle e_m^{(\ell)}\rangle_\ell, \quad \ell \neq j, \quad (14)$$

is used, as well as the fact that all cylinders are identical, so that relation (13) gives

$$\langle e_n^{(j)}\rangle_j = e^{i\mathbf{k}\cdot\mathbf{d}_j} i^n e^{-in\alpha} + \frac{c}{\pi a^2}\sum_m T_{mm}\int_{S'} d\mathbf{d}_\ell G_{mn}^{(j\ell)} \times \langle e_m^{(\ell)}\rangle_\ell. \quad (15)$$

The averaged modal coefficient $\langle e_n^{(j)}\rangle_j$ is decomposed into two refracted plane waves, one propagates, under angle α_r, in the direction of increasing x, and the other one propagates, under the same angle, in the direction of decreasing x

$$\langle e_n^{(j)}\rangle_j = i^n e^{-in\alpha_r} X_n^+ e^{i\mathbf{K}^+\cdot\mathbf{d}_j} + i^{-n} e^{+in\alpha_r} X_n^- e^{i\mathbf{K}^-\cdot\mathbf{d}_j}, \quad (16)$$

where $\mathbf{K}^+$ and $\mathbf{K}^-$ are effective wave vectors

$$\mathbf{K}^+ = (K_x = K\cos\alpha_r, K_y = K\sin\alpha_r),$$
$$\mathbf{K}^- = (-K_x, K_y), \quad (17)$$

which verify the Snell–Descartes law

$$K_y = k_y. \quad (18)$$

Both $\mathbf{K}^+$ and $\mathbf{K}^-$, as well as the $X_n^\pm$ coefficients, are unknowns in system (15). Due to the invariance of the slab in the y direction, y_j will be supposed equal to zero from now on, with no lack of generality.

B. Dispersion equation of the coherent waves

Relations (16) and (17) are now introduced into system (15), so that

$$X_n^+ e^{iK_x x_j} + (-1)^n e^{+2in\alpha_r} X_n^- e^{-iK_x x_j} = e^{ik_x x_j} e^{+in(\alpha_r-\alpha)} + \frac{c}{\pi a^2}\left[\sum_p T_{(p+n)(p+n)} i^p e^{-ip\alpha_r} X_{p+n}^+ I_p^+ + \sum_p T_{(p+n)(p+n)} i^{-p}(-1)^n e^{ip\alpha_r} e^{2in\beta} X_{p+n}^- I_p^-\right], \quad (19)$$

with

$$I_p^\pm = \int_{S'} e^{ip\theta_{j\ell}} H_p^{(1)}(kd_{\ell j}) e^{i\mathbf{K}^\pm\cdot\mathbf{d}_\ell} d\mathbf{d}_\ell. \quad (20)$$

As in Ref. 6, the Green's theorem is applied to the surface integrals in Eq. (20)

$$I_p^\pm = \frac{1}{k^2-K^2}\int_{L'} e^{ip\theta_{j\ell}}[H_p^{(1)}(kd_{\ell j})\hat{n}\cdot\boldsymbol{\nabla}(e^{i\mathbf{k}^\pm\cdot\mathbf{d}_\ell}) - e^{i\mathbf{K}^\pm\cdot\mathbf{d}_\ell}\hat{n}\cdot\boldsymbol{\nabla}(H_p^{(1)}(kd_{\ell j}))]dl, \quad (21)$$

with L' a closed curve, composed of the circumference $L_{\beta a}$ of a circle of radius βa centered at O_j, the two straight lines that define the slab region ($x_\ell = 0$ and $x_\ell = d$), and two arcs that connect the previous two straight lines at infinity. $\hat{n}$ is the local unit vector normal to L'. The two integrals defined in Eq. (21), equal to (see Appendix A)

$$I_p^+ = \frac{2i^{-p}}{(k^2-K^2)}\left\{ -\frac{i}{k_x}(K_x+k_x)e^{ik_xx_j}e^{ip\alpha} + (-1)^p e^{+i(K_x+k_x)d}\frac{i}{k_x}(K_x-k_x)e^{-ip\alpha}e^{-ik_xx_j} \right.$$
$$\left. -\pi b e^{+iK_xx_j}e^{+ip\alpha_r}[KJ_p'(K\beta a)H_p^{(1)}(k\beta a) - kJ_p(K\beta a)H_p^{(1)'}(k\beta a)] \right\},$$
$$I_p^- = \frac{2i^{-p}}{(k^2-K^2)}\left\{ \frac{i}{k_x}(K_x-k_x)e^{ik_xx_j}e^{ip\alpha} + (-1)^{p+1}\frac{ie^{+i(k_x-K_x)d}}{k_x}(K_x+k_x)e^{-ip\alpha}e^{-ik_xx_j} \right. \tag{22}$$
$$\left. +(-1)^{p+1}\pi b e^{-iK_xx_j}e^{-ip\alpha_r}[KJ_p'(K\beta a)H_p^{(1)}(k\beta a) - kJ_p(K\beta a)H_p^{(1)'}(k\beta a)] \right\}$$

are introduced into system (19).

This system must be verified, whatever the position of cylinder j, i.e., for any x_j. Equating separately the factors multiplying $e^{iK_xx_j}$ and $e^{-iK_xx_j}$ yields the following linear homogeneous sytems known as Lorentz–Lorenz laws:

$$X_n^{\pm} = -\frac{2\beta c}{(k^2-K^2)a^2}\sum_p T_{(p+n)(p+n)}X_{p+n}^{\pm}$$
$$\times[KaJ_p'(K\beta a)H_p^{(1)}(k\beta a)$$
$$-kaJ_p(K\beta a)H_p^{(1)'}(k\beta a)]. \tag{23}$$

Both systems have a determinant, depending on the effective wave number K, that must be equal to zero

$$\mathrm{Det}\left(I + \frac{2\beta c}{(k^2-K^2)a^2}AT \right) = 0, \tag{24}$$

with I the identity matrix, T the transition matrix, and A the square matrix defined by its elements

$$A_{nm} = [KaJ_{n-m}'(K\beta a)H_{n-m}^{(1)}(k\beta a)$$
$$-kaJ_{n-m}(K\beta a)H_{n-m}^{(1)'}(k\beta a)]. \tag{25}$$

Relation (24) is the dispersion equation of the effective medium in the slab. As the transition matrix depends on frequency only via the reduced frequency ka, so does the effective wave number K.

Contrary to the dispersion equation found by Waterman and Truell, Eq. (24) is not an explicit relation on the effective wave number; it has to be numerically solved. It is exactly the same as that found by Fikioris and Waterman, as well as by Varadan *et al.*

For cylindrical scatterers such as ours, the dispersion equation of Waterman and Truell[8] gives

$$\left(\frac{K}{k}\right)^2 = \left(1 - \frac{2ic}{\pi(ka)^2}\sum_n T_{nn} \right)^2$$
$$-\left(\frac{2ic}{\pi(ka)^2}\sum_n (-1)^p T_{nn} \right)^2. \tag{26}$$

The first infinite summation in Eq. (26) corresponds to forward scattering by one cylinder, and the second one to backward scattering. If backward scattering is neglected, Eq. (26) reduces to

$$\frac{K}{k} = 1 - \frac{2ic}{\pi(ka)^2}\sum_n T_{nn}. \tag{27}$$

Equation (27) is identical to Foldy's dispersion equation for point scatterers that would radiate a far-field amplitude equal to the forward far-field amplitude radiated by a cylinder.

It will be numerically observed, in Secs. IV and V, that Eq. (24) has (at least) one solution, which is very close to that given by Eq. (26). As the resonance frequencies of one scatterer are linked to the poles of its transition matrix T (cf. Appendix B), they correspond, on the plot of the forward- (or backward) scattered amplitude versus frequency, to sharp variations. This will be, of course, also observed on the plot, versus frequency, of both the real and imaginary parts of the effective wave number.

C. Amplitudes of the coherent waves

Let us suppose that the equation of dispersion, Eq. (24), has been solved, providing only one physically acceptable solution K. The $X_n^{\pm}$ are still unknowns that will be determined in the following. Systems (23) provide

$$\forall n \neq 0, \quad \sum_{m\neq 0}\left(\delta_{nm} - \frac{2\beta c}{\kappa^-\kappa^+a^2}A_{nm}T_{mm} \right)X_m^{\pm}$$
$$= +\frac{2\beta c}{\kappa^-\kappa^+a^2}A_{n0}T_{00}X_0^{\pm}, \tag{28}$$

with

$$\kappa^{\pm} = K_x \pm k_x. \tag{29}$$

Those two systems may be solved using Cramer's method. Let D be the common determinant of the associated homogeneous systems and D_m the same one, except that the elements of column m, row n, are replaced by A_{n0}, so that

$$\forall n \neq 0, \quad X_n^{\pm} = \frac{2\beta c}{\kappa^-\kappa^+a^2}\frac{D_n}{D}T_{00}X_0^{\pm} = \eta D_n T_{00}X_0^{\pm}, \tag{30}$$

with $D_n = D_{-n}$. The last step now consists of calculating $X_0^{\pm}$.

So far, we have equated separately the factors that multiply $e^{iK_x x_j}$ and $e^{-iK_x x_j}$ in system (19). Equating also those multiplying $e^{ik_x x_j}$, on one hand, and those multiplying $e^{-ik_x x_j}$, on the other hand (extinction theorem), leads to two equations

$$\frac{2ic}{\pi k_x a}\left[\frac{1}{\kappa^- a}\sum_m T_{mm}e^{-im(\alpha_r-\alpha)}X_m^+ - \frac{1}{\kappa^+ a}\times\sum_m T_{mm}e^{im(\alpha+\alpha_r)}(-1)^m X_m^-\right]=-1, \tag{31}$$

$$\kappa^- e^{+iK_x d}\sum_m T_{mm}(-1)^m e^{-im(\alpha+\alpha_r)}X_m^+ - \kappa^+ e^{-iK_x d}\times\sum_m T_{mm}e^{+im(\alpha_r-\alpha)}X_m^-=0, \tag{32}$$

which, combined with Eq. (30), gives

$$X_0^+=\frac{i\pi k_x a}{2cT_{00}}\frac{\kappa^- a}{(1+\eta f)}\frac{1}{1-Q^2e^{2iK_x d}},$$

$$X_0^-=Qe^{2iK_x d}X_0^+, \tag{33}$$

with

$$Q=\frac{1+\eta g}{1+\eta f}\frac{\kappa^-}{\kappa^+},\quad f=\sum_{n=-N_0,n\neq 0}^{N_0} D_n T_{nn}e^{in(\alpha_r-\alpha)},$$

$$g=\sum_{n=-N_0,n\neq 0}^{N_0}(-1)^n D_n T_{nn}e^{-in(\alpha_r+\alpha)}. \tag{34}$$

D. The reflection and transmission coefficients of the slab

The scattered field, outside the slab, is given by

$$\phi_{\text{diff}}=N\int_S p(\mathbf{d}_j)\sum_n T_{nn}\langle e_n^{(j)}\rangle_j H_n^{(1)}(kr_j)e^{in\theta_j}d\mathbf{d}_j. \tag{35}$$

Use of Eq. (16) gives

$$\phi_{\text{diff}}=\frac{c}{\pi a^2}\sum_n T_{nn}i^n e^{-in\alpha_r}X_n^+ L_j^+ + \frac{c}{\pi a^2}\sum_n T_{nn}i^{-n}e^{+in\alpha_r}X_n^- L_j^-, \tag{36}$$

with

$$L_n^{\pm}=\int_S H_n^{(1)}(kr_j)e^{in\theta_j}e^{i\mathbf{K}^{\pm}\cdot\mathbf{d}_j}d\mathbf{d}_j. \tag{37}$$

In the same way as before, the Green's theorem is applied to the surface integral in Eq. (37), so that

$$L_n^{\pm}=\frac{1}{k^2-K^2}\left[\left(\mp iK_x+\frac{\partial}{\partial x_j}\right)+e^{\pm iK_x d}\left(\pm iK_x-\frac{\partial}{\partial x_j}\right)\right]\times\int_{-\infty}^{+\infty}e^{iK_y y_j}[H_n^{(1)}(kr_j)e^{in\theta_j}]dy_j. \tag{38}$$

Once again, the last integral in Eq. (38) is evaluated, as in Appendix A, by use of the Sommerfeld integral representation of the Hankel function. The result depends on the position of the observation point P (via r_j and θ_j), i.e., on whether the reflected or the transmitted field is sought.

1. The reflection coefficient R

The observation point P corresponds to $x<0$, and the calculation of the integral in Eq. (38) gives

$$\int_{-\infty}^{+\infty}e^{iK_y y_j}[H_n^{(1)}(kr_j)e^{in\theta_j}]dy_j=\frac{2i^n}{k_x}e^{ik_x(x_j-x)}e^{ik_y y}e^{-in\alpha}. \tag{39}$$

Equation (36) then reads

$$\phi_{\text{diff}}(x<0,y)=Re^{-ik_x x}e^{+ik_y y}, \tag{40}$$

with

$$R=\frac{2ic}{\pi k_x a}\left[\frac{1-e^{+i\kappa^+ d}}{\kappa^+ a}\sum_n T_{nn}(-1)^n e^{-in(\alpha+\alpha_r)}X_n^+ + \frac{-1+e^{-i\kappa^- d}}{\kappa^- a}\sum_n T_{nn}e^{+in(-\alpha+\alpha_r)}X_n^-\right]. \tag{41}$$

Evaluation of the infinite series, with use of Eqs. (30), (33), and (34), finally provides

$$R=\frac{-Q(1-e^{2iK_x d})}{1-Q^2e^{2iK_x d}}. \tag{42}$$

2. The transmission coefficient T

The observation point P corresponds to $x>d$, and the calculation of the integral in Eq. (38) gives

$$\int_{-\infty}^{+\infty}e^{iK_y y_j}[H_n^{(1)}(kr_j)e^{in\theta_j}]dy_j=\frac{2i^{-n}}{k_x}e^{-ik_x x_j}e^{ik_x x}e^{ik_y y}e^{in\alpha}. \tag{43}$$

Equation (36) then reads

$$\phi_{\text{diff}}(x>d,y)=\frac{2ic}{\pi k_x a}\left[\frac{1-e^{+i\kappa^- d}}{\kappa^- a}\sum_n T_{nn}e^{-in(-\alpha+\alpha_r)}X_n^+ + \frac{-1+e^{-i\kappa^+ d}}{\kappa^+ a}\sum_n T_{nn}(-1)^n \times e^{+in(\alpha+\alpha_r)}X_n^-\right]e^{ik_x x}e^{ik_y y}. \tag{44}$$

The transmission coefficient T, however, is related to the total field ϕ_T at $x>d$, as

$$\phi_T(x>d,y)=\phi_{\text{diff}}(x>d,y)+\varphi_{\text{inc}}(x>d,y)=Te^{ik_x x}e^{ik_y y}. \tag{45}$$

Evaluation of the infinite series in Eq. (44), with use of Eqs. (30), (33), (34), finally provides

$$T=\frac{(1-Q^2)e^{iK_x d}}{1-Q^2e^{2iK_x d}}e^{-ik_x d}. \tag{46}$$

Now, let us define τ as

$$\tau = \frac{1-Q}{1+Q} = \frac{Z_2}{Z_1}\frac{\cos\alpha}{\cos\alpha_r}, \tag{47}$$

so that the expressions found for R and T are those of the reflection and transmission coefficient of a fluid slab (medium 2),[16] of acoustic impedance Z_2, embedded in another fluid (medium 1), of acoustic impedance $Z_1=\rho_1 c_1$.

Equation (47) defines the acoustic impedance Z_2 of the effective medium in the slab. As the effective wave number K depends on frequency, so does the refraction angle α_r, and the acoustic impedance Z_2. This, in turn, defines the frequency-dependent density of the effective medium, from

$$Z_2 = \rho_{\text{eff}} c_{\text{eff}} = \rho_{\text{eff}} \frac{\omega}{\Re e(K)}, \tag{48}$$

with $\Re e$ standing for "real part of."

The next section is devoted to the numerical results obtained when the scatterers are solid steel cylinders in water.

IV. THE SCATTERERS ARE STEEL CYLINDERS: NUMERICAL RESULTS

Steel is characterized by its density $\rho=7816$ kg/m^3, the velocity of the longitudinal waves $c_L=6000$ m/s, that of the shear waves $c_s=3100$ m/s, water by its density $\rho_1=1000$ kg/m^3, and by the velocity of sound $c_1=1480$ m/s. In all cases considered, parameter β is fixed to 2. The hole correction, Eq. (3), is known[17,18] to be less realistic than Percus–Yevick's approximation, especially as the concentration increases. Like Tsang *et al.*,[17] we found indeed that increasing concentration c could lead to obtain solutions of the dispersion equation that would correspond to an amplified coherent wave. This occurred, in the frequency range we considered, for concentration values greater than 0.25. This is the reason why concentration c is fixed to 0.1 in the following.

A. The coherent wave properties

A numerical resolution of the dispersion equation has provided one (and only one) solution $K=K'+iK''$, with K' and K'' positive real numbers such that $K''/K'<1/(2\pi)$, i.e., the wavelength of the coherent wave is greater than its mean free path[3] $L_s=1/K''$. The effective attenuation in the slab is then K'', and the effective velocity is given by

$$c_{\text{eff}} = \frac{\omega}{K'}. \tag{49}$$

The solutions of the dispersion equations Eq. (26) and Eq. (27) are practically identical in the frequency range we consider. Both Fig. 3 and Fig. 4 show that the difference between the solution of Waterman and Truell's Eq. (26), plotted in thin lines, and that of Eq. (24), plotted in thick lines, is also rather negligible.

Figure 3 presents the evolution of the reduced effective velocity c_{eff}/c_0 with the reduced frequency ka, obtained from the resolution of Eq. (24) and from that of Waterman and Truell's dispersion equation. The presence of scatterers is seen to reduce slightly the propagation speed. Not surprisingly, the largest differences between c_{eff} and c_0 occur at particular frequencies which are resonance frequencies of the cylinders. However, these differences never exceed 5%.

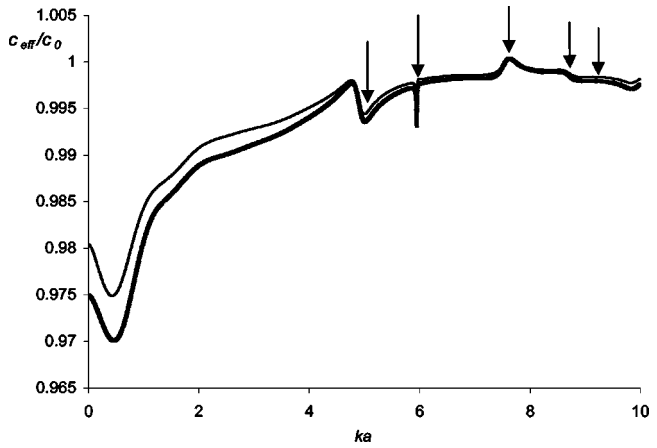


FIG. 3. The scatterers are steel cylinders. Concentration $c=0.1$. Evolution, with reduced frequency ka, of the ratio of the effective velocity of the coherent wave to that of sound in water. The arrows point at the resonance frequencies of a steel cylinder. Thin line: Waterman and Truell's solution. Thick line: solution of Eq. (24).

The evolution with frequency of the reduced effective attenuation, or loss tangent, K''/K', is shown in Fig. 4. At low frequency, the K''/K' ratio is roughly an increasing linear function of frequency, with a very slightly greater slope in Waterman and Truell's case than in ours. The other two curves in Fig. 4 show the ratio of the mean elastic free path to the radius of the cylinders. At low frequency, as the loss tangent increases linearly, the elastic mean free path diminishes from infinity (at zero frequency) to around $33a$, at $ka=1$. Then, as the frequency increases, the mean free path goes on decreasing, except at the first two resonance frequencies of the cylinders. Usually, a damped wave is considered to be propagating on distances no larger than three times its mean free path. The mean value of the elastic mean free path, here, is around $15a=7.5\beta a$, which means that the coherent wave is about to collapse after having encountered, roughly, no more than 24 scatterers.

In the following sections, we shall study the reflection and transmission coefficients of the slab. Considering our

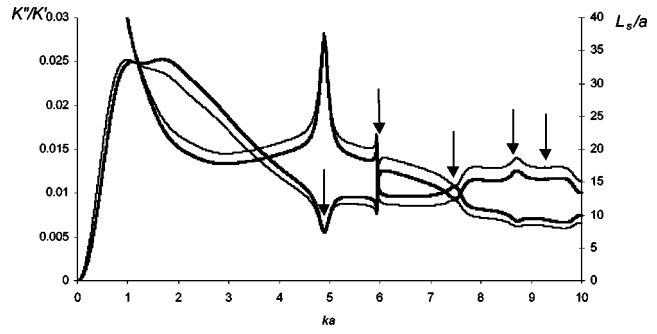


FIG. 4. The scatterers are steel cylinders. Concentration $c=0.1$. Evolution, with reduced frequency ka, of the effective loss tangent of the coherent wave (curves begining at zero frequency), and of the elastic mean free path normalized by the radius a of the cylindrical scatterers. The arrows point at the resonance frequencies of a steel cylinder. Thin line: Waterman and Truell's solution. Thick line: solution of Eq. (24).

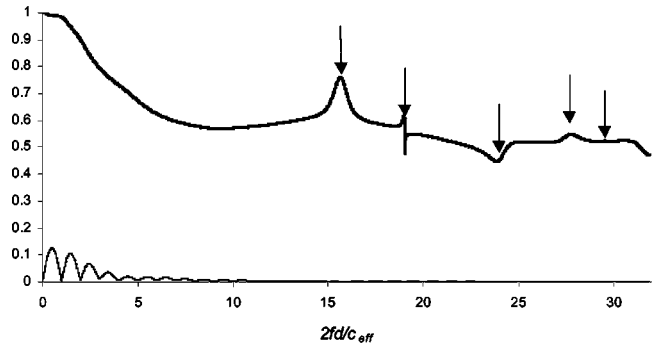


FIG. 5. The scatterers are steel cylinders. Concentration $c=0.1$. Normal incidence on the slab. Moduli of the reflection coefficient (thin line) and of the transmission coefficient (thick line) versus reduced frequency. The slab thickness d equals $10a$. The arrows point at the resonance frequencies of a steel cylinder.

last remark on the mean free path, the slab thickness d we shall consider will be fixed to $10a$.

B. Reflection and transmission by the slab at normal incidence ($\alpha=\alpha_r=0$)

The moduli of the reflection coefficient R and of the transmission coefficient T are plotted in Fig. 5, versus $2fd/c_{\text{eff}}$, with f the frequency. There is little reflection by the slab compared to the transmission. Energy is seen to be lost due to incoherent scattering. The curve of the modulus of the reflection coefficient presents damped periodic oscillations. The reflection is minimum at frequencies such that the coherent transmitted wave in the slab is a standing wave. As the reflection coefficient is roughly zero at $2fd/c_{\text{eff}}\geqslant 12$, i.e. $ka\geqslant 3.73$, while the first resonance of a steel cylinder occurs at $ka\simeq 5$, no resonance of the cylinders may be observed on that curve. This is not the case of the transmission coefficient. Its modulus undergoes visible variations at each resonance frequency of the cylinders.

The reflection and transmission coefficients, Eqs. (42) and (46), are those of a fluid slab of acoustic impedance Z_2, defined in Eq. (47). These coefficients may be written in an alternative way, using the Debye series,[16] which allows the description of the fluid slab as a Fabry–Perot interferometer. These series are expressed via the local reflection and transmission coefficients R_{pq} and T_{pq} $(p,q=1,2)$ at each local boundary between fluids p and q

$$R_{pq}=\frac{Z_q-Z_p}{Z_q+Z_p},\quad T_{pq}=\frac{2Z_p}{Z_q+Z_p},\quad p,q=1,2, \tag{50}$$

$$R=R_{12}+T_{12}T_{21}R_{21}e^{2iKd}\sum_{n=0}^{+\infty}R_{21}^{n}e^{2niKd},$$

$$T=T_{12}T_{21}e^{iKd}\sum_{n=0}^{+\infty}R_{21}^{2n}e^{2niKd}. \tag{51}$$

In the case of Fig. 5, the exact coefficients, determined after Eqs. (42) and (46), may be seen to be identical to those calculated from Eq. (51) while keeping only, in the infinite series, the $n=0$ term. Moreover, the product $T_{12}T_{21}$ is merely equal to 1 on the whole frequency range investigated. This means that no slab effect appears on the transmission coefficient: the coherent wave propagates through only once. This is due to the relatively low value of the ratio of the mean free path to the slab thickness. The slab effect, in contrast, appears on the reflection coefficient, as $R=R_{12}$ is not a valuable approximation of it. From a practical point of view, this means that the effective wave number K of the coherent wave may be determined experimentally by use of Sachse and Pao's method[19] on the transmitted field, as done, for example, in Ref. 20, in which no local transmission coefficients, such as T_{12}, T_{21}, are taken into account.

In the next section, we are interested in the effect a resonant coupling between scatterers may have on the coherent wave. By resonant coupling, we mean that a circumferential wave that is responsible for a resonance of the scatterers has a skin depth, in the surrounding liquid, large enough to include some part of the closest neighboring cylinder. This gives rise to a new resonance, which is part of the signature of the scatterer composed of those cylinders that are included in the skin depth of the wave.

Solid cylinders, even close to each other, are difficult to couple, as all the circumferential waves that are responsible for the resonances are internal waves, i.e., most of their energy is inside the solid. Thin empty elastic shells, on the contrary, may be coupled at low frequency,[21] as the circumferential wave that is responsible for their resonances carries more energy in the fluid part of the shell outer interface than in its solid part.

V. THE SCATTERERS ARE THIN EMPTY ALUMINUM SHELLS: NUMERICAL RESULTS

Aluminum is characterized by its density $\rho=2790$ kg/m^3, the velocity of the longitudinal waves $c_L=6120$ m/s, and that of the shear waves $c_s=3020$ m/s. The shell inner radius is $b=0.9a$, with a the outer radius. Parameter β is still fixed to 2, but the concentration c is now 0.25. The mean distance between the scatterers, then, is reduced, compared to the previous case. This increases the possibility for the shells to couple via the circumferential A wave (Scholte–Stoneley wave).[21] The frequency range investigated is the same as in the previous section; in this range, all resonances of the shells are due to the A wave. The dispersion equation, Eq. (24), provides two solutions, and the corresponding loss tangents are plotted, versus ka, in Fig. 6, along with that corresponding to Waterman and Truell's solution. The sharp variations observed on each curve occur at a resonance frequency of the shells. The plot range of the K''/K' ratio extends from zero (no damping) to 0.2: when $K''/K'=1/(2\pi)\simeq 0.16$, the mean free path equals the wavelength, and the wave is (*quasi*) nonpropagative. Both the first solution of Eq. (24) and Waterman and Truell's solution are seen to be propagative in the whole frequency range, except at two resonance frequencies (ka around 0.3 and 1.2), while the second solution of Eq. (24) is nonpropagative for ka values less than 3.5, except at those same particular resonance frequencies.

Figure 7 shows the velocity dispersion curves of the same three solutions. Both the first solution of Eq. (24) and Waterman and Truell's solution correspond, this time, to an effective velocity greater than that of the incident wave.

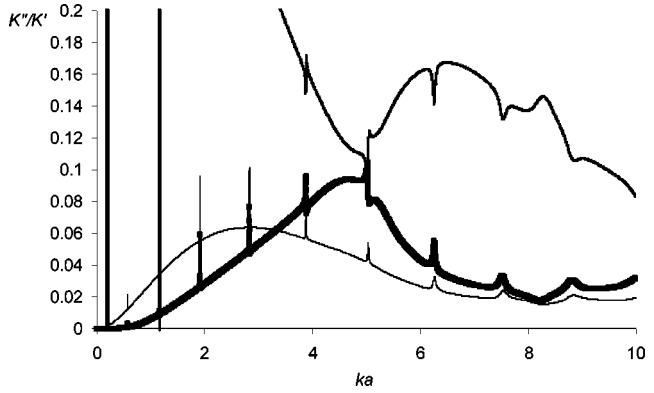


FIG. 6. The scatterers are thin empty aluminum shells, $b/a=0.9$, concentration $c=0.25$. Evolution, with reduced frequency ka, of the effective loss tangent of the coherent wave. Thin line: Waterman and Truell's solution. Thick line: second solution of Eq. (24). Very thick line: first solution of Eq. (24).

When propagative ($ka \geqslant 3.5$), the second solution of Eq. (24) propagates with a lower velocity than that of sound in water.

These results are similar to those found by Jing *et al.*[22,23] on colloidal suspensions of spheres, in which they observed two quasimodes of finite life time. The one they call ''the high-frequency mode'' corresponds to our first solution of Eq. (24), which has the closest dispersion curve to that of sound in the homogeneous fluid. They observe gaps in its dispersion curve, corresponding to a single sphere resonance. In our case, indeed, the mean free path of the first solution of Eq. (24) is seen, in Fig. 8, to decrease at each resonance frequency of a single shell. The low-frequency mode of Jing *et al.* corresponds to our second solution of Eq. (24). Its velocity is lower than that of sound in the homogeneous fluid, and its mean free path, while quite low in our case, increases at each resonance frequency of a shell. This mode is due to the coupling of neighboring shells, via the Scholte–Stoneley A wave, just as in the case considered by Jing *et al.*

The mean free path of both solutions is lower than ten

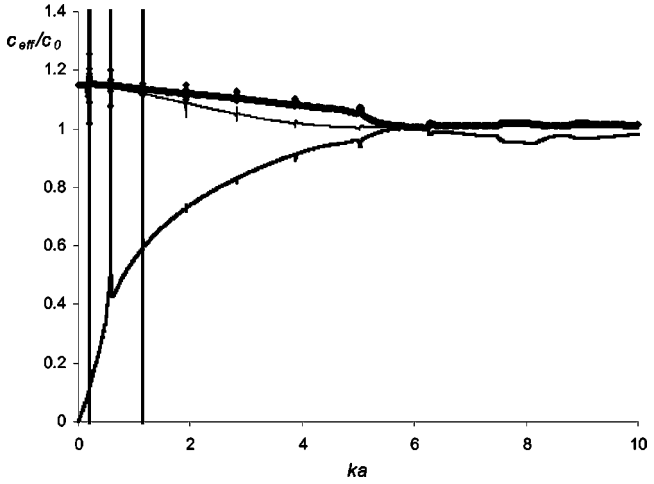


FIG. 7. The scatterers are thin empty aluminum shells, $b/a=0.9$, concentration $c=0.25$. Evolution, with reduced frequency ka, of the ratio of the effective velocity of the coherent wave to that of sound in water. Thin line: Waterman and Truell's solution. Thick line: second solution of Eq. (24). Very thick line: first solution of Eq. (24).

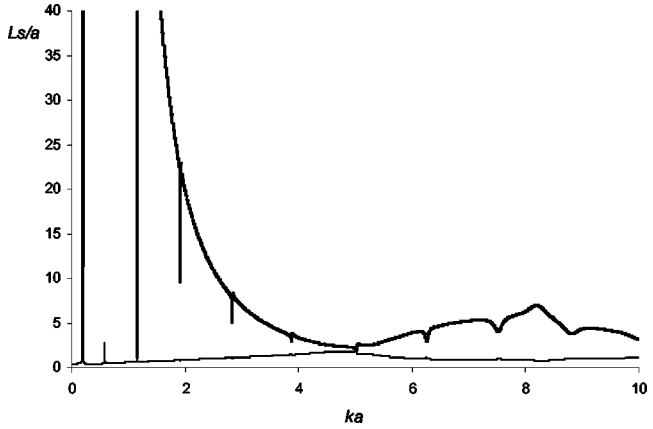


FIG. 8. The scatterers are thin empty aluminum shells, $b/a=0.9$, concentration $c=0.25$. Evolution, with reduced frequency ka, of the elastic mean free path normalized by the radius a of the shells. Thin line: second solution of Eq. (24). Thick line: first solution of Eq. (24).

times the radius of the shells, for reduced frequencies ka greater than 2.5. At such frequencies, then, the study of the reflection/transmission by the slab is no longer interesting. At lower frequencies, however, and particularly at resonance frequencies of the shells, the analysis of the reflection and transmission coefficients of the slab should take into account the presence of two coherent waves, which is not done in the present study. This work is currently in progress.

VI. CONCLUSION

We have presented an extension of Fikioris and Waterman's formalism to describe the reflection/transmission process from a slab-like region in which multiple scattering occurs. The dispersion equation of the coherent wave inside the slab must be solved numerically, unlike that found by Waterman and Truell. The hole correction has been used, so that there is no possibility for interpenetration of scatterers. The dispersion equation may have, *a priori*, more than one solution. If only one corresponds to a mean free path of the coherent wave larger than one wavelength, the slab region may be described, from the coherent wave point of view, as an effective dissipative fluid medium, and its reflection and transmission coefficients have been formally written as those of a fluid plate. For solid elastic cylinders such as scatterers, both coefficients have been shown to be numerically equivalent to the first term(s) of their Debye series, as the damping process of the coherent wave kills the multiple reflections in the slab. For thin hollow shells, a second solution of the dispersion equation may be found, at concentrations large enough for the shells to be coupled via the radiation of a circumferential Scholte–Stoneley A wave on each shell. This process gives rise to a propagating coherent wave, with a velocity slightly lower than that of the homogeneous incident wave, and a mean free path rather low, compared to the radius of the shells, except at a few resonance frequencies of the shells. At those frequencies, then, two different coherent waves propagate in the slab, and the theoretical derivation of the reflection and transmission of the slab, presented in Sec. III, is no longer valid.

APPENDIX A: DETERMINATION OF

$$I_p^{\pm}=\frac{1}{k^2-K^2}\int_{L'}e^{ip\theta_{j\ell}}[H_p^{(1)}(kd_{\ell j})\hat{n}\cdot\boldsymbol{\nabla}(e^{i\mathbf{K}^{\pm}\cdot\mathbf{d}_\ell})$$
$$-e^{i\mathbf{K}^{\pm}\cdot\mathbf{d}_\ell}\hat{n}\cdot\boldsymbol{\nabla}(H_p^{(1)}(kd_{\ell j}))]dl$$

L' is a closed curve composed of the circumference $L_{\beta a}$ of a circle of radius βa centered at O_j, the two straight lines that define the slab region ($x_\ell=0$ and $x_\ell=d$), and two arcs that connect the previous two straight lines at infinity

$$I_p^{\pm}=I_{L_{\beta a}}^{\pm}+I_{x=0}^{\pm}+I_{x=d}^{\pm}+I_{\text{arc}^{\text{sup}}}^{\pm}+I_{\text{arc}^{\text{inf}}}^{\pm}. \tag{A1}$$

It is rather easy to find that

$$I_{L\beta a}^{\pm}=2\pi\frac{\beta a}{K^2-k^2}e^{\pm iK_xx_j}i^{\mp p}e^{\pm ip\alpha_r}[KJ_p'(K\beta a)H_p^{(1)}(k\beta a)$$
$$-kJ_p(K\beta a)H_p^{(1)'}(k\beta a)], \tag{A2}$$

as well as to show that both integrals, $I_{\text{arc}^{\text{sup}}}^{\pm}$ and $I_{\text{arc}^{\text{inf}}}^{\pm}$, tend towards 0 as the arc radius tends to infinity.

The last two integrals, on $x=0$ and $x=d$, may be written as

$$I_{x=0}^{\pm}=\frac{1}{k^2-K^2}\left[\mp iK_x+\frac{\partial}{\partial x_\ell}\right]_{x_\ell=0}J_p(x_\ell),$$

$$I_{x=d}^{\pm}=\frac{e^{\pm iK_xd}}{k^2-K^2}\left[\pm iK_x-\frac{\partial}{\partial x_\ell}\right]_{x_\ell=d}J_p(x_\ell), \tag{A3}$$

with

$$J_p(x_\ell)=\int_{-\infty}^{+\infty}e^{iK_yy_\ell}[H_p^{(1)}(kd_{\ell j})e^{ip\theta_{j\ell}}]dy_\ell. \tag{A4}$$

Use of the Snell–Descartes law gives

$$J_p(x_\ell)=\int_{-\infty}^{+\infty}e^{ik\sin\alpha y_\ell}[H_p^{(1)}(kd_{\ell j})e^{ip\theta_{j\ell}}]dy_\ell$$
$$=\frac{1}{\pi}\int_W\int_{-\infty}^{+\infty}e^{ik_yy_\ell}e^{ikd_{\ell j}\cos\gamma}e^{ip[\gamma-(\pi/2)]}e^{ip\theta_{j\ell}}dy_\ell\,d\gamma, \tag{A5}$$

with W the Sommerfeld's integration path. $J_p(x_\ell)$ depends on whether x_ℓ is greater or less than x_j.

1. First case: $x_\ell<x_j$

As y_ℓ varies from $-\infty$ to $+\infty$, so does $\theta_{j\ell}$ from $+\pi/2$ to $-\pi/2$, through 0. Letting $\gamma_1=\gamma+\theta_{j\ell}$, with $-\pi<\theta_{j\ell}<\pi$, one gets

$$J_p(x_\ell<x_j)=\frac{1}{\pi}\int_W\int_{-\infty}^{+\infty}e^{ik\sin\alpha y_\ell}e^{ikd_{\ell j}\cos\gamma_1\cos\theta_{j\ell}}$$
$$\times e^{ikd_{\ell j}\sin\gamma_1\sin\theta_{j\ell}}e^{ip[\gamma_1-(\pi/2)]}dy_\ell\,d\gamma_1,$$

$$J_p(x_\ell<x_j)=\frac{1}{\pi}\int_W e^{ip[\gamma_1-(\pi/2)]}e^{ik(x_j-x_\ell)\cos\gamma_1}$$
$$\times\int_{-\infty}^{+\infty}e^{ik\sin\alpha y_\ell}e^{-ik\sin\gamma_1y_\ell}dy_\ell\,d\gamma_1,$$

$$J_p(x_\ell<x_j)=\frac{1}{\pi}\frac{2\pi}{k}\int_W e^{ik(x_j-x_\ell)\cos\gamma_1}e^{ip[\gamma_1-(\pi/2)]}$$
$$\times\delta(\sin\alpha-\sin\gamma_1)d\gamma_1, \tag{A6}$$

$$J_p(x_\ell<x_j)=\frac{2}{k}\int_W e^{ik(x_j-x_\ell)\cos\gamma_1}e^{ip[\gamma_1-(\pi/2)]}$$
$$\times\frac{1}{\cos\alpha}\delta(\alpha-\gamma_1)d\gamma_1,$$

$$J_p(x_\ell<x_j)=\frac{2}{k\cos\alpha}e^{ik(x_j-x_\ell)\cos\alpha}e^{ip[\alpha-(\pi/2)]}$$
$$=\frac{2}{k_x}i^{-p}e^{ik_x(x_j-x_\ell)}e^{ip\alpha}.$$

Introduction of Eq. (A6) into the first Eq. (A3) gives

$$I_{x=0}^{\pm}=\frac{2i^{-p+1}}{k_x(k^2-K^2)}(\mp K_x-k_x)e^{ik_xx_j}e^{ip\alpha}. \tag{A7}$$

2. Second case: $x_\ell>x_j$

As y_ℓ varies from $-\infty$ to $+\infty$, so does $\theta_{j\ell}$ from $+\pi/2$ to $-\pi/2$, through π. Letting $\gamma_1=\gamma+\theta_{\ell j}$, with $-\pi<\theta_{\ell j}<\pi$, and $e^{ip\theta_{j\ell}}=(-1)^pe^{ip\theta_{\ell j}}$, one gets

$$J_p(x_\ell)=(-1)^p/\pi\int_W\int_{-\infty}^{+\infty}e^{ik\sin\alpha y_\ell}e^{ikd_{\ell j}\cos\gamma_1\cos\theta_{\ell j}}$$
$$\times e^{ikd_{\ell j}\sin\gamma_1\sin\theta_{\ell j}}e^{ip[\gamma_1-(\pi/2)]}dy_\ell\,d\gamma_1,$$

which, in the same way as in the previous case, finally gives

$$J_p(x_\ell>x_j)=(-1)^pi^{-p}\frac{2}{k_x}e^{-ik_x(x_j-x_\ell)}e^{-ip\alpha}, \tag{A8}$$

and introduction of Eq. (A8) into the second Eq. (A3) provides

$$I_{x=d}^{\pm}=(-1)^p\frac{2i^{-p+1}}{k_x}\frac{e^{\pm iK_xd}}{k^2-K^2}(\pm K_x-k_x)e^{-ip\alpha}e^{-ik_x(x_j-d)}. \tag{A9}$$

Now, Eq. (A1) together with Eq. (A2) and either Eq. (A7) or Eq. (A9), leads to the final result, Eq. (21), in the main text.

APPENDIX B: THE TRANSITION MATRIX OF A CYLINDRICAL SCATTERER

For a vacuum-filled shell, use of the boundary conditions at $r=a$ and $r=b$ leads to the following linear system of equations

$$MX=S, \tag{B1}$$

with the M_{ij} elements of matrix M given, for each mode $n(n\in\mathbb{Z})$, by

$$M_{11}=(\rho_0/\rho)kaH_n^{(1)\prime}(ka), \quad M_{12}=k_La J_n(k_La),$$

$$M_{13}=k_La N_n(k_La), \quad M_{14}=nJ_n(k_La),$$

$$M_{15}=nN_n(k_Ta), \quad M_{21}=-\frac{\rho_0^2(k_La)^2}{c_0^2\rho^2}H_n^{(1)}(k_La),$$

$$M_{22}=m_1(a)=-\frac{1}{c_L^2-2c_T^2}k_La J_n(k_La) + \frac{2}{c_T^2}[(n^2-(k_La)^2)J_n(k_La)-k_La J_n'(k_La)],$$

$$M_{23}=h_1(a)=-\frac{1}{c_L^2-2c_T^2}k_La N_n(k_La) + \frac{2}{c_T^2}[(n^2-(k_La)^2)N_n(k_La)-k_La N_n'(k_La)],$$

$$M_{24}=u_1(a)=\frac{2}{c_T^2}n[k_Ta J_n'(k_Ta)-J_n'(k_Ta)],$$

$$M_{25}=v_1(a)=\frac{2}{c_T^2}n[k_Ta N_n'(k_Ta)-N_n'(k_Ta)]$$

$$M_{31}=M_{41}=M_{51}=0, \quad M_{32}=m_1(b), \tag{B2}$$

$$M_{33}=h_1(b), \quad M_{34}=u_1(b), \quad M_{35}=v_1(b),$$

$$M_{42}=m_2(a)=2n[1-k_La J_n'(k_La)],$$

$$M_{43}=h_2(a)=2n[1-k_La N_n'(k_La)],$$

$$M_{44}=u_2(a)=-(n^2-(k_Ta)^2)J_n(k_Ta)+2J_n'(k_Ta),$$

$$M_{45}=v_2(a)=-(n^2-(k_Ta)^2)N_n(k_Ta)+2N_n'(k_Ta),$$

$$M_{52}=m_2(b), \quad M_{53}=h_2(b), \quad M_{54}=u_2(b),$$

$$M_{55}=v_2(b),$$

with k_L (respectively, k_T) the wave number of the longitudinal (respectively, shear) waves in the solid, ρ_0 and ρ the densities of the surrounding fluid and of the solid, a the outer radius of the shell, and b its inner one.

If the components S_j of the source vector in Eq. (B1) are

$$S_j=-\Re e(M_{j1}), \quad (j=1,5), \tag{B3}$$

the nth element T_{nn} of the diagonal transition matrix is the first component of the unknown vector X; as such, it is obtained from system (B1) by use of the Cramer's rule

$$T_{nn}=\frac{D_n^{[1]}}{D_n}, \tag{B4}$$

with D_n the determinant of matrix M, and $D_n^{[1]}$ the same one, except that column 1 is replaced by vector S.

In the case of a cylinder, rather than a shell, system (B1) reduces to a linear system of three equations and three unknowns, obtained by suppressing rows 3 and 5, in the whole system, together with columns 3 and 5 in matrix M.

The reduced resonance frequencies $(ka)_{\text{res}}$ are the real parts of the complex roots of the following equation:

$$D_n(ka)=0. \tag{B5}$$

[1] L. L. Foldy, "The multiple scattering of waves. I. General theory of isotropic scattering by randomly distributed scatterers," Phys. Rev. **67**(3,4), 107–119 (1945).

[2] D. Sornette, "Acoustic waves in random media," Acustica **67**, 199–215 (1989); **67**, 251–265 (1989); **68**, 15–25 (1989).

[3] A. Tourin, M. Fink, and A. Derode, "Multiple scattering of sound," Waves Random Media **10**, R31–R60 (2000).

[4] M. Lax, "Multiple scattering of waves," Rev. Mod. Phys. **23**(4), 287–310 (1951).

[5] S. K. Bose and A. K. Mal, "Longitudinal shear waves in a fiber-reinforced composite," Int. J. Solids Struct. **9**, 1075–1085 (1973).

[6] V. K. Varadan, V. V. Varadan, and Y.-H. Pao, "Multiple scattering of elastic waves by cylinders of arbitrary cross section. I. SH waves," J. Acoust. Soc. Am. **63**(5), 1310–1319 (1978).

[7] R. B. Yang and A. K. Mal, "Multiple scattering of elastic waves in a fiber-reinforced composite," J. Mech. Phys. Solids **42**(12), 1945–1968 (1994).

[8] P. C. Waterman and R. Truell, "Multiple scattering of waves," J. Math. Phys. **2**(4), 512–537 (1961).

[9] M. Lax, "Multiple scattering of waves. II. The effective field in dense systems," Phys. Rev. **85**(4), 621–629 (1952).

[10] V. Twersky, "On scattering of waves by random distributions. I. Free-space scatterer formalism," J. Math. Phys. **3**(4), 700–715 (1962).

[11] J. G. Fikioris and P. C. Waterman, "Multiple scattering of waves. II. 'Hole corrections' in the scalar case," J. Math. Phys. **5**(10), 1413–1420 (1964).

[12] C. Aristégui and Y. C. Angel, "New results for isotropic point scatterers: Foldy revisited," Wave Motion **36**(4), 383–399 (2002).

[13] Y. C. Angel, C. Aristégui, and J. Y. Chapelon, "Reflection and transmission of plane waves by anisotropic line scatterers," Poromechanics II, edited by Auriault, Geindreau, Royer, Bloch, Boutin, and Lewandowska (Sweets & Zeitlinger, Lisse, 2002), pp. 607–618.

[14] A. Ishimaru, *Wave Propagation and Scattering in Random Media* (Academic, New York, 1978).

[15] M. Abramovitz and I. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1964).

[16] J. M. Conoir, in *La Diffusion Acoustique par des Cibles Élastiques de Forme Géométrique Simple*. Théories et Expériences (Acoustic Scattering by Elastic Targets of Simple Geometric Form), edited by N. Gespa (CEDOCAR, Paris, 1987), Chap. 5.

[17] L. Tsang and J. Au Kong, *Scattering of Electromagnetic Waves—Advanced Topics*, Wiley Series in Remote Sensing (Wiley, New York, 2001).

[18] S. Torquato, *Random Heterogeneous Materials—Microstructure and Macroscopic Properties*, Interdisciplinary Applied Mathematics Volume 16 (Springer, New York, 2002).

[19] W. Sachse and Y.-H. Pao, "On the determination of phase and group velocities of dispersive waves in solids," J. Appl. Phys. **49**(8), 4320–4327 (1978).

[20] F. Vander Meulen, G. Feuillard, O. Bou Matar, F. Levassort, and M. Lethiecq, "Theoretical and experimental study of the influence of the particle size distribution on acoustic wave properties of strongly inhomogeneous media," J. Acoust. Soc. Am. **110**(5), 2301–2307 (2001).

[21] P. Y. Le Bas, F. Luppé, J. M. Conoir, and H. Franklin, "*N*-shell cluster in water: Multiple scattering and splitting of resonances," J. Acoust. Soc. Am. **115**(4), 1460–1467 (2004).

[22] X. Jing, P. Sheng, and M. Zhou, "Theory of acoustic excitations in colloidal suspensions," Phys. Rev. Lett. **66**(9), 1240–1243 (1991).

[23] X. Jing, P. Sheng, and M. Zhou, "Acoustic and electromagnetic quasimodes in dispersed random media," Phys. Rev. A **46**(10), 6513–6534 (1992).

Material property measurement using the quasi-Scholte mode—A waveguide sensor

F. B. Cegla, P. Cawley, and M. J. S. Lowe
Department of Mechanical Engineering, Imperial College London, London SW7 2AZ, United Kingdom

(Received 8 July 2004; revised 2 November 2004; accepted 9 November 2004)

In the food industry and other industries, rheological measurements and determination of particle sizes in suspensions and emulsions is of great importance for process and quality control. Current test cell based ultrasonic methods exist but are often inconvenient. An attractive alternative could be to insert a simple measurement ''dipstick'' into the fluid; this paper presents an initial study of the feasibility of using measurements of the velocity and attenuation of the quasi-Scholte mode on a plate to obtain the longitudinal velocity and attenuation of an embedding medium. The attenuation of the quasi-Scholte mode is caused by two mechanisms: shear leakage and attenuation due to the bulk longitudinal attenuation of the embedding material. In a calibration test the bulk longitudinal velocity and viscosity of glycerol were determined experimentally. Measurements agreed well with results from conventional methods and literature data. Quantitative results and an independent validation for honey, a very viscous fluid, are also presented. For Newtonian liquids like glycerol and honey, the shear leakage and longitudinal bulk attenuation are both related to viscosity. To demonstrate the sensitivity to nonviscous attenuation mechanisms, qualitative measurement results on fluids that mainly exhibit attenuation due to scattering are presented. [DOI: 10.1121/1.1841631]

PACS numbers: 43.35.Mr [YHB] Pages: 1098–1107

I. INTRODUCTION

Every day millions of liters of complex fluids are produced in the food industry and other industries. Examples of these fluids are honey, ketchup, mayonnaise, and chocolate sauce. These products contain many different ingredients mixed together; some are dissolved in solution while others are present as solid or liquid particles. The different phases of the ingredients give the product a unique texture that influences the taste.[1] Thus, to obtain the best results during production of foods, process control and quality control techniques are needed to optimize the production. Currently ultrasonic methods can be used for particle size distribution measurements,[2] which is a step towards characterizing the texture of a fluid.

In the frequency domain the mechanical properties of isotropic materials can be fully described by density and a complex wave number for shear and longitudinal waves. Conventional ultrasonic material property measurements are carried out in test cells. The measured velocity and attenuation can be converted back into a complex wave number that describes the material properties. These techniques are standard and widely used.[3–5] For accurate results the manufactured test cells have to contain precisely aligned ultrasonic transducers and the measured total attenuation should lie between 0.5–1 np (neper is a unit of attenuation that corresponds to a reduction in amplitude by factor e^{-x}, where x is in nepers) in order to avoid large error magnification.[6] In a pulse–echo measurement the reflector properties have to be determined accurately, and also beam-spreading losses have to be minimized. Other practical drawbacks are the difficulty of integrating such a test cell into a reaction vessel that contains a stirring mechanism. The investigation of highly attenuative materials is also difficult since a very small separation between transducer and reflector is required, which does not allow high flow rates through the cell.

Alternative material property measurements using guided waves have recently been of interest. Lynnworth[7] was the first to describe a fluid level meter using a waveguide. Kim and Bau[8] further developed the idea and measured density and viscosity using a waveguide. Vogt[9,10] described the theory behind the use of waveguide rods and showed how to extract the shear velocity as a function of frequency (related to viscosity in the case of a Newtonian fluid) as well as the longitudinal velocity of an embedding liquid using two different guided-wave modes. The technique was also extended to monitor the cure of epoxies. However, all these waveguide techniques are utilizing the effect of energy leakage from the waveguide into the surrounding medium. Leakage occurs when the phase velocity of the guided wave is supersonic with respect to the bulk velocity of the embedding medium. Leakage is the generation of waves in the embedding material caused by the displacement and stress fields of the waves traveling in a waveguide.[11] The generated waves radiate energy away from the waveguide. The resulting attenuation of the guided wave can be related to the impedance of the embedding medium.[12] Methods involving leaky waves measure the effect of the fluid acoustic impedance on the waveguide surface; however, in most fluids the bulk attenuation is small and thus the impedance depends mainly on the bulk velocity and not the attenuation (e.g., in glycerol the real part of the complex longitudinal velocity is approximately 2000 m/s, while the imaginary part is about 2 m/s). Therefore, techniques involving leaky waves are very insensitive to the bulk longitudinal attenuation of most liquids.

An attractive alternative to determine the longitudinal bulk velocity and attenuation of the embedding material is

the use of a measurement ''dipstick'' that employs a ''nonleaky'' guided wave mode. The term ''nonleaky'' here is strictly speaking only valid for an ideal fluid (see Sec. III). In the nonleaky case the wave energy in the embedding material is confined to a region close to the waveguide surface. Any attenuation will be due to energy dissipation in the material rather than energy leakage away from the waveguide system. The guided-wave mode that this paper focuses on is very similar to the Scholte interface wave that is used in geophysics.[13] In NDE the Scholte wave is seldom used because of the high proportion of energy that it carries in the fluid;[14] exactly this property makes it attractive to extract bulk longitudinal properties of liquids. The actual mode that is investigated in this paper is a plate mode. It is termed the quasi-Scholte mode, which asymptotes to the Scholte-interface wave behavior at high frequencies (This is similar to quasi-Rayleigh waves turning into Rayleigh waves at high frequencies). Since the quasi-Scholte mode is closely related to the Scholte interface wave, the characteristics of the Scholte wave are discussed initially.

II. PROPERTIES OF THE SCHOLTE WAVE

At the interface of a liquid and solid there exist two propagating solutions to the governing equation: the leaky Rayleigh wave and the Stoneley–Scholte wave.[15] This paper focuses on the latter and will for the sake of clarity refer to it as the Scholte wave. The Scholte wave is a special case of a Stoneley wave (on a solid/liquid interface) which was first pointed out by Scholte.[16] The properties of an interface wave can be theoretically modeled using the partial wave technique as described in Refs. 14 and 17. This yields the phase velocity (or wave number)–frequency relationship that describes the mode. In this special case of a wave propagating along the interface of an elastic solid and an ideal fluid the mode is nondispersive, i.e., the phase velocity does not change with frequency.

The mode shape of the Scholte wave reveals that most of the displacements occur in the fluid parallel to the propagation direction. A small component of displacement normal to the propagation direction exists as well as small displacements in the solid. The displacements decay exponentially away from the surface, which makes the Scholte wave an evanescent wave guided along the interface. The fraction of the total wave energy that travels in the fluid as compared to the solid depends on the materials that form the interface. In general, for a stiff and dense solid compared to the fluid, most of the energy will travel in the fluid, while for a lower modulus and density solid more energy will travel in the solid.

The theoretical description of the Scholte wave has been thoroughly investigated by many researchers[15,18,19] and the reader is referred to these papers and the book by Rose[14] for more details. It can be shown that the phase velocity of the Scholte wave is limited by the lower of the fluid and the transverse bulk velocity of the elastic medium ($c_{\text{Scholte}} < c_{\text{liquid}}, c_{t,\text{solid}}$). The dependence of the Scholte wave velocity on the density and longitudinal velocity ratio of the neighboring materials can be displayed on a contour map as shown in Ref. 20. Similar contour maps to that of Ref. 20 can be created to visualize the distribution of energy flow in the two materials. The quasi-Scholte plate mode that is mainly dealt with in this paper asymptotes to a pure Scholte wave at high frequencies.

Many ways of exciting Scholte waves have been reported in the literature (see, e.g., Refs. 21–25). Nasr *et al.*[21] and McLean *et al.*[22] use a comb-like transducer (PVDF or micromachined) at the interface of the two materials to excite Scholte waves. Desmet *et al.*[23] use laser excitation to create the Scholte wave at the interface. DeBilly *et al.*[24] and Matula *et al.*[25] use the concept of mode conversion from a Rayleigh (or $A0$ wave on a plate) to the Scholte wave. They generate a clean mode on an unimmersed section of the waveguide using conventional techniques, the wave then propagates into the liquid and part of its energy converts into the Scholte mode. This latter approach of mode conversion from an $A0$ plate wave was also chosen for the experiments in this paper. The method is reliable, cheap, and easily set up. It also separates the transducer from the embedding medium and is in essence already a dipstick.

A. Modeling a viscous fluid

Real fluids have dissipation mechanisms. This is usually modeled by making the fluids viscous; here, the fluid is assumed to be Newtonian. The viscous medium is modeled as a solid with appropriate parameters. More theoretical details on this approach are described in Refs. 26–28.

In order to obtain the Scholte equation for the viscous case, the real bulk velocities have to be replaced by complex bulk velocities, which are given by the following expression:

$$c_{\text{bulk}} = \frac{\omega}{k} = \frac{\omega}{k_r - ik_i} = \frac{c_p}{1 - i\dfrac{k_i}{k_r}} = \frac{c_p}{1 - i\dfrac{\kappa}{2\pi}}, \tag{1}$$

where c_{bulk} is the complex bulk velocity, ω is the angular frequency, k is the complex wave number with real part k_r and imaginary part k_i, c_p is the phase velocity of the wave (i.e., ω/k_r) and κ is the attenuation per wavelength. Both c_p and κ are real quantities. Thus

$$\alpha_n = \frac{c_l}{1 - i\dfrac{\kappa_l}{2\pi}}, \tag{2}$$

$$\beta_n = \frac{c_s}{1 - i\dfrac{\kappa_s}{2\pi}}, \tag{3}$$

where α_n is the complex longitudinal bulk velocity, β_n is the complex transverse bulk velocity, and the subscript n indicates the material. c_l is the longitudinal bulk wave phase velocity, κ_l is the longitudinal bulk wave attenuation (per wavelength), c_s stands for the shear bulk wave phase velocity, and κ_s for the shear bulk wave attenuation (per wavelength).

The phase velocities and attenuation of the fluid are real quantities and can be derived from the material properties by the following expressions:

$$c_l = \left(\frac{K_f}{\rho} \right)^{1/2}, \tag{4}$$

$$\kappa_l = \left(\frac{4\pi\omega\nu}{3C_l^2} \right), \tag{5}$$

$$c_s = (2\omega\nu)^{1/2}, \tag{6}$$

$$\kappa_s = 2\pi, \tag{7}$$

where K_f is the real part of the fluid bulk modulus, ρ is the density of the fluid, ω is the angular frequency, and ν is the kinematic viscosity of the fluid.

Equations (4) and (5) assume low viscosities. This allows simplifications such as the use of the real part of the bulk modulus in Eq. (4) and the expression in Eq. (5).[27] This introduces errors of about 1% in velocity and 3.3% in attenuation for a fluid with a viscosity of 100 Pas at 1 MHz. For fluids like glycerol the error is of the order of 0.002% for the velocity and 0.01% for the attenuation; these errors are much smaller than the measurement uncertainties.

By utilizing the above equations in the analysis described in the references of Sec. II where applicable, a new characteristic equation and mode shape can be found. The Scholte wave in this system becomes dispersive and attenuated. However, the dispersion effect is relatively weak and for small viscosities the velocity characteristics are virtually the same as in the nonviscous case except that now the wave is attenuated.

In experiments a finite thickness plate rather than a solid half-space was used; the system consisted of a plate surrounded by a liquid. The analysis of the system is summarized in the following section.

III. THE QUASI-SCHOLTE PLATE MODE

A. Modeling of the quasi-Scholte plate mode

The quasi-Scholte mode was modeled using the global matrix approach. For more details on the exact modeling procedure the reader is referred to Lowe.[17] A matrix containing the general solutions for wave propagation in the materials and all the boundary conditions (continuity of stresses and displacements at the interfaces) is set up and a solution sought. The matrix equation for a plate immersed in two half-spaces of viscous fluid is

$$[A] \cdot \begin{pmatrix} L_{f-} \\ T_{f-} \\ L_{s+} \\ L_{s-} \\ T_{s+} \\ T_{s-} \\ L_{f+} \\ T_{f+} \end{pmatrix} = 0 \tag{8}$$

The terms of matrix A are given in the Appendix; L and T are the amplitudes of the partial longitudinal and transverse waves. The subscripts f and s refer to the fluid and solid, respectively, and the $+$ and $-$ signs indicate the direction of the partial wave.

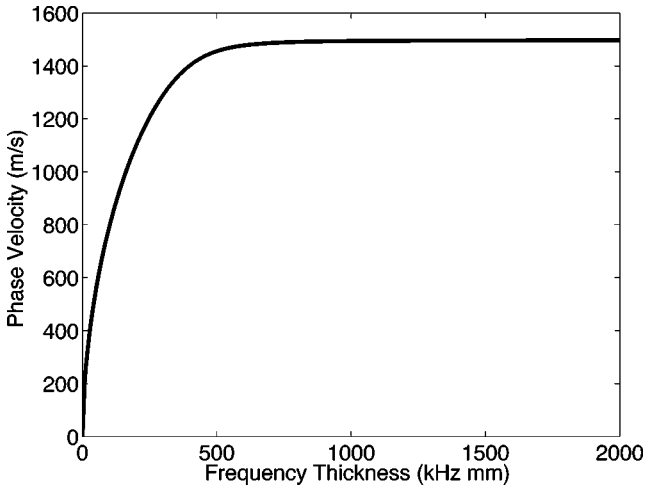


FIG. 1. The phase velocity dispersion of the quasi-Scholte mode on an aluminum plate surrounded by water (see the text for material properties).

A solution exists if matrix A in Eq. (8) is singular, i.e., when the determinant of the matrix equals zero. There are several wave numbers that satisfy the equation at each frequency. One of these solutions in the wave number frequency space is the quasi-Scholte mode. The complex wave number gives the phase velocity and attenuation of the mode. Also, by substituting it into Eq. (8) and setting one of the partial wave amplitudes to unity, the mode shape of the mode can be evaluated at each frequency.

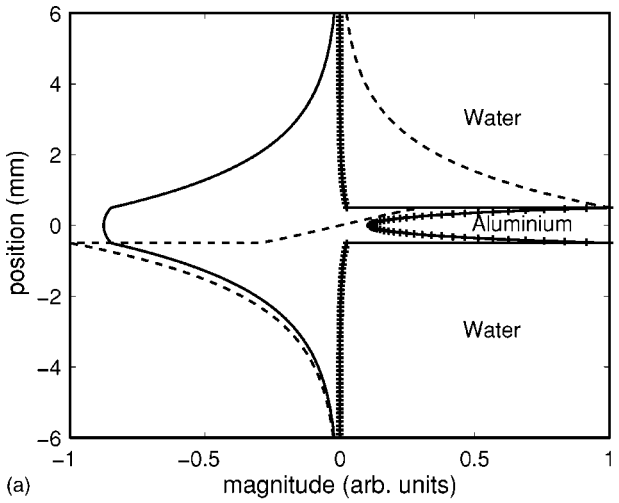
The solution techniques for multiple material layer stacks of solid materials and viscous fluids are well known and have been implemented in a software package called DISPERSE.[27] In the next section the DISPERSE software was used to determine the dispersion curves and mode shapes of the quasi-Scholte mode.

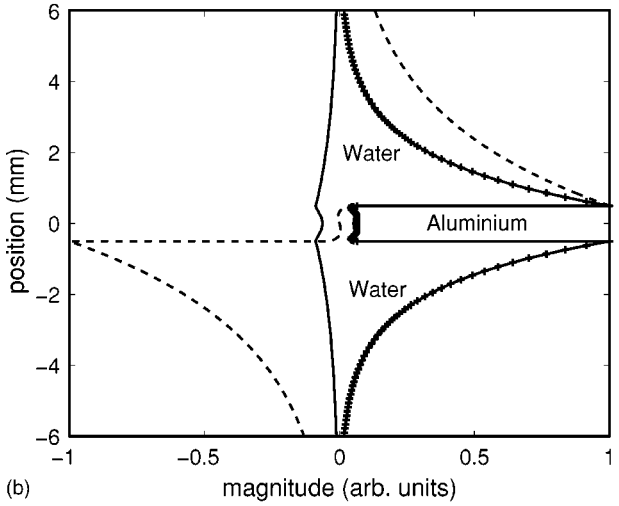
B. Properties of the quasi-Scholte plate mode

This section discusses the quasi-Scholte mode on an embedded plate. First an ideal fluid is used as the embedding medium and then a viscous liquid. The velocity dispersion characteristics are virtually unchanged by the introduction of viscosity, but the attenuation is noticeably affected.

An aluminum plate ($c_l = 6320$ m/s, $c_s = 3130$ m/s, $\rho_{\mathrm{Al}} = 2700$ kg/m^3, thickness$=1$ mm) is embedded in an infinite space of water ($c_l = 1500$ m/s, $\rho_{\mathrm{water}} = 1000$ kg/m^3). Figure 1 shows the phase velocity dispersion curve of the quasi-Scholte mode. The phase velocity of the quasi-Scholte mode rises with frequency from zero and gradually asymptotes to the nondispersive Scholte wave velocity of two elastic half-spaces. A physical explanation of the asymptotic behavior is the decrease of mode wavelength with increasing frequency. At a certain point the wavelength will be small compared to the thickness of the plate. Since the Scholte wave displacements decay in an exponential manner away from the interface, the mode resembles a mode on an infinite half-space at high frequencies, whereas at lower frequencies the two interfaces of the plate interact.

The similarities of the quasi-Scholte mode and the $A0$ mode in a free plate are highlighted by their mode shapes. Figure 2(a) shows that the out-of-plane displacement compo-





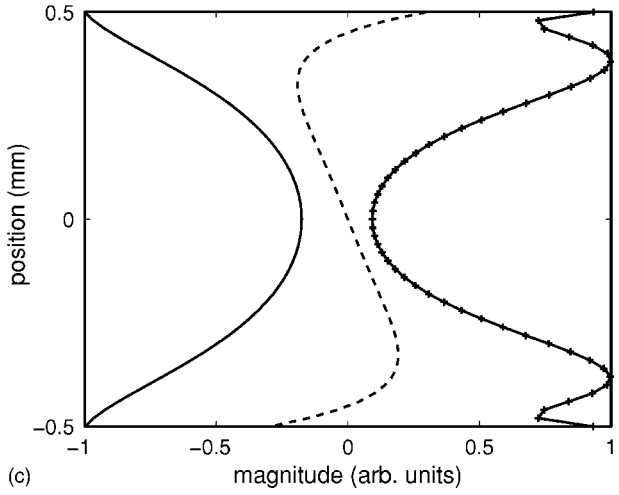


FIG. 2. Mode shapes of the quasi-Scholte mode (water/aluminum) at frequency thicknesses (a) 0.1 MHz-mm; (b) 1 MHz-mm; (c) 2 MHz-mm (in-plate only) [(———) out-of-plane displacement; (---) in-plane displacement; (+++) strain energy density].

nent at 100 kHz-mm is almost constant across the section of the plate, and the strain energy density indicates that most of the energy is propagating in the plate. At higher frequencies most of the energy is traveling in the fluid as shown in Fig. 2(b). Figure 2(c) shows the mode shape in the plate alone at

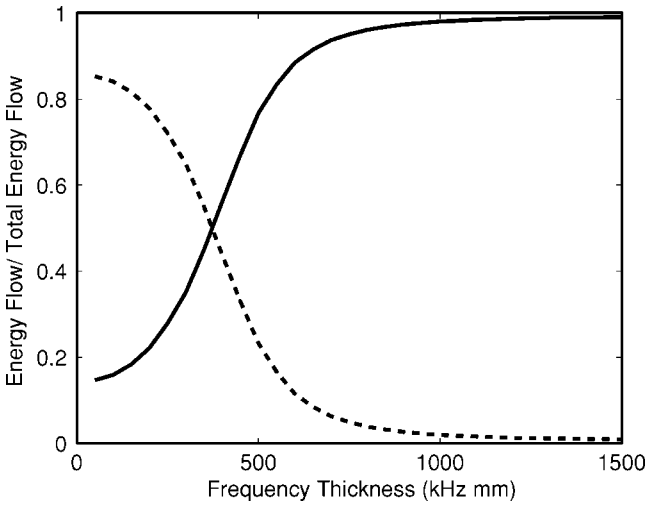


FIG. 3. Quasi-Scholte energy flow localized in the fluid (water) and plate (aluminum) as a function of frequency [(———) energy flow in fluid; (---) energy flow in plate].

2 MHz-mm. Displacements decay away from the surfaces and are a minimum at the center of the plate. This illustrates the effect of the quasi-Scholte mode asymptoting to an interface wave (Scholte wave) at high frequencies, as discussed above. The phenomenon is also well known in plate waves where the $A0$ and $S0$ modes asymptote to the Rayleigh wave solution.[29]

An important property of the quasi-Scholte mode is its change in energy partition between fluid and solid with frequency. Figure 3 shows the fraction of energy flow that is carried in the plate and fluid at different frequencies. At low frequencies ($<$500 kHz-mm) the energy flows predominantly in the plate, while at high frequencies ($>$1 MHz-mm) most of the energy travels in the fluid. The crossover point of the two curves occurs at about 400 kHz-mm. This means that the mode can be made more or less sensitive to the embedding medium properties. If highly attenuating materials are surrounding the waveguide, careful frequency selection will allow an appropriate fraction of energy to travel in the plate so that a reasonable propagation range can be obtained. However, in the unusual case that the material damping is so high that the damping distance is smaller than the travel path of the wave, the method fails. In this case a leaky mode that measures the impedance of the embedding medium should be used, as now the imaginary part of the bulk velocity has become dominant (see Sec. I).

There are two attenuation mechanisms for the quasi-Scholte and Scholte modes. The first is the leakage of shear waves from the guided wave into the embedding viscous fluid, and the second is the attenuation of longitudinal bulk waves in the viscous fluid. Energy that is guided along the interface in the form of an evanescent longitudinal wave in the fluid will be attenuated by fluid longitudinal bulk attenuation. In the case of a purely Newtonian viscous fluid this attenuation is due to viscosity and can be modeled by Eq. (5) (assuming low viscosities[30]). Also, for Newtonian fluids both attenuation mechanisms depend on the viscosity only [see Eqs. (5), (6), (7)]. At low frequencies the shear leakage is dominant, while at higher frequencies the evanescent longi-

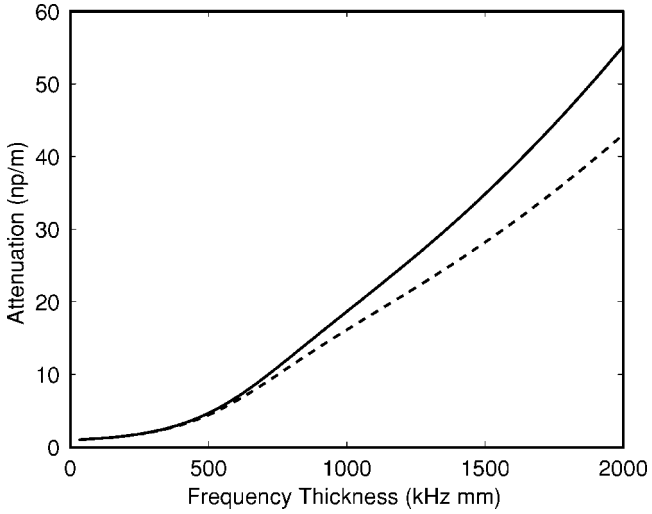


FIG. 4. Total attenuation (———) of the quasi-Scholte mode and attenuation due to shear leakage only (---) as a function of frequency for an aluminum plate (ρ=2700 kg/m^3, C_l=6320 m/s, C_s=3130 m/s) immersed in glycerol (ρ=1258 kg/m^3, C_l=1900 m/s, η=1 Pas).

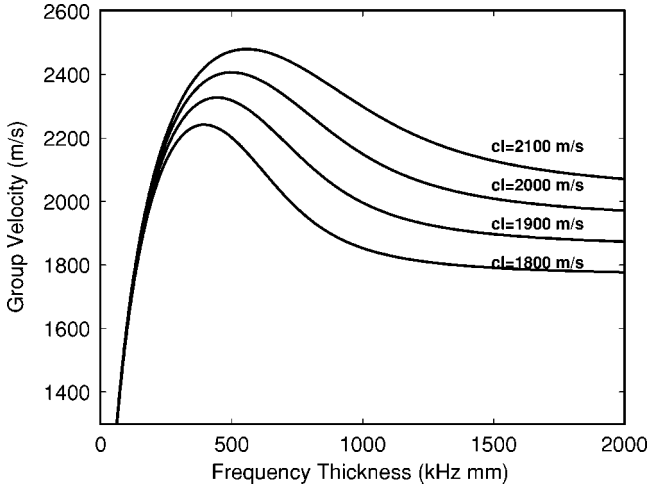


FIG. 5. Quasi-Scholte mode group velocities as a function of frequency for different longitudinal bulk velocities (c_l) (The other properties are as in Fig. 4).

tudinal wave attenuation contributes significantly to the quasi-Scholte mode attenuation. Figure 4 shows the total attenuation of the quasi-Scholte mode on an aluminum plate (c_l=6320.0 m/s, c_s=3130 m/s, ρ_{al}=2700 kg/m^3, thickness=1 mm) immersed in glycerol (c_l=1860 m/s, ρ_{glycerol}=1258 kg/m^3 η_{glycerol}=1 Pas). A line showing the attenuation due to shear leakage only is also shown (the lines are traced in DISPERSE; for the shear leakage-only case the longitudinal attenuation (κ_l) is set to zero). The difference between the two curves is linked to the longitudinal bulk attenuation of the fluid. Thus, if the longitudinal bulk velocity and the shear properties of the fluid are known, it is possible to measure the longitudinal bulk attenuation of the fluid surrounding the waveguide.

Figure 4 considers a viscous liquid, and thus the shear properties of the liquid are described by Eqs. (6) and (7) so the longitudinal attenuation and the shear velocity of the embedding medium depend solely on the viscosity and frequency. Therefore, the longitudinal bulk velocity and viscosity of the fluid can be extracted from the measured quasi-Scholte mode group velocity and attenuation. Figure 5 shows the quasi-Scholte mode group velocity as a function of frequency for different longitudinal velocities of the embedding fluid. For small viscosities the group velocity of the quasi-Scholte mode is almost independent of viscosity; the group velocity changes by about 10 m/s ($\sim$0.5% of C_l) when the viscosity changes from 0 to 1 Pas.

Figure 6 displays the theoretical attenuation of the quasi-Scholte mode as a function of frequency for different viscosities at a fixed longitudinal bulk velocity of the embedding medium. The assumption of a Newtonian viscous fluid links the bulk shear velocity and bulk longitudinal attenuation of the fluid to the viscosity [see Eqs. (5)–(7)]; they therefore are not independent properties. However, the longitudinal velocity and attenuation of more complex embedding fluids can still be extracted from the properties of the quasi-Scholte mode if the shear velocity of the embedding medium is known or determined using a method such as that described by Vogt.[9]

IV. EXPERIMENTAL SETUP AND RESULTS

A. The setup

The experimental setup used was designed to excite the quasi-Scholte mode on a plate immersed in a fluid and to measure its group velocity and attenuation. Figure 7 shows a schematic of the apparatus. A Panametrics 500-kHz shear transducer was attached to the center of an end of a 0.94-mm-thick aluminum plate; the plate was 100 mm wide and 200 mm long. A vessel containing a fluid sample was placed beneath the plate on a table of variable height. By changing the height of the table the plate could conveniently be immersed in the fluid to different depths; the angle between the fluid surface and the plate was 90 deg. The room temperature was recorded. The signal was sent and received by a waveform generator (Macro Design Ltd.); a LeCroy 9400A stor-

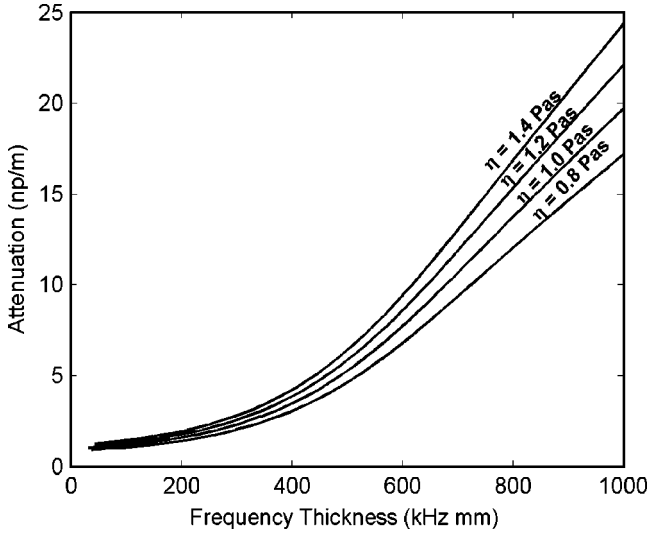


FIG. 6. Quasi-Scholte mode attenuation as a function of frequency for different viscosities (η) at c_l=1800 m/s (The other properties are as in Fig. 4).

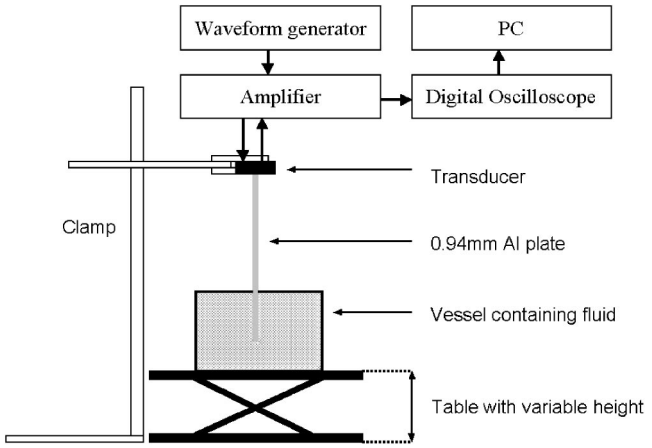


FIG. 7. Experimental setup.

age oscilloscope was used to store the signal and the data were then transferred to a PC for processing. The transducer was excited with a Hanning windowed 10-cycle tone burst. The excited $A0$ wave traveled along the plate until it hit the fluid surface. Here, part of the wave was reflected back as an entry reflection and the rest converted into modes of the fluid–plate system: a highly attenuated leaky $A0$ mode and the quasi-Scholte mode; after a short propagation distance only the quasi-Scholte mode remained. It was reflected by the end of the plate and converted back into an $A0$ mode at the point where the plate leaves the liquid. (Some energy was also reflected back into the fluid.) The wave then propagated along the plate until it reached the transducer again. A typical time trace is displayed in Fig. 8. The entry reflection, plate end reflection, and a reverberation in the fluid are clearly visible. Data for tone bursts at different center frequencies and two different immersion depths were collected.

The group velocity and attenuation of the quasi-Scholte mode were then extracted from the collected data using the following equations:

$$C_g = \frac{2(x_2 - x_1)}{\left(\Delta T + \frac{2(x_2 - x_1)}{C_{gA0}}\right)}, \tag{9}$$

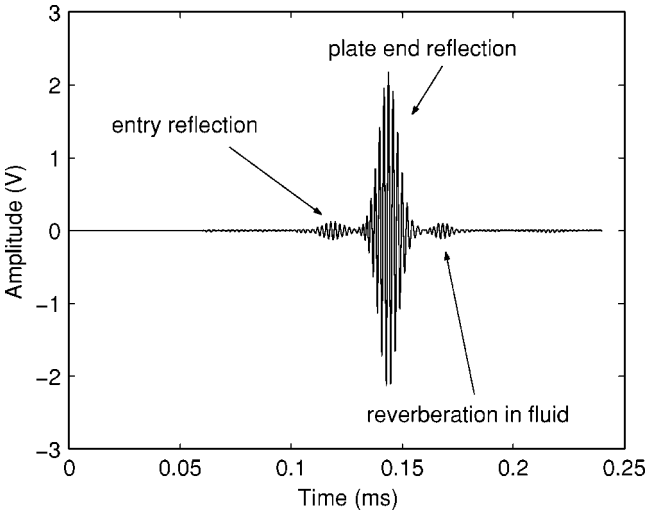


FIG. 8. Time trace at 500 kHz with plate 30 mm immersed in glycerol.

$$\alpha = -\frac{1}{2(x_2 - x_1)} \ln\left(\frac{S_2(\omega)}{S_1(\omega)}\right), \tag{10}$$

where $x_1 < x_2$ are two different immersion depths, ΔT is the time difference between the arrival of the wave packages at the two immersion states, C_{gA0} is the group velocity of the $A0$ plate mode (without fluid loading), and S_1 and S_2 are the signal amplitudes at the respective immersion depths.

In order to eliminate the effects of dispersion on the measured attenuation, the magnitude of the Fourier transform at the center frequency of each tone burst was used in Eq. (10). The difference in arrival time ΔT was evaluated by cross correlating the signal envelopes at the two immersion depths. The maximum of this function corresponds to the time shift between the two signals.[25]

The wave number in both systems (plate only and plate–fluid) is very similar so that beam spreading effectively stays constant. It was verified that the plate was large enough to ensure that reflections from the plate edges were not present in the recorded signal.

To obtain the viscosity and bulk longitudinal velocity of the embedding viscous fluid at each measurement frequency, the viscosity and bulk longitudinal velocity in the disperse model are adjusted so that the theoretically modeled quasi-Scholte mode attenuation agrees with the measured attenuation within 0.01 np/m and the modeled group velocity with the measured group velocity within 0.1 m/s. Then, the mean value of the bulk velocity and viscosity and their standard errors are calculated from the set of results obtained at different frequencies. This approach, however, is only valid for purely viscous liquids where the longitudinal bulk attenuation and the shear velocity depend entirely on the viscosity. In suspensions, for example, the longitudinal bulk attenuation of the suspension does not only depend on viscosity; thus, the quasi-Scholte mode attenuation also will not entirely depend on viscosity.

B. Results

1. Glycerol

The density of a glycerol (98%–99% pure) sample was measured to be 1258 kg/m^3 and the aluminum plate properties were evaluated experimentally as $\rho = 2700$ kg/m^3, $C_l = 6320$ m/s, and $C_s = 3130$ m/s. The temperature was recorded to be 25 °C. Results were extracted from the time traces as described in subsection *A*. Figure 9(a) shows the measured group velocity of the quasi-Scholte mode as a function of frequency, and theoretically modeled curves (DISPERSE[27]) for different fluid bulk velocities are also displayed. For the low viscosities of glycerol, the group velocity dispersion relation of the quasi-Scholte mode is essentially only dependent on the longitudinal velocity as described in Sec. II. Thus, it is possible to extract the longitudinal bulk velocity of the glycerol as 1899 ± 2 m/s. Figure 9(b) shows the measured and theoretically predicted attenuation for the quasi-Scholte mode. The mean viscosity evaluated from the measured attenuation data is 0.64 ± 0.03 Pas.

To validate these results, a measurement of the bulk longitudinal velocity was carried out. This yielded a value of 1900 m/s. Also, an alternative method of measuring viscosity

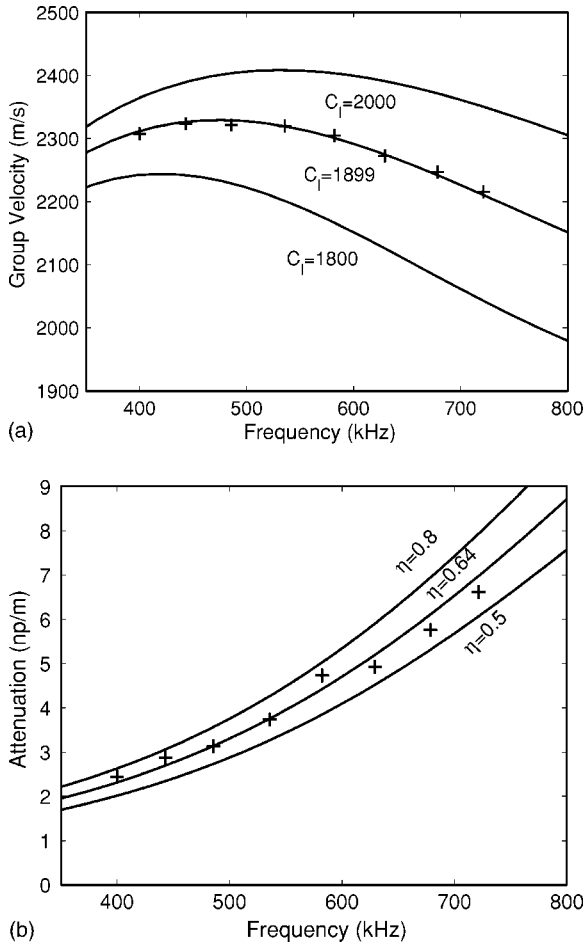


FIG. 9. Measured (+) and for different properties theoretically predicted (−) (a) group velocity and (b) attenuation of the quasi-Scholte mode on a 0.94-mm-thick aluminum plate immersed in glycerol.

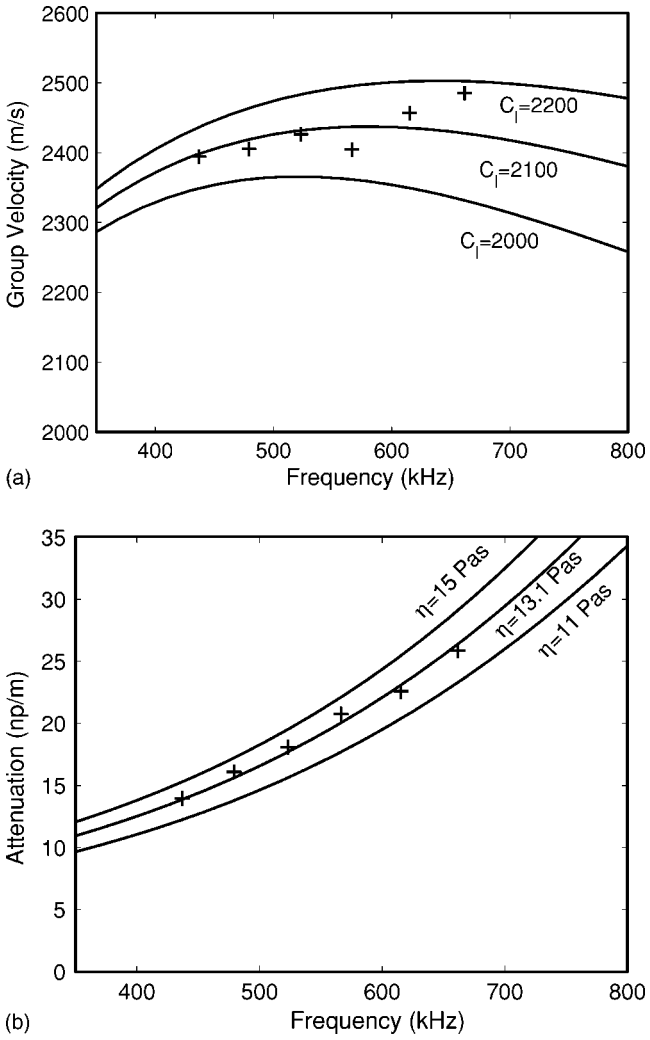


FIG. 10. Measured (+) and for different properties theoretically predicted (−) (a) group velocity and (b) attenuation of the quasi-Scholte mode on a 0.94-mm-thick aluminum plate immersed in honey.

was utilized; this involved the measurement of the $SH0$ shear horizontal wave mode attenuation when immersed in glycerol. This mode is the equivalent mode in a plate to the torsional $T(0,1)$ mode in a rod that was used by Vogt *et al.*[10] to measure viscosity. Using this approach the viscosity was determined to be 0.62 ± 0.02 Pas. The glycerol viscosities measured using both methods agree within (4%) and the range of their standard errors overlaps. The measured results also compare well to literature data. The viscosity of 98% glycerol was determined from Kaye[31] to be 0.614 Pas at 25 °C. The viscosity of glycerol strongly depends on temperature and the water content of the sample. A change in temperature of 1 °C can change the viscosity by 0.1 Pas and a change in water content by 1% can shift the viscosity by 0.2 Pas.[31] Therefore, exact agreement with literature data would not be expected since glycerol is hygroscopic and was open to air between tests.

The extracted longitudinal bulk velocity and viscosity from the experiments agree well with results from established techniques (bulk velocity measurement and attenuation of the $SH0$ mode). The independently measured velocities agree extremely well. The most significant error is believed to be due to positioning errors of the adjustable table and the plate. They are of the order of 1%. Literature values for the longitudinal bulk velocity of glycerol at 25 °C are inconsistent and can be found to be 1920 m/s[31] and 1904 m/s.[32] Despite these inconsistencies these values agree well (within 1%) with the measurements. The differences can possibly be due to differences in water content of the glycerol samples used.

The consistency of the quasi-Scholte mode measurements compared to that of an established technique was also investigated. Sets of five different measurements on different days and at different temperatures were carried out with the $SH0$ and quasi-Scholte mode techniques; the two measurements on a given day were carried out almost simultaneously. The mean difference in measured viscosity between the two methods was 1.4%, confirming the consistency of the proposed quasi-Scholte mode method.

2. Honey

The quasi-Scholte mode was also used to measure the properties of honey. Figure 10(a) shows the measured group velocity and Fig. 10(b) shows the measured attenuation of the quasi-Scholte mode of the 0.94-mm aluminum plate immersed in honey [Gale's Clear Honey, Premier Ambient Products (UK) Ltd., Spalding PE12 9EQ]. The average bulk

velocity of the honey was determined to be 2100±17 m/s and the average viscosity was evaluated at 13.1±0.2 Pas. A viscosity measurement using the $SH0$ mode resulted in 13.6 ±0.6 Pas.

The results for honey show that the method can also retrieve data from very viscous fluids and both methods agree well again (3.5% difference). Despite its many different constituents the behavior of honey is described well by a Newtonian fluid.[33] Viscosities of less than 100 Pas are currently assumed (see Sec. II A) in the modeling of the quasi-Scholte mode. Errors will be introduced in the analysis for highly viscous liquids. Honey is still modeled well by assuming low viscosity; however, high-viscosity formulations for Eqs. (4)–(7) can be implemented[27] for more viscous fluids. High-viscosity fluids will also exhibit significant shear leakage that can be partially limited by altering measurement frequency and propagation length. Highly viscous liquids will also affect the quasi-Scholte mode group velocity, which complicates the extraction of bulk velocity and attenuation.

3. Suspension

Further measurements were performed on water and a suspension of silicon dioxide (1% volume fraction, mean diameter 1–5 μm) in water. The particles were suspended by stirring and by the use of an ultrasonic bath before the measurement. A thinner stainless-steel plate ($\rho=7467$ kg/m^3, $c_l=6000$ m/s, $c_s=2700$ m/s) was immersed in the fluid. A higher frequency shear transducer (Panametrics 5 MHz) was coupled to the plate as in the previous setup and the same procedure was followed. Figure 11 shows the measured group velocity and attenuation of the quasi-Scholte mode.

Water is modeled as a viscous fluid with viscosity of 1 mPas, density of 1000 kg/m^3, and longitudinal bulk velocity $c_l=1500$ m/s. The water results agree well with the predicted curves, and also the quasi-Scholte mode group velocity of the silicon dioxide suspension agrees well with group velocity predictions for water. The bulk velocity and viscosity of the suspension are effectively the same as that of water because of the low volume fraction (1%) of the suspended particles. A slight decrease in velocity of the order of 1% is noticeable in the velocity measurement in Fig. 11(a). However, the quasi-Scholte mode attenuation in the suspension is far higher than expected for its viscosity. This increase in attenuation in the suspension compared to water suggests that the quasi-Scholte mode is sensitive to nonviscous longitudinal bulk attenuation mechanisms in the fluid (scattering from particles in this case). This is not the case in the conventional guided-wave techniques,[9] which only measure shear velocity in the fluid. Only this qualitative result is presented here; however, future work will be undertaken to quantitatively extract the longitudinal bulk attenuation from suspensions and emulsions.

A combination of the quasi-Scholte mode with a torsional (or $SH0$ wave) dipstick and a densitometer[8] will enable the complete characterization of a fluid where shear velocity, longitudinal velocity, longitudinal attenuation, and density are measured as independent properties of the fluid. (Shear attenuation is not determined; however, it is unimportant considering the short penetration depth of shear waves in liquids.)

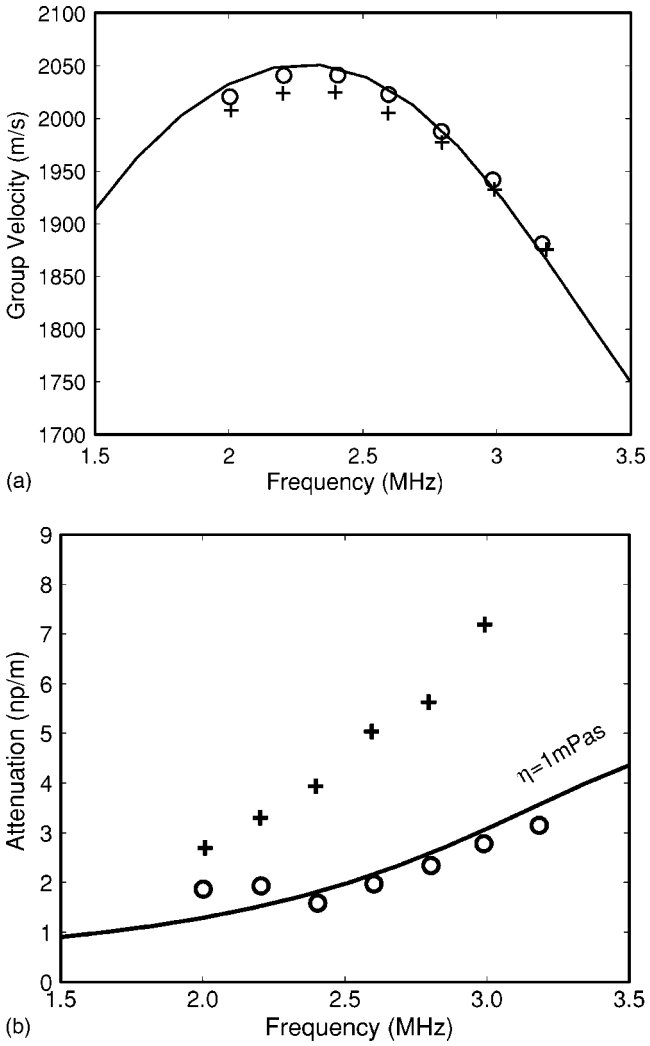


FIG. 11. Measured and theoretically predicted (——) (a) group velocity and (b) attenuation of the quasi-Scholte mode on a 0.11-mm-thick stainless-steel plate immersed in water (○) and a silicon dioxide suspension (+).

V. CONCLUSIONS

A new method to extract fluid properties using the quasi-Scholte mode has been investigated. Experiments were carried out to test the method on glycerol and honey. Good results were obtained and validated by a conventional measurement technique. The measurement of glycerol properties using the quasi-Scholte mode and the validation method agreed to within 4% for the viscosity, and the longitudinal bulk wave velocities agreed to well within 1%, which is of the same order of magnitude as the measurement uncertainties. The measured values also compared well to literature data. The quasi-Scholte mode method, in contrast to other guided-wave techniques, is sensitive to attenuation mechanisms that are not related to viscosity. This is illustrated by qualitative measurements on water and a suspension of silica particles in water. Future work will address problems encountered with very viscous fluids, the extension of the

method to a wider frequency range and the change of geometry from a plate to a narrow dipstick. The quantitative extraction of longitudinal attenuation from fluids with nonviscous attenuation mechanisms such as suspensions and emulsions will also be considered. This opens the possibility of measuring particle size distribution. The method has many practical advantages and can be developed into a dipstick that completely characterizes fluids.

APPENDIX: DETAILS OF THE GLOBAL MATRIX

The components of matrix A in Eq. (8) are given here. Shown as $A=[B\vdots C]$, where

$$B=\begin{bmatrix} k & -k_{2Tf} & k & kg_{SL} \\ -k_{2Lf} & -k & k_{2Ls} & -k_{2Ls}g_{SL} \\ i\rho_f B_f & 2i\rho_f\beta_f^2 k k_{2Tf} & i\rho_s B_s & i\rho_s B_s g_{SL} \\ -2i\rho_f\beta_f^2 k k_{2Lf} & i\rho_f B_f & 2i\rho_s\beta_s^2 k k_{2Ls} & -2i\rho_s\beta_s^2 k k_{2Ls} g_{SL} \\ 0 & 0 & kg_{SL} & k \\ 0 & 0 & k_{2Ls}g_{SL} & -k_{2Ls} \\ 0 & 0 & i\rho_s B_s g_{SL} & i\rho_s B_s \\ 0 & 0 & 2i\rho_s\beta_s^2 k k_{2Ls} g_{SL} & -2i\rho_s\beta_s^2 k k_{2Ls} \end{bmatrix} \tag{A1}$$

$$C=\begin{bmatrix} k_{2Ts} & -k_{2Ts}g_{ST} & 0 & 0 \\ -k & -kg_{ST} & 0 & 0 \\ -2i\rho_s\beta_s^2 k k_{2Ts} & 2i\rho_s\beta_s^2 k k_{2Ts} g_{ST} & 0 & 0 \\ i\rho_s B_s & i\rho_s B_s g_{ST} & 0 & 0 \\ k_{2Ts}g_{SS} & -k_{2Ts} & k & k_{2Tf} \\ -kg_{ST} & -k & k_{2Lf} & -k \\ -2i\rho_s\beta_s^2 k k_{2Ts} g_{ST} & 2i\rho_s\beta_s^2 k k_{2Ts} & i\rho_f B_f & -2i\rho_f\beta_f^2 k k_{2Tf} \\ i\rho_s B_s g_{ST} & i\rho_s B_s & 2i\rho_f\beta_f^2 k k_{2Lf} & i\rho_f B_f \end{bmatrix} \tag{A2}$$

where k is the guided-wave wave number to be solved for, ρ_s is the solid plate material density, ρ_f is the fluid density, and the remaining quantities are defined by the following equations:

$$k_{2Lf,s}=\pm\left(\frac{\omega^2}{\alpha_{f,s}^2}-k^2\right)^{1/2}, \tag{A3}$$

$$k_{2Tf,s}=\pm\left(\frac{\omega^2}{\beta_{f,s}^2}-k^2\right)^{1/2}, \tag{A4}$$

$$B_{f,s}=\omega^2-2\beta_{f,s}^2k^2, \tag{A5}$$

$$G_{SL}=e^{i(k_{2Ls}h)}, \tag{A6}$$

$$G_{ST}=e^{i(k_{2Ts}h)}, \tag{A7}$$

where $k_{2Lf,s}$ is the component of the longitudinal bulk wave wave number in the direction normal to the plate surface, $k_{2Tf,s}$ is the component of the shear bulk wave wave number in the direction normal to the plate surface, $B_{f,s}$, G_{SL}, and G_{ST} are multiplication factors; h stands for the plate thickness. α and β are the complex bulk velocities [see Eqs. (2) and (3)] of the medium. The media are indicated by the subscripts f and s, which also stand for fluid and solid in the subscripts of other quantities.

Equations (A3) and (A4) contain a square-root expression; this results in several possible solutions, some of which may be unphysical (e.g., waves with negative attenuation). Therefore, care should be taken to choose the right sign of the expression under the square root before entering Eqs. (A3) and (A4) into matrix A.

[1] E. P. Kora, E. Latrille, and I. Souchon, "Texture–flavor interactions in low fat stirred yogurt: How mechanical treatment, thickener concentration and aroma concentration affect perceived texture and flavor," J. Sensory Studies **18**(5), 367–390 (2003).

[2] A. K. Holmes and R. E. Challis, "The applicability of acoustic wave propagation models to silica sols and gels," J. Colloid Interface Sci. **216**(1), 50–58 (1999).

[3] *Nondestructive Testing Handbook*, edited by A. Vary (A.S.N.T., Columbus, OH, 1991), Vol. 7.

[4] J. Krautkramer and H. Krautkramer, *Ultrasonic Testing of Materials* (Springer, Berlin, 1983).

[5] M. J. W. Povey, *Ultrasonic Techniques for Fluids Characterization* (Academic, New York, 1997).

[6] A. N. Kalashnikov and R. E. Challis, "Errors in the measurement of ultrasonic absorption for materials evaluation," in *Review of Quantitative Nondestructive Evaluation* (American Institute of Physics, Melville, NY, 2002), Vol. 21B, pp. 1997–2004.

[7] L. C. Lynnworth, "New designs for magnetostrictive probes using extensional, torsional and flexural waves," in Ultrasonics Symposium Proceedings, Vol. IEEE No. 78CH1344-1SU, 1978.

[8] J. O. Kim and H. H. Bau, "On line, real-time densimeter—Theory and optimization," J. Acoust. Soc. Am. **85**, 432–439 (1989).

[9] T. Vogt, ''Determination of Material Properties Using Guided Waves,'' Ph.D. thesis, Imperial College London, 2002.
[10] T. Vogt, M. J. S. Lowe, and P. Cawley, ''Ultrasonic waveguide techniques for the measurement of material properties,'' in *Review of Quantitative Nondestructive Evaluation* (American Institute of Physics, Melville, NY, 2002), Vol. 21, pp. 1–8.
[11] V. G. Mozhaev and M. Weihnacht, ''Subsonic leaky Rayleigh waves at liquid–solid interfaces,'' Ultrasonics **40**, 927–933 (2002).
[12] T. Vogt, M. J. S. Lowe, and P. Cawley, ''Cure monitoring using ultrasonic guided waves in wires,'' J. Acoust. Soc. Am. **114**(3), 1303–1313 (2003).
[13] A. Caiti, T. Akal, and R. D. Stoll, ''Estimation of shear wave velocity in shallow marine sediments,'' IEEE J. Ocean. Eng. **19**(1), 58–72 (1994).
[14] J. L. Rose, *Ultrasonic Waves in Solid Media* (Cambridge University Press, Cambridge, 1999).
[15] C. Glorieux, K. V. Rostyne, K. Nelson, W. Gao, W. Lauriks, and J. Thoen, ''On the character of acoustic waves at the interface between hard and soft solids and liquids,'' J. Acoust. Soc. Am. **110**(3), 1299–1306 (2001).
[16] J. G. Scholte, ''On the Stoneley wave equation,'' in *Geophysics* (North-Holland, Amsterdam, 1941), Vol. XLV, pp. 20–25.
[17] M. J. S. Lowe, ''Matrix techniques for modelling ultrasonic waves in multilayered media,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**(4), 525–542 (1995).
[18] M. M. Vol'kenshtein and V. M. Levin, ''Structure of a Stoneley wave at an interface between a viscosus fluid and a solid,'' Sov. Phys. Acoust. **34**(4), 351–355 (1987).
[19] N. Favretto-Anres and G. Rabau, ''Excitation of the Stoneley–Scholte wave at the boundary between an ideal fluid and a viscoelastic solid,'' J. Sound Vib. **203**(2), 193–208 (1997).
[20] G. D. Meegan, M. F. Hamilton, Y. A. II'nskii, and E. A. Zabolotskaya, ''Nonlinear Stoneley and Scholte waves,'' J. Acoust. Soc. Am. **106**(4), 1712–1723 (1999).
[21] S. Nasr, J. Duclos, and M. Leduc, ''PVDF transducers generating Scholte waves,'' Electron. Lett. **24**(6), 309–311 (1988).
[22] J. McLean and F. L. Degertekin, ''Directional scholte wave generation and detection using interdigital capacitive micomachined ultrasonic transducers,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **51**, 756–764 (2004).
[23] C. Desmet, V. Gusev, W. Lauriks, C. Glorieux, and J. Thoen, ''Laser-induced thermoelastic excitation of Scholte wave,'' Appl. Phys. Lett. **68**(21), 2939–2941 (1996).
[24] M. De Billy and G. Quentin, ''Experimental study of the Scholte wave propagation on a plane surface partially immersed in a liquid,'' J. Appl. Phys. **54**(8), 4314–4322 (1983).
[25] T. J. Matula and P. L. Marston, ''Energy branching of a subsonic flexural wave on a plate at an air–water interface. Observation of the wave field near the interface and near the plate,'' J. Acoust. Soc. Am. **97**(3), 1389–1398 (1995).
[26] P. B. Nagy and R. M. Kent, ''Ultrasonic assessment of Poisson's ratio in thin rods,'' J. Acoust. Soc. Am. **98**(5), 2694–2701 (1995).
[27] M. J. S. Lowe and B. N. Pavlakovic, DISPERSE User Manual, version 2.0.11d. Technical report, Imperial College of Science, Technology and Medicine, London, UK; ⟨www.ndt.imperial.ac.uk, 2001⟩.
[28] A. H. Nayfeh and P. B. Nagy, ''Excess attenuation of leaky Lamb waves due to viscous fluid loading,'' J. Acoust. Soc. Am. **101**(5), 2649–2658 (1997).
[29] M. J. S. Lowe, ''Plate waves for the NDT of diffusion bonded titanium,'' Ph.D. thesis, Imperial College, 1992.
[30] P. B. Nagy and A. H. Nayfeh, ''Viscosity-induced attenuation of longitudinal guided waves in fluid-loaded rods,'' J. Acoust. Soc. Am. **100**(3), 1501–1508 (1996).
[31] G. W. C. Kaye and T. H. Laby, *Tables of Physical and Chemical Constants*, 16th ed. (Longman, Harlow, New York, 1995).
[32] *Handbook of Chemistry and Physics*, 70th ed., edited by D. R. Lide (CRC Press, Boca Raton, 1989).
[33] A. Kulmyrzaev and D. J. McClements, ''High frequency dynamic shear rheology of honey,'' J. Food. Eng. **45**, 219–224 (2000).

Finite elements methods for modeling the guided waves propagation in structures with weak interfaces

Bernard Hosten[a] and Michel Castaings
Laboratoire de Mécanique Physique, Bordeaux 1 University, UMR C.N.R.S. 5469, 351 cours de la Libération, 33405 Talence Cedex, France

(Received 3 September 2004; revised 2 November 2004; accepted 5 November 2004)

This paper describes two methods using a finite element (FE) code for modeling the effects of weak interfaces on the propagation of low-order Lamb modes. The variable properties of the interfaces are modeled by either a thin layer or a uniform repartition of compression and shear springs that insure the continuity of the stresses and impose a discontinuity in the displacement field. The method is tested by comparison with measurements that were presented in a previous paper [J. Acoust. Soc. Am. **113**(6) 3161–3170 (2003)]. The interface was the contact between a rough elastomer with high internal damping loaded against one surface of a glass plate. Both normal and shear stiffnesses of the interface were quantified from the attenuation of A_0 and S_0 Lamb waves caused by leakage of energy from the plate into the elastomer and measured at each step of a compressive loading. The FE model is made in the frequency domain, thus allowing the viscoelastic properties of the elastomer to be modeled by using complex moduli as input data. By introducing the interface stiffnesses in the code, the predicted guided waves attenuations are compared to the experimental results to validate the numerical FE methods.
[DOI: 10.1121/1.1841731]

PACS numbers: 43.35.Mr, 43.35.Cg, 43.35.Zc [YHB] Pages: 1108–1113

I. INTRODUCTION

The nondestructive evaluation and testing of interface quality is a key issue for the development of stratified structures in various domains, for instance, the aeronautics or car industry. Due to the complex geometry of some structures, numerical methods must often be used instead of closed-form solutions, to model the propagation of guided waves and their diffraction by obstacles, inhomogeneities, or defects.

In solid stratified media, it is generally assumed that the displacements and stresses fields are continuous through interfaces, except for weak interfaces due to various conditions of bad adhesion. The adhesion is said to be cohesive or adhesive depending on the area where the weakness happens. For instance, in glued bonds,[1] the cohesion is realized by a layer of glue that can be included in the finite element model with its viscoelastic properties. Then, the weakness can be represented by the variations of these properties due to various causes such as the level of curing, the porosity, etc. However, the weakness can happen at the very thin interfaces, for instance between the glue and the adherents depending on the surface's preparation (oil contamination, roughness, etc.) or in diffusion bonds.[2] Whatever the nature of the weakness (slip, kissing or partial bonds,[3] rough interfaces[4]) the interface can be represented as a very thin layer. In this case, with a finite element, models require very large numbers of elements and exaggerated computation times. In some cases, the intermediate thin layer and two interface layers must be taken into account[5] and should be included in a numerical simulation.

This paper presents two approaches for modeling the weakness of an interface with a commercially available FE software:[6] (1) the thin layer model[7] with an adequate thickness and (2) boundary equations that take into account the displacement discontinuities at interfaces. The term ''adequate'' is specified with a semianalytic computation that gives the largest thickness of the thin layer compatible with the wavelengths of the modes propagating in the surrounding media. The boundary equations method consists of considering a surface repartition of springs between the assembled media which are meshed by finite elements. This technique was recently applied to composites, taking into account the existence of imperfect interfaces for predicting the macroscopic response.[8]

Both models are validated by comparison with experimental results on the propagation of guided waves in the presence of an interface with variable stiffness.[9]

II. THIN LAYER MODEL

The use of a thin layer for modeling the effects of weak interface on the propagation of waves in stratified materials was introduced in 1980 by Schoenberg.[7] This model is recalled here with some simplifications by introducing a fictitious material.

In a plane of symmetry of an anisotropic material, with the plane strain hypothesis ($u_3=0$), the displacement field $\mathbf{U}=\{u_1,u_2\}$ is linked to the stress tensor by

$$\sigma_1=C_{11}u_{1,1}+C_{12}u_{2,2}, \quad \sigma_2=C_{12}u_{1,1}+C_{22}u_{2,2},$$

$$\sigma_6=C_{66}(u_{1,2}+u_{2,1}), \tag{1}$$

[a]Electronic mail: b.hosten@lmp.u-bordeaux1.fr

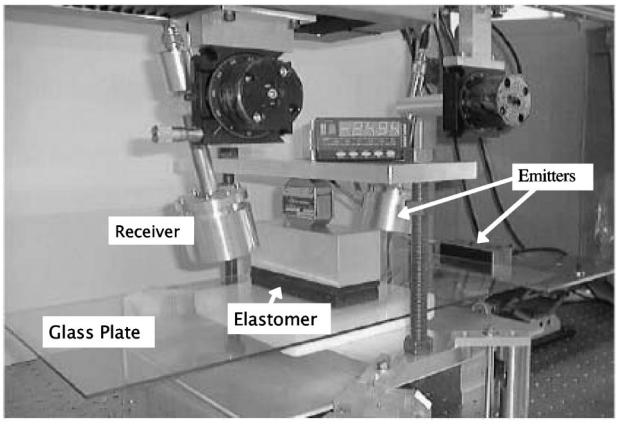


FIG. 1. Experimental arrangement to launch and receive Lamb modes in a glass plate and to measure the transmission coefficient when an elastomer is pressed against the plate.

where $C_{ij}=C'_{ij}+iC''_{ij}$, C'_{ij} and C''_{ij} represent the elastic and viscoelastic moduli, respectively, and $u_{i,j}$ are the partial derivatives along the directions j.

A fictitious material layer can be used to simulate a material that represents two springs, one perpendicular to the interface with stiffness k_L and another along the interface with stiffness k_T. In a stratified structure, with direction 1 being normal to interfaces, the stress σ_2 is not continuous through the interfaces. Therefore, its value has no importance in the fictitious material layer and can be set to zero by imposing $C_{12}=0$ and $C_{22}=0$. If the thickness of this layer is small enough, the gradients of displacements u_1 and u_2 in direction 1 can be approximated by

$$u_{1,1}=\Delta u_1/h \quad \text{and} \quad u_{2,1}=\Delta u_2/h, \tag{2}$$

where Δu_1 is the normal displacement jump between the two surfaces of the spring layer. Then, the relation between the stress normal to the interface and the displacement jump defines the stiffness k_L

$$\sigma_1=\Delta u_1 C_{11}/h=k_L\Delta u_1. \tag{3}$$

Since the thickness of the fictitious material layer is small, its density has no effect on the propagation and can be set to zero. Then, the two equilibrium equations become

$$\sigma_{1,1}+\sigma_{6,2}=0 \quad \text{and} \quad \sigma_{6,1}=0. \tag{4}$$

For a plane wave propagating along the plate, all the fields depend on the direction 2 through the term exp $\times(-2\pi i x_2/\lambda)$, where λ is the wavelength of the guided wave. Then, similarly to the normal stress, the shear stress is linked to the tangential displacement jump Δu_2 by

$$\sigma_6=C_{66}(-2\pi i u_1/\lambda+\Delta u_2/h). \tag{5}$$

To define the equivalent ''shear spring'' by the equation

$$\sigma_6=\Delta u_2 C_{66}/h=k_T\Delta u_2, \tag{6}$$

the term $2\pi u_1/\lambda$ must be small in comparison to the term $\Delta u_2/h$. With this condition, it can be shown that the term $\sigma_{6,2}$ in Eq. (4) is also small and the model achieves the conditions of constant stresses through the layer. A necessary

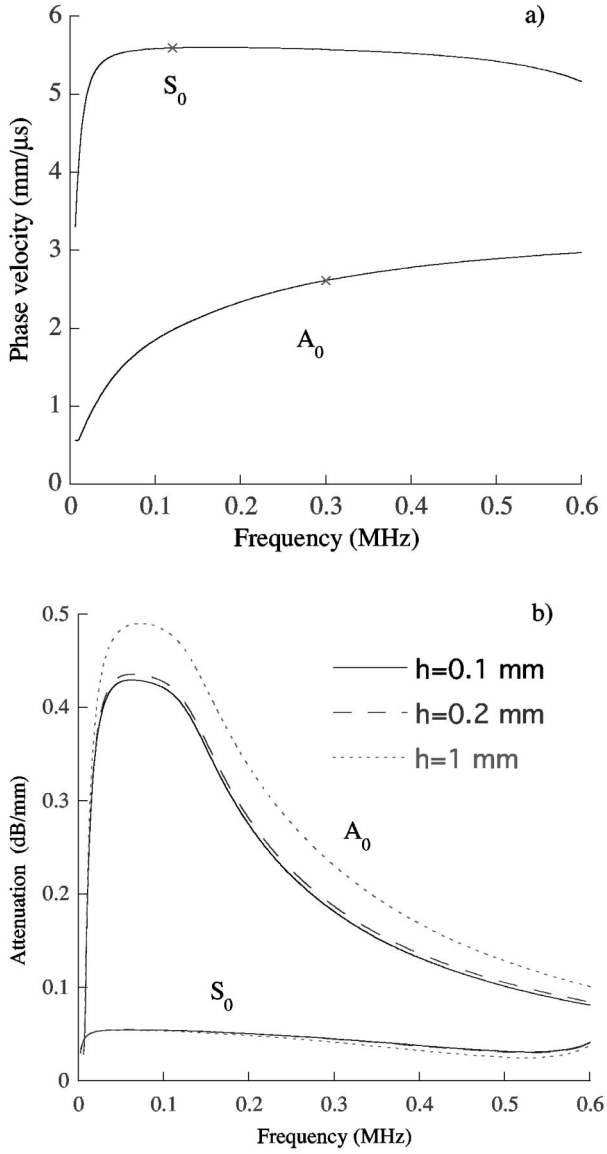


FIG. 2. Dispersion curves for the structure glass/interface/elastomer ($k_L=4.30$ and $k_T=3.30$ GPa/mm). (a) Phase velocity; (b) Attenuation of the guided modes in glass due to the leak into the elastomer.

condition is that the thickness is smaller than the wavelength of the guided mode

$$h \ll \lambda/2\pi. \tag{7}$$

This condition depends on the wavelength of the mode and on the relative values of u_1 and u_2. With a semianalytic model it is easy to use a very thin layer to satisfy this condition. However, the use of FE software imposes numerous very small elements in the thin layer, thus increasing the computation time excessively.

III. SIMULATION OF EXPERIMENTAL SETUP

To use the experimental results of a real situation as a reference, the FE method is applied to simulate the situation shown in Fig. 1 and previously described.[9] It is shortly described in this section.

The measurements were performed in a structure made of a 26.4-mm-thick highly attenuative elastomer block mea-

TABLE I. Variation of interface stiffnesses (from Ref. 9).

Pressure (MPa)	0.00	0.01	0.03	0.05	0.10	0.16	0.21	0.26	0.31	.36	0.42
k_L (GPa/mm)	0	1.10	1.75	2.28	2.80	3.15	3.50	3.80	4.00	4.10	4.30
k_T (GPa/mm)	0	0.19	0.29	0.39	0.65	0.90	1.00	1.15	1.50	2.00	3.30

suring 160 by 60 mm that was compressively loaded against a glass plate, measuring $3.9\times300\times600$ mm. The elastomer block was loaded through a thick (60-mm) aluminum block. This acted as an effectively rigid block, and so the load was applied evenly across the top surface of the elastomer. The glass plate was supported by a 40-mm-thick block of rigid foam. This foam was very porous and had a very low acoustic impedance, and so simulated air backing of the glass plate.

The transmission coefficient was then calculated by dividing the spectra of the transmitted pulse by a reference measurement taken when no elastomer was present. In this way measured transmission coefficient reductions are due either to the leakage of energy into the surrounding media or to reflections back from the edges of the elastomer.

Based on the k_L and k_T sensitivity study results,[9] measurement positions were chosen centered on 120 kHz for S_0 and 300 kHz for A_0 [Fig. 2(a)]. The results of the best estimation of k_L and k_T are shown in Table I.

IV. SEMIANALYTIC COMPUTATION OF DISPERSION CURVES WITH WEAK INTERFACES

The elastic properties of the glass plate and the viscoelastic properties of the elastomer are given in Table II. From these properties, the complex dispersion curves of the glass plate connected to a semi-infinite elastomer half-space can be computed for the two lower order modes A_0 and S_0, with a semianalytic model based on the global transfer matrix method[10] or the stiffness matrix method.[11] Here, it is computed with the surface impedance matrices method.[12] It was found that the phase velocities of these two modes are not sensitive to the interface thickness, h, as long as this is smaller than 1 mm. The attenuation was found not sensitive to this thickness below a threshold of around $h=0.2$ mm. Figure 2(b) shows significant differences of the attenuation when the computations are made with a thickness ($h=1$ mm) greater than this threshold. This is an easy and quick test to estimate the upper limit of the thickness interface for satisfying the condition given at Eq. (7). In Fig. 2, the values of k_L and k_T are the largest used in this work. Then, for any other values, the approximation will be even better.

For this computation, the layer thickness was set to $h=10\,\mu$m, which imposes the values of $C11=0.043$ GPa and $C66=0.033$ GPa. $C12$ and $C22$ were set to zero. If we assume that the interface material is isotropic, the values of $C22$ and $C12$ should be, respectively, 0.043 and -0.023 GPa and lead to a negative Young's modulus. These values indicate that the interface is not really composed of a ''classical'' material, and the term ''fictitious'' is quite appropriate. The same computation with various combinations of $C12$ and $C22$ values leads to the same dispersion curves, showing that they have no influence as soon as the thickness is small enough. In the case of a real layer made of glue, for instance, we should use values representative of the material properties.

V. FINITE ELEMENT SIMULATION

A. Wave propagation in viscoelastic materials

The propagation of plane waves in a plate made of an orthotropic viscoelastic material of density ρ is considered, when the plane of propagation is a plane of symmetry, where directions 1 and 2 (Fig. 3) represent the axes of symmetry. With the plane strain conditions ($\varepsilon_{13}=\varepsilon_{23}=\varepsilon_{33}=0$), the two-dimensional general equation is written as follows in the frequency domain:

$$\begin{aligned} &C_{11}u_{1,11}+C_{66}u_{1,22}+(C_{12}+C_{66})u_{2,12}=-\rho\omega^2u_1,\\ &C_{22}u_{2,22}+C_{66}u_{2,11}+(C_{21}+C_{66})u_{1,12}=-\rho\omega^2u_2, \end{aligned} \tag{8}$$

where $\mathbf{U}=\{u_1,u_2\}$ is the Fourier transform of the displacement vector. One and 2 define the plane of propagation with 1 normal to the plate and 2 the direction of propagation. The above differential equations must be written in the following FEMLAB form:[6]

$$\nabla\cdot(c\nabla u)-au=0, \tag{9}$$

where c is a 2×2 matrix composed of four submatrices, such that

$$\begin{aligned} &c_{11}=\begin{pmatrix} C_{11} & 0\\ 0 & C_{66}\end{pmatrix},\quad c_{12}=\begin{pmatrix} 0 & C_{12}\\ C_{66} & 0\end{pmatrix},\\ &c_{21}=\begin{pmatrix} 0 & C_{66}\\ C_{12} & 0\end{pmatrix},\quad c_{22}=\begin{pmatrix} C_{66} & 0\\ 0 & C_{22}\end{pmatrix}, \end{aligned} \tag{10}$$

and a is a 2×2 matrix given by

$$a=\begin{pmatrix} -\rho\omega^2 & 0\\ 0 & -\rho\omega^2\end{pmatrix}. \tag{11}$$

TABLE II. Properties of the materials.

Material	Thickness (mm)	Density (g/cm^3)	C_{11} (GPa)	C_{66} (GPa)	C_{22} (GPa)	C_{12} (GPa)
Glass	3.9	2.5	85	30	85	25
Elastomer	∞	1.25	4.8+I 0.4	0.4+I 0.04	4.8+I 0.4	4.0+I 0.32
Interface layer	h	0	k_L^*h	k_T^*h	0	0

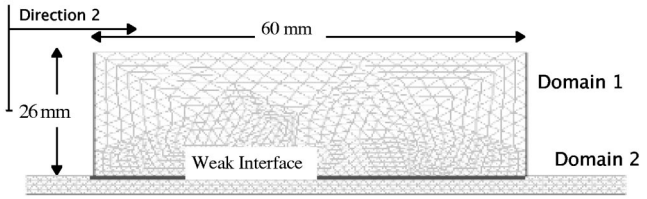


FIG. 3. Meshing around the weak interface between the glass plate and the polymer.

In the Fourier domain, the complex viscoelastic properties of the elastomer given in Table II are easily introduced as well as fictitious complex viscoelastic properties of both ends of the glass plate to realize absorbers.

For simulating the propagation of waves in an unbounded domain, "perfectly matched layers" are commonly used.[13] Another solution is to add a zone to the studied zone, in which the real parts of the C_{ij} are the same and their imaginary parts are progressively enlarged.[14] Unfortunately, this solution increases the number of elements at the boundary between the studied zone and the absorber. Here, the solution is even simpler since no external zone is added; the material by itself is made absorbing at its extremities. The imaginary parts of the moduli in the glass plate, the length of which is comprised between $-La$ and L, are given by the equations

$$
\begin{aligned}
&0>x_2<L-La \quad C''_{ij}=0,\\
&x_2<0 \quad C''_{ij}=A\left(\frac{-x_2}{La}\right)^3 C'_{ij},\\
&x_2>L-La \quad C''_{ij}=A\left(\frac{x_2-(L-La)}{La}\right)^3 C'_{ij}.
\end{aligned}
\tag{12}
$$

The length of the absorber La and the factor A can be adjusted to minimize the reflections at the edges of the plate.

B. Displacement discontinuities for simulating a weak interface

The FE software permits inclusion of various physics models for connecting them together in a "multiphysics" mode.[6] Here, two identical wave propagation models are applied, one to the polymer region with displacements u_1^1, u_2^1 and the other to the glass plate with displacements u_1^2, u_2^2 (Fig. 3). The connection is done by the following boundary conditions at the interface:

$$
\begin{aligned}
&\begin{bmatrix}\sigma_1^1\\ \sigma_6^1\end{bmatrix}=\begin{bmatrix}k_L(u_1^2-u_1^1)\\ k_T(u_2^2-u_2^1)\end{bmatrix} \quad (\text{domain } 1) \quad \text{and}\\
&\begin{bmatrix}\sigma_1^2\\ \sigma_6^2\end{bmatrix}=\begin{bmatrix}k_L(u_1^1-u_1^2)\\ k_T(u_2^1-u_2^2)\end{bmatrix} \quad (\text{domain } 2).
\end{aligned}
\tag{13}
$$

In the case of a fluid–solid interface, k_T is null as well as the shear stress.

C. Simulation of the A_0 and S_0 emissions

For the experimental generation detection of A_0 at 300 kHz, air-coupled capacitive transducers were used (Fig. 1). In the FE method, the action of the transmitter is simulated by locally imposing a stress corresponding to that applied by the transducer inclined at 8 deg., with respect to the normal of the plate.

As S_0 is made up of mostly in-plane displacement, a piezoelectric contact transducer (active element size was 100 mm long by 40 mm wide) was placed against the edge of the plate (Fig. 1). In the FE method, this is simply simulated by imposing a uniform displacement normal to an edge of the plate.

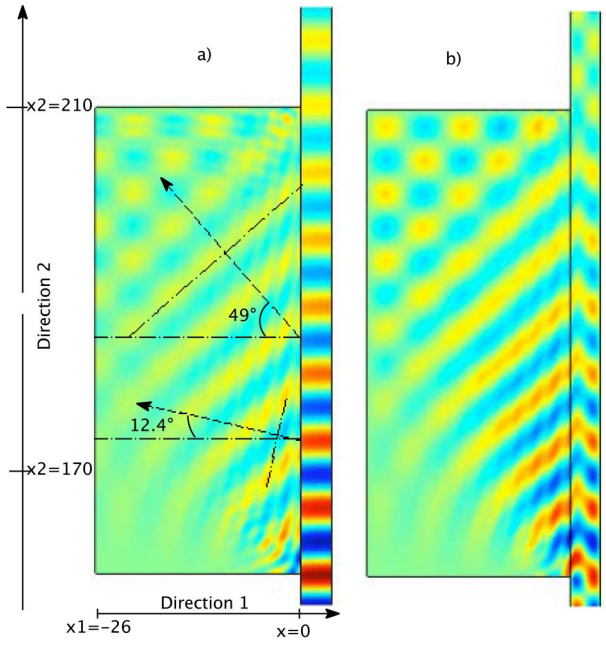


FIG. 4. Snapshots of the real part of (a) the normal displacements (direction 1) u_1^1 and u_1^2 and (b) the stress fields σ_1^1 and σ_1^2, for A_0-300-kHz mode and $k_L=4.3$, $k_T=3.3$ GPa/mm.

D. Results for A_0 mode simulation

Figures 4 and 5 present some results of the A_0 mode simulation, with values of the interface stiffnesses $k_L=4.3$ and $k_T=3.3$ GPa/mm. The two domains, glass plate and elastomer, are represented by 19 776 triangular elements corresponding to 82 516 degrees of freedom. The computation time is 13 s with a Macintosh (G5 biprocessor 2 MHz). The snapshots clearly show the discontinuity in the displacements u_1^1 and u_1^2 normal to the plate and the continuity of the stresses σ_1^1 and σ_1^2. The attenuation in the elastomer is not enough to suppress the diffraction features in the corner at $x_2=210$ mm. Figure 4 shows the real parts of the fields and that the main wavefronts in the elastomer are oriented at around 49 deg. from the direction 1, normal to the interface. According to Snell's laws, the elastomer longitudinal bulk wave velocity (1.96 mm/μS) and the velocity of the mode A_0 in the glass plate (2.61 mm/μS), this angle corresponds to the longitudinal bulk mode generation in the elastomer. Another array of small wavefronts can be seen with an angle of 12.6 deg. corresponding to the shear bulk wave velocity (0.57 mm/μS).

The boundary equations (13) are now replaced by a thin layer made of fictitious material, as described in Sec. II. The

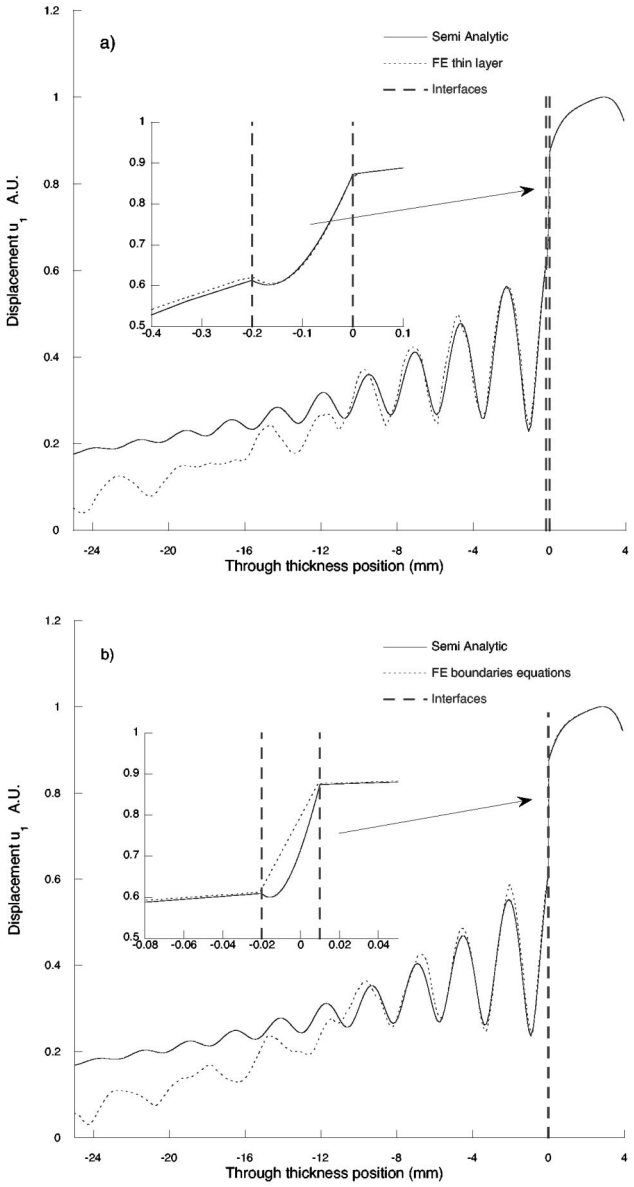


FIG. 5. Modulus of the amplitude of through-thickness displacement field at position $x_2=170$ mm. Comparison between semianalytic and FE computations $k_L=4.3$, $k_T=3.3$ GPa/mm. (a) Thin layer method; (b) Boundary equations method.

thickness of this layer is chosen equal to 0.2 mm, as this limit value was defined in Sec. IV. This geometry, and also data given in Table II, were used for FE simulation. This relatively large thickness does not lead to an excessive number of elements: 19 514 and 80 294 degrees of freedom. For this particular example, the computation time was similar for both methods.

Figure 5 presents the comparison between the displacements computed with the semianalytic and FE methods, plotted versus the through-thickness position. In Fig. 5(a), the interface layer thickness is fixed at 200 μm both for the semianalytic model and the FE thin layer method. In Fig. 5(b), the interface layer thickness considered in the semianalytic model is made equal to the size of the elements meshing the boundary, i.e., 30 μm. The zooms show that the three

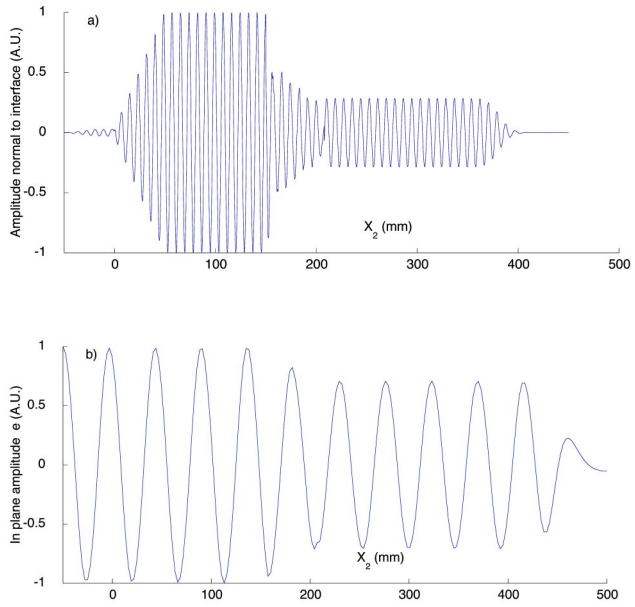


FIG. 6. Real part of the displacement at the surface of the glass plate. (a) Out-of-plane displacements of mode A_0-300 kHz; (b) In-plane displacements of mode S_0-120 kHz.

methods give the same displacement discontinuity and agree very well.

The amplitude in the elastomer is decreasing from the interface due to the viscoelastic attenuation and presents some oscillations. This is not due to reflections from the elastomer borders, since the semianalytic computation simulates a semi-infinite elastomer and also presents the same oscillations. This is due to interactions with the shear bulk mode that is also generated. Using the semianalytic model, if the value of the imaginary part of the elastomer Coulomb modulus (C_{66}) is progressively increased, these oscillations disappear.

E. Transmission coefficients

The displacement fields at the surface of the glass plate for A_0 and S_0 modes are shown in Fig. 6, where the various zones can be seen. For the A_0-300-kHz mode [Fig. 6(a)], the absorption zones are $-50<x_2<0$ and $x_2>350$ mm, the excitation zone is $0<x_2<50$ mm, and the leakage-in-elastomer zone is $150<x_2<210$ mm. For the S_0-120-kHz mode [Fig. 6(b)], the left absorption zone is suppressed since the excitation is made by imposing a unit amplitude displacement at the left edge and the length La of the right-hand-side absorber is increased to optimize the efficiency at this lower frequency. In both cases, the transmission coefficients are easily defined by the ratio of the amplitudes after and before the leakage zone.

Finally, Fig. 7 shows the comparison between the transmission coefficients of A_0 and S_0 modes measured experimentally and computed with the FE methods in the range of interface stiffness values (Table I). The small differences can be attributed to the measurement errors or to the estimation errors of stiffness values or to the limitation of the 2D simulation. In any case, they are very small.

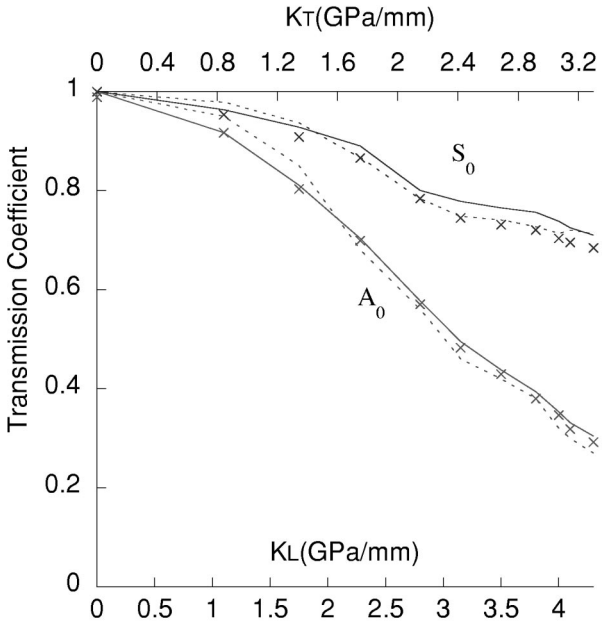


FIG. 7. Comparison between experimental (x) and FE predicted transmission coefficients. Dashed lines: thin layer method. Solid lines: boundary equations method.

VI. CONCLUDING REMARKS

Two FE methods has been successfully used for simulating the propagation of guided modes in a glass plate and their attenuation by a viscoelastic elastomer block loaded against the plate. They were tested for taking into account the variable interface stiffness between a glass plate and a rough elastomer block.

The thin layer method can be used for modeling the perfect/weak interface if the layer thickness is smaller than a threshold that depends on the wavelengths of the modes propagating in the surrounding media.

The boundary equations method, introducing displacement discontinuities between two domains, gives similar results. Moreover, this method does not require adjusting a parameter like the layer thickness in the previous method, and it is more valuable since it can be used for simulating boundary conditions between a fluid and a solid.

These two methods have been used for simulating a specific application with the motivation of experimental verifications. However, they could be applied to any situation involving weak interfaces and bulk or guided modes propagation testing.

[1] R. Seifried, L. Jacobs, and J. Qu, ''Propagation of guided waves in adhesive bonded components,'' NDT & E Int. **37**, 317–328 (2002).
[2] A. Lavrentyev and J. Beals, ''Ultrasonic measurement of the diffusion bond strength,'' Ultrasonics **38**, 513–516 (2000).
[3] P. Nagy, ''Ultrasonic classification of imperfect interfaces,'' J. Nondestruct. Eval. **113/4**, 127–140 (1992).
[4] A. Baltazar, S. Rokhlin, and C. Pecorari, ''On the relationship between ultrasonic and micromechanical properties of contacting rough surfaces,'' J. Mech. Phys. Solids **50**, 1397–141 (2002).
[5] A. Baltazar, L. Wang, B. Xie, and S. Rokhlin, ''Inverse ultrasonic determination of imperfect interfaces and bulk properties of a layer between two solids,'' J. Acoust. Soc. Am. **114**(3), 1424–1434 (2003).
[6] FEMLAB, User's Guide and Introduction. Version 3.0 by—COMSOL AB 2004 ⟨http://www.comsol.com/⟩. Website last viewed on 2 November 2004.
[7] M. Schoenberg, ''Elastic wave behavior across linear slip interfaces,'' J. Acoust. Soc. Am. **68**(5), 1516–1521 (1980).
[8] L. Benabou, M. Nait Abdelaziz, and N. Benseddiq, ''Effective properties of a composite with imperfectly bonded interface,'' Theor. Appl. Fract. Mech. **41**, 15–20 (2004).
[9] B. Drinkwater, M. Castaings, and B. Hosten, ''The measurement of *A*0 and *S*0 Lamb wave attenuation to determine the normal and shear stiffnesses of a compressively loaded interface,'' J. Acoust. Soc. Am. **113**(6), 3161–3170 (2003).
[10] M. Lowe, ''Matrix techniques for modeling ultrasonic waves in multilayered media,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**(4), 525–542 (1995).
[11] S. Rokhlin and L. Wang, ''Stable recursive algorithm for elastic wave propagation,'' J. Acoust. Soc. Am. **12(3)**, 822–834 (2002).
[12] B. Hosten and M. Castaings, ''Surface impedance matrices to model the propagation in multilayered media,'' Ultrasonics **41**, 501–507 (2003).
[13] U. Basu and A. K. Chopra, ''Perfectly matched layers for time-harmonic elastodynamics of unbounded domains: Theory and finite-element implementation,'' Comput. Methods Appl. Mech. Eng. **192**, 1337–1375 (2003).
[14] M. Castaings, C. Bacon, B. Hosten, and M. V. Predoi, ''Finite element predictions for the dynamic response of thermo-viscoelastic material structures,'' J. Acoust. Soc. Am. **115**(3), 1125–1133 (2004).

Frequency dependence of elastic properties of acoustic foams

M. Etchessahar[a] and S. Sahraoui[b]
Laboratoire d'Acoustique de l'Université du Maine-UMR CNRS 6613, Avenue Olivier Messiaen, 72085 Le Mans Cedex 9, France

L. Benyahia and J. F. Tassin
Laboratoire Polymères, Colloides, Interfaces-UMR CNRS 6120, Avenue Olivier Messiaen, 72085 Le Mans Cedex 9, France

(Received 18 September 2002; revised 8 December 2004; accepted 8 December 2004)

Polyurethane (PU) and other plastic foams are widely used as passive acoustic absorbers. For optimal design, it is often necessary to know the viscoelastic properties of these materials in the frequency range relevant to their application. A nonresonance technique (dynamic stiffness) based on a forced vibrations procedure is used to investigate the frequency dependent complex shear modulus of a PU foam. This modulus is first measured, in a quasistatic configuration, in the frequency range (0.016–16 Hz) at different temperatures between 0 and 20 °C. It is afterwards predicted over a wide frequency range (0.01–3000 Hz) using the frequency-temperature superposition principle. The validation of this principle is discussed through quasistatic experiments. Under the assumption that Poisson's ratio of polymeric foams is real and frequency independent on the frequency range used, the frequency dependence of the complex shear modulus obtained is used to predict the complex stiffness of the acoustic foam on a wide frequency range. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1857527]

PACS numbers: 43.35.Mr, 43.50.Gf [YHM] Pages: 1114–1121

I. INTRODUCTION

In the last few years, porous materials such as polyurethane (PU) foams have been widely used for passive sound absorption and noise control in automotive and aircraft applications. At low frequencies, the porous medium is characterized by geometrical parameters such as porosity, flow resistivity, tortuosity (Allard, 1993) and by the mechanical parameters of the frame. When the skeleton is set in motion, the dynamic behavior is described by Biot's theory (Biot, 1956) where the viscoelastic properties of the solid phase are involved in a wide frequency range.

Several techniques exist for the experimental determination of frequency dependent moduli of viscoelastic solids (Corsaro and Sperling, 1990, Oyadiji and Tomlimson, 1985). These techniques can be broadly classified as resonant and nonresonant and consist generally in measurements on a narrow frequency range associated with measurements at different temperatures. The complex moduli are built on a large frequency range using the frequency-temperature equivalence properties of polymeric viscoelastic materials. The choice of the appropriate experimental technique is usually governed by the geometry and loss factor of the material, and by the frequency range of interest.

To characterize the viscoelastic properties of PU acoustic foams, only a few experimental techniques are available in the literature and they consist of nonresonant techniques (Mariez *et al.*, 1996, Sahraoui *et al.*, 2001, Panneton and Langlois, 2001), standing wave resonance of a longitudinally excited rod with end mass (Pritz, 1994, Sfaoui, 1995) or mass-spring resonance (Pritz, 1990). They all perform measurements on a narrow frequency range. An extension of those measurements on a wide frequency range has been made by theoretical modelization of the viscoelastic behavior of foams (Pritz, 1996). When the tests are not performed in vacuum, the dynamic stiffness does not only depend on the mechanical properties of the skeleton but also on the satured fluid and on the coupling effects (Pritz, 1990). The influence of these last phenomena will grow with increasing flow resistivity and frequency. A torsion test is then suitable for neglecting the elastic and viscous forces as it ensures a constant volume of material.

The objective of this paper is to determine the complex shear modulus of a PU foam over the wide frequency range 0.01–3000 Hz needed for acoustic applications. We show that the frequency dependence of the complex shear modulus can be useful for further mechanical characterization. Such a frequency range at a given temperature results from shifting data measured on a much narrower frequency range but at various temperatures. The results are compared with data directly obtained with dynamical methods at low frequencies. Therefore, two experimental systems will be used for the characterization and two others for the validation.

In the first stage, the complex shear modulus of a PU foam usually used in sound absorbing applications is built experimentally over a large frequency range (0.01–3000 Hz). To do so, the sample is tested in torsion by means of a standard torsional rheometer (system 1), at low frequency (less than 16 Hz), at different temperatures (from 0 to 20 °C). The frequency–temperature superposition principle is used to compute the complex shear modulus at a reference temperature (20 °C) over a broader frequency range. A partial

[a]Present address: PSA Peugeot Citroen, route de Gisy, 78943 Vélizy-villacoublay Cedex, France.
[b]Electronic mail: sohbi.sahraoui@univ-lemans.fr

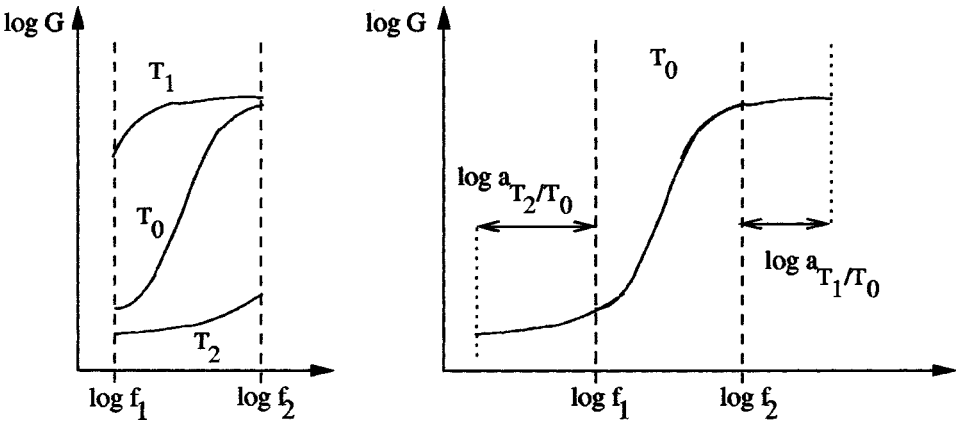


FIG. 1. Frequency–temperature superposition principle ($T_2 > T_0 > T_1$).

experimental validation is proposed by the use of quasistatic shear measurements (system 2) of the same material on the frequency range 1–100 Hz.

In the second stage, we show that the frequency dependence of the complex shear modulus can be useful for further mechanical characterization. While on theoretical grounds Poisson's ratio of viscoelastic materials could be complex and frequency dependent, we assume here that it is real and frequency independent. This assumption has been validated for foams either theoretically (Warren and Kraynik, 1988) or experimentally on the frequency range 1–100 Hz (Mariez *et al.*, 1996). Under this assumption, all foam's mechanical moduli present the same frequency dependence. As an application, we will deduce the stiffness of a foam sample on a wide frequency range (0.01–3000 Hz) from a quasistatic test (system 3) and the frequency dependence of the shear modulus determined previously. The method is validated at intermediate frequencies by means of transmissibility measurements of the mass-loaded sample (system 4).

II. THEORY

The frequency–temperature superposition principle, based on the reduced variables method, is useful for the study of the viscoelastic properties of polymeric materials. Indeed, the construction of master curves at a given temperature resulting from tests performed at other temperatures allows us to extend the range of the reduced frequency to a range that remains inaccessible to conventional dynamic rheometers (Ferry, 1961). A brief description of the principles of this method is given in the following and the response of material to dynamic mechanical solicitations is recalled.

A. Linear dynamic rheometry

If a material undergoes a sinusoidal strain of amplitude γ_0 at angular frequency ω:

$$\gamma = \gamma_0 \sin(\omega t), \tag{1}$$

its response, in terms of stress, is also sinusoidal at the same frequency but phase shifted:

$$\sigma = \sigma_0 \sin(\omega t + \delta), \tag{2}$$

where σ_0 is the amplitude of the stress, and δ is the phase angle. Asymptotic values 0 and $\pi/2$ of δ correspond, respectively, to the elastic solid and to the Newtonian fluid. For intermediate values, the material is said to be viscoelastic. Employing complex notations, the relationship between stress and strain is given through the complex modulus $G^*(\omega)$:

$$\sigma^* = G^*(\omega)\gamma^* \tag{3}$$

with

$$G^*(\omega) = G'(\omega) + jG''(\omega) \tag{4}$$

where $G'(\omega)$ and $G''(\omega)$ are, respectively, the so-called storage and loss moduli. At a given temperature, these moduli are only frequency dependent and are directly related to the amplitude σ_0 and to the phase angle δ by

$$G'(\omega) = \frac{\sigma_0}{\gamma_0}\cos(\delta), \tag{5}$$

$$G''(\omega) = \frac{\sigma_0}{\gamma_0}\sin(\delta). \tag{6}$$

The damping factor or loss angle can be calculated as

$$\eta = \tan\delta = G''(\omega)/G'(\omega). \tag{7}$$

To interpret the results in terms of molecular or structural parameters, the measurements must be done in the viscoelastic linear domain (Ferry, 1961). In this case, the response of the material remains independent of the solicitation amplitude.

TABLE I. Acoustical parameters of the tested PU foam.

Porosity ϕ	Flow resistivity σ (N m^{-4} s)	Tortuosity α_∞	Viscous length Λ (μm)	Thermal length Λ' (μm)	Density ρ (kg m^{-3})
0.97	5500	1.2	220	490	28.1

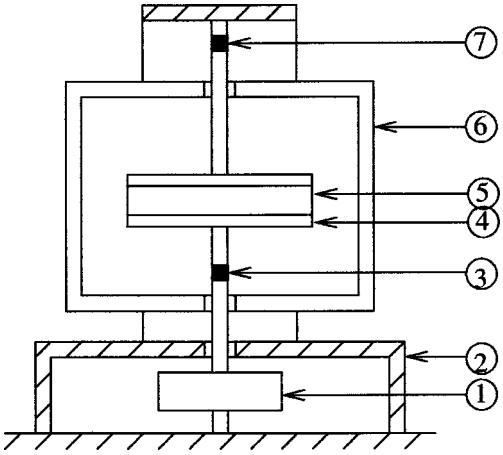


FIG. 2. Nonresonant experimental setup for dynamic torsional measurements (system 1). (1) Torsional excitator; (2) supporting frame; (3) angular displacement transducer; (4) supporting plate; (5) foam samples; (6) climatic environmental chamber; (7) force rebalanced torque transducer.

B. The frequency–temperature superposition

This method consists in translating horizontally and vertically the curves of $G'(\omega)$ and $G''(\omega)$ versus frequency, obtained on a reduced frequency range (typically between 0.016 and 16 Hz) at different temperatures, in order to superimpose them at a given temperature T_0 (Fig. 1). It is then equivalent to increase temperature or to reduce frequency and to decrease temperature or to increase frequency. This equivalence between frequency and temperature, relating the concept of reduced variables, results analytically in the following:

$$G'(T_i,\omega)=\frac{\rho_0 T_0}{\rho_i T_i}G'(T_0,a_{T_i/T_0}\omega)$$
$$=b_{T_i/T_0}G'(T_0,a_{T_i/T_0}\omega), \tag{8}$$
$$G''(T_i,\omega)=\frac{\rho_0 T_0}{\rho_i T_i}G''(T_0,a_{T_i/T_0}\omega)$$
$$=b_{T_i/T_0}G''(T_0,a_{T_i/T_0}\omega), \tag{9}$$

where a_{T_i/T_0} and b_{T_i/T_0} represent, respectively, the horizontal and the vertical shifts. All temperatures T are in degrees kelvin. Often b_{T_i/T_0} remains close to 1 because of the weak variation of the density ρ. The horizontal shift factor a_{T_i/T_0} is equal to the ratio of relaxation times at temperatures T_i and T_0. It is thus a sensitive measurement of the temperature dependence of relaxation times. The horizontal shift factors are determined experimentally. The curve at temperature T_i of one modulus (for example, the elastic modulus) is shifted horizontally as long as it perfectly superimposes with the corresponding curve at temperature T_0. The a_{T_i/T_0} factor found is then used to shift all curves at temperature T_i of other moduli. When such a technique is used, curves at temperatures T_i and T_0 must have a significant overlap in order to well determine the a_{T_i/T_0}. The shifting factors can follow different theoretical expressions depending on the temperature range used. In a large variety of polymeric systems, the WLF equation is frequently used. The frequency dependence of the shift factor is written as (Williams *et al.*, 1955)

$$\log a_{T/T_0}=\frac{-C_1^0(T-T_0)}{C_2^0+T-T_0}, \tag{10}$$

where the viscoelastic coefficients C_1^0 and C_2^0 depend both on the material and on the temperature T_0. The WLF equation is usually valid in the range $T_g<T<T_g+100\,^{\circ}\mathrm{C}$ where T_g is the glass transition temperature. In a higher range of temperature, the temperature dependence of the a_{T_i/T_0} shift factors can follow other types of laws (Ferry, 1961) like the Arrhenius equation:

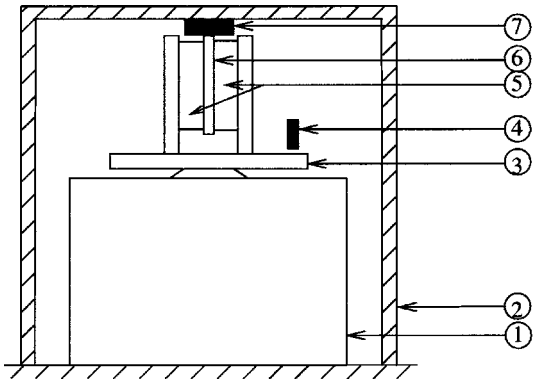


FIG. 3. Experimental setup for quasistatic shearing measurements (system 2). (1) Electrodynamic shaker; (2) supporting frame; (3) driving plates; (4) inductive displacement sensor; (5) foam samples; (6) middle plate; (7) piezoelectric force transducer.

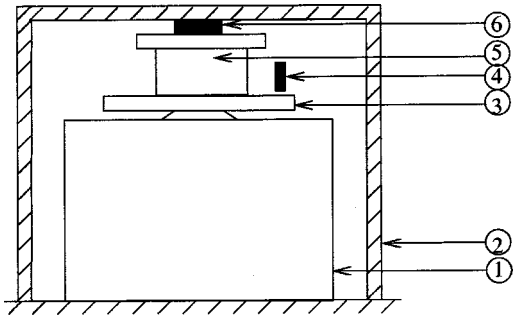


FIG. 4. Experimental setup for quasistatic compression measurements (system 3). (1) Electrodynamic shaker; (2) supporting frame (3) driving plate; (4) base accelerometer; (5) foam sample; (6) piezoelectric force transducer.

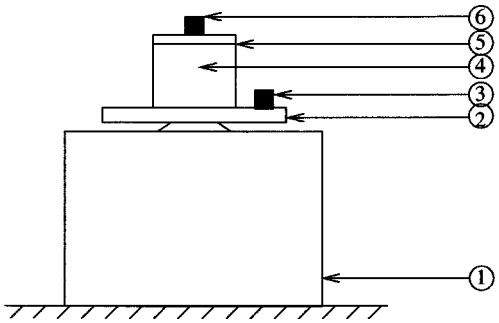


FIG. 5. Experimental setup for dynamic resonant measurements (system 4). (1) Electrodynamic shaker; (2) driving plate; (3) base accelerometer; (4) foam sample; (5) light aluminum plate; (6) top accelerometer.

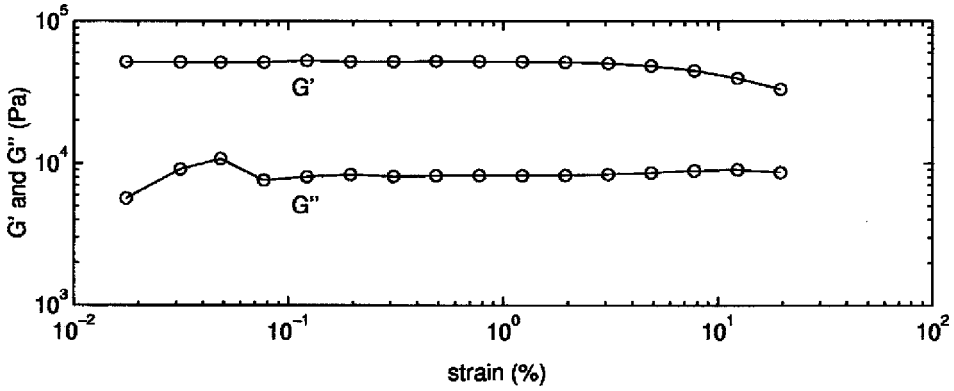


FIG. 6. Shear modulus at frequency $f=16$ Hz and temperature $T=20$ °C.

$$\log a_{T/T_0} = \frac{-E}{2.3RT}, \tag{11}$$

where E is the activation energy, R is the gas constant, and T the absolute temperature. Because these specific coefficients proper to the studied material (namely C_1^0 and C_2^0 or E) are often not known, the experimental determination is generally preferred.

III. MATERIAL PROPERTIES AND EXPERIMENTAL ARRANGEMENT

The acoustical properties of the tested foam sample are shown in Table I. The experimental setup can be divided into two parts: shear (systems 1 and 2) and compression (systems 3 and 4) set-ups.

A. Shear test

The experimental determination procedure is based on dynamic measurements of the shear modulus at different temperatures combined with data reduction using frequency–temperature superposition as described earlier. A nonresonant torsional experimental technique has been chosen as it ensures an excellent frequency resolution (Ferry, 1961). The frequency range of test is below the first resonant frequency of the sample.

1. Dynamic rheometer

Experimental measurements are performed with a Rheometrics RDAII apparatus. The experimental configuration employed is shown in Fig. 2. In order to avoid any slip, the cylindrical sample of foam of 50 mm diameter and 10 mm thickness is glued between two parallel plates via a two sided adhesive tape. One end of the sample is connected via a rigid fixture to a torque transducer while the other end is harmonically excited in torsion with a constant angular frequency. The stress σ_0 and the shear strain γ_0, calculated from measurements of the output torque and the applied angular displacement, are used through Eqs. (5) and (6) to calculate the real and imaginary parts, G' and G'', respectively, of the

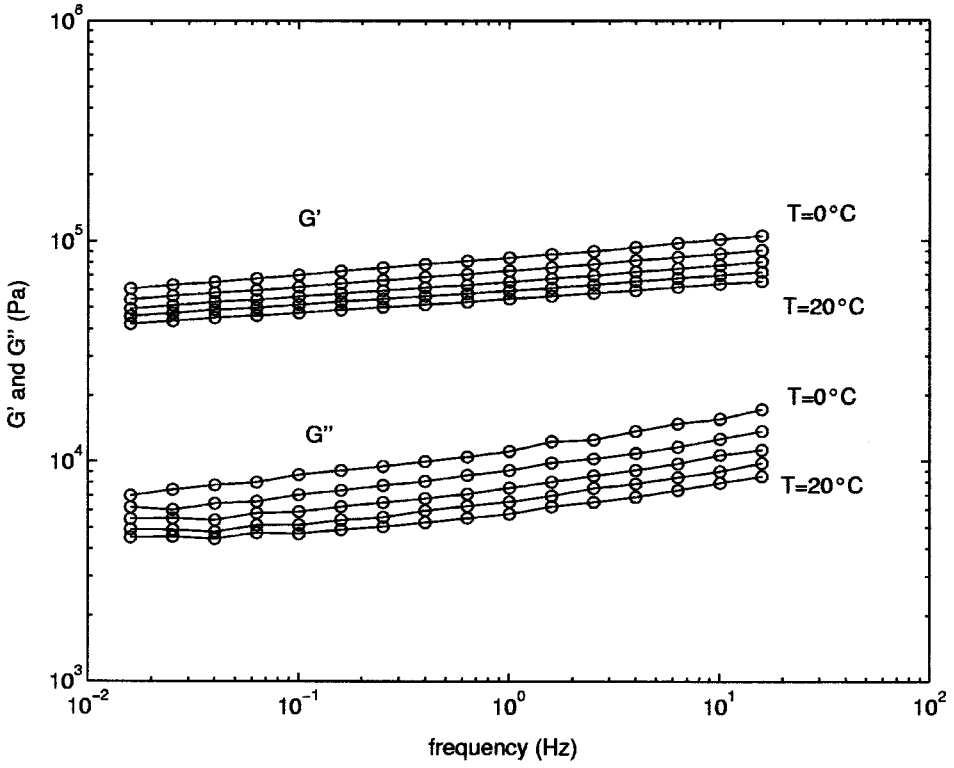


FIG. 7. Real and imaginary parts of shear modulus at temperatures 0, 5, 10, 15, and 20 °C.

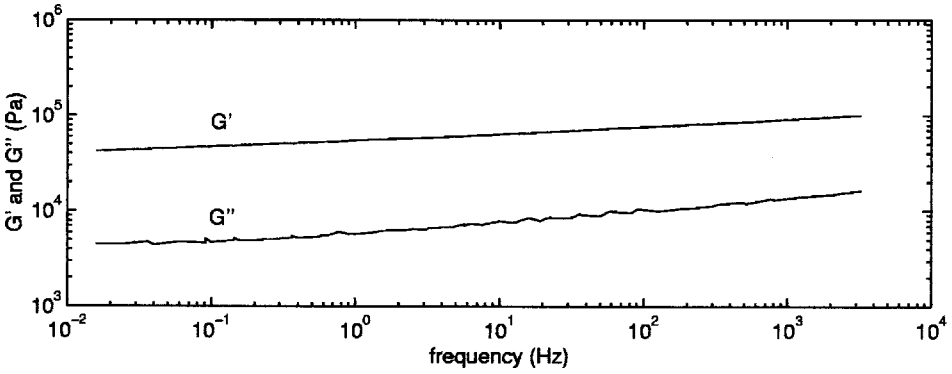


FIG. 8. Real and imaginary parts of reduced shear modulus at temperature $T_0 = 20\,^{\circ}\mathrm{C}$ over more than 5 decades.

complex shear modulus. Measurements are performed between 0.016 and 16 Hz in a temperature controlled environment.

2. Quasistatic experimental configuration

The experimental configuration is shown in Fig. 3. Two samples of foam of the same geometry (thickness $e = 10\,\mathrm{mm}$ and diameter $L = 50\,\mathrm{mm}$) are sandwiched between three parallel plates. The two external plates are firmly attached to each other and harmonically translated by an electrodynamic shaker using a 1–100 Hz sweep-sine. An inductive displacement sensor is used to measure the displacement u_L of the driving plates. This displacement is kept constant on all the frequency range. A piezoelectric force transducer placed between the top of the middle plate and the supporting frame is used to measure the transmitted force $F^*(\omega)$. A FFT analyzer computes the following mechanical impedance:

$$H^*(\omega) = \frac{F^*(\omega)}{u_L}. \tag{12}$$

Under the assumption that the thickness e of the sample is small compared to the other dimensions of the sample, the complex shear modulus $G^*(\omega)$ is given by

$$G^*(\omega) = \frac{H^*(\omega)e}{2S}, \tag{13}$$

where S is the area of the sample in contact with the middle plate.

B. Compression test

1. Quasistatic compression setup

The experimental setup (Fig. 4) is based on the setup described for quasistatic shear tests. The measurement method is described in Sahraoui *et al.* (2001).

2. Dynamic resonant setup

The experimental setup is shown in Fig. 5. The cylindrical foam sample is placed between the driving plate of an electrodynamic shaker and a light aluminum plate. Two loading masses (mass of the plate+mass of the accelerometer) of 6.4 and 16.3 g were used. The two accelerometers used to measure the input and the aluminum plate accelerations are connected to a FFT analyzer which computes the frequency response (transmissibility) of the mass loaded foam. The resonant system is excited by a 100–1000 Hz swept sine.

IV. RESULTS AND DISCUSSION

A. Dynamic complex shear moduli master curve

The sample was excited in torsion (with system 1) by a swept sine in the frequency range 0.016–16 Hz at 0.1% of dynamic strain in order to remain in the linear viscoelastic range, where the moduli are independent of the shear amplitude (Fig. 6).

The real G' and imaginary G'' parts of the complex shear modulus are presented in Fig. 7 for temperatures 0, 5, 10, 15, and 20 °C. Dependence on both frequency and temperature can be observed. An increase of G' and G'' with decreasing temperature as well as a power law dependence with frequency can be noticed. From these curves, a master

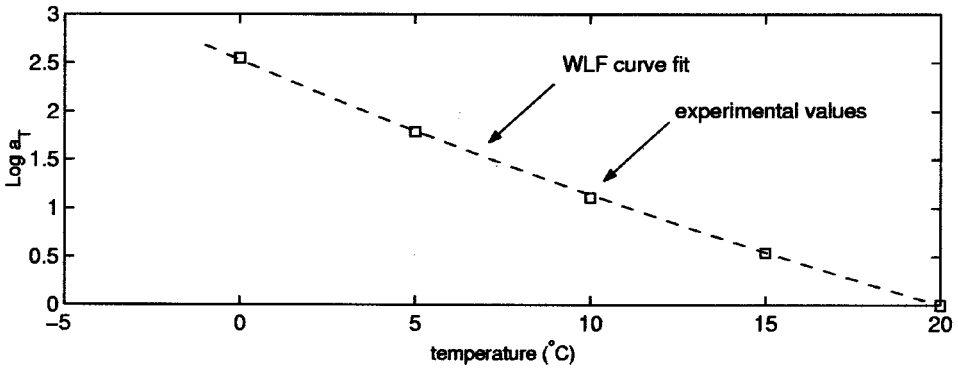


FIG. 9. Fit of the experimenatl values of a_T vs temperature by the WLF equation with coefficients $C_1^0 = 11.557$ and $C_2^0 = 111.4\,^{\circ}\mathrm{C}$.

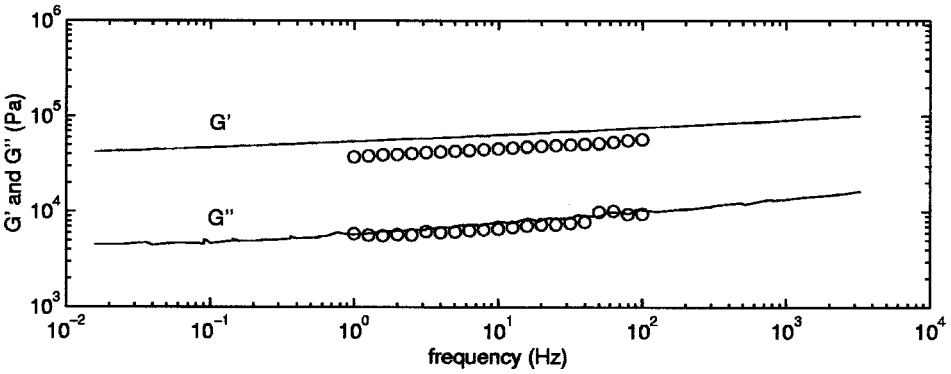


FIG. 10. Real and imaginary parts of the shear modulus at 20 °C: comparison between results found with the method of reduced variables and quasistatic shear measurements (system 2).

curve covering several decades of frequency at a single temperature is obtained using the experimental shifting procedure described Sec. II B.

The master curve is built at a reference temperature of 20 °C (Fig. 8) keeping in mind the fact that comparison with other experimental data obtained at room temperature will be considered. One can note that in the investigated temperature and frequency ranges, the frequency–temperature superposition is well adapted to the material. The master curve shows an extension of the frequency range (up to 3000 Hz) which cannot be reached with the existing apparatus. Here, factors b_{T/T_0} were about 1 since the five tested temperatures were very close from one another.

From the horizontal shift factors a_{T/T_0}, the viscoelastic coefficients C_1^0 and C_2^0 were calculated by means of a fit with the WLF equation (Fig. 9). With these viscoelastic coefficients, master curves of G' and G'' can now be predicted at other temperatures in the range 0–20 °C. It is important to point out that WLF coefficients may vary from one polymer to another so that coefficients corresponding to a different polymer can lead to large errors in the extrapolated values.

B. Frequency–temperature superposition principle validation

This section presents a partial validation of the experimental complex shear modulus obtained by means of the method of reduced variables. The complex shear modulus is measured on a new quasistatic apparatus which covers the frequency range 0.1–100 Hz. These results are compared to those obtained in Sec. IV A. Experiments were made at temperature $T_0 = 20$ °C and 0.08% of dynamic strain on similar samples from the same material. Figure 10 compares the complex shear moduli obtained by the two methods. We can notice on one hand that the two imaginary parts do exactly superimpose to one another, and on the other hand that even though the amplitudes of the two real parts are slightly different, they have the same frequency dependence. This last point might be enough to conclude that the frequency–temperature superposition principle is valid for foams since frequency dependence is an intrinsic property of a solid. The differences between the two measurements may result from the distinctive excitations used for the two methods: the first one is in torsion while the second one is in translation, and we already know that foams are mechanically anisotrops. Furthermore, foams are sensitive to external parameters such as the applied static load (Rodriguez-Pérez and Saja, 2000), which was not adjustable in this last experiment.

C. Application: Determination of the stiffness in compression test

According to Sahraoui *et al.* (2001) and Pritz (1994), the experimental determination of the complex compression stiffness is a key step in the calculation of the Young's modulus of foams. This experimental determination becomes difficult due to dynamic or coupling effects between the skeleton and the saturing fluid (Dauchez *et al.*, 2002). The coupling effects have a greater influence with higher flow resistivity and frequency. In the following we propose to determine experimentally the stiffness of the studied sample on a wide frequency range (0.01–3000 Hz) using the frequency dependence function of the complex shear modulus and a measured quasistatic value of this stiffness. The quasi-

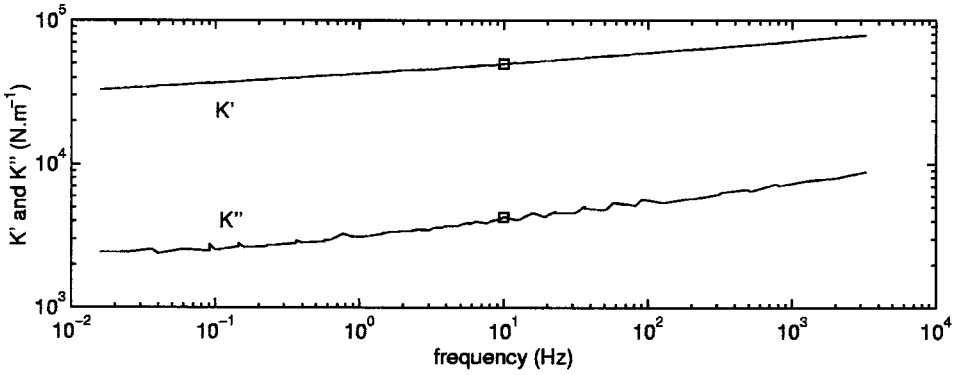


FIG. 11. Real and imaginary parts of predicted complex stiffness at 20 °C. The squares represents the measured quasistatic values at 10 Hz.

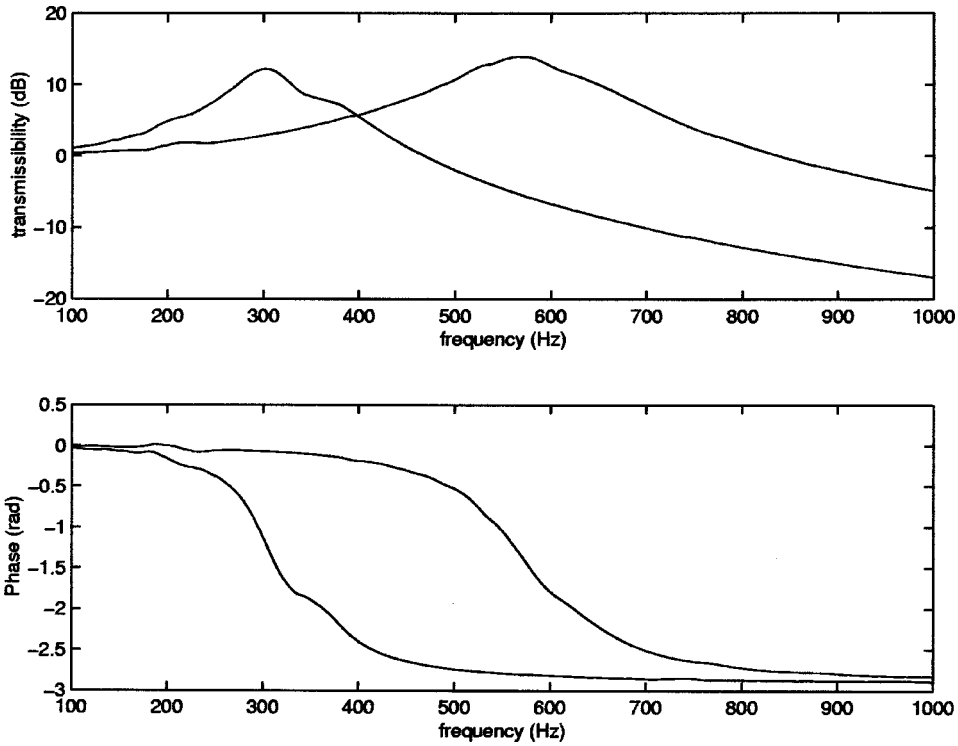


FIG. 12. Experimental frequency response of the mass loaded foam sample on the frequency range 100–1000 Hz. On the top picture, left curve: $m=16.3$ g, right curve: $m=6.4$ g.

static stiffness is measured with a classical compression test system and the measured quantity will be further called ''complex compression stiffness.''

1. Determination of stiffness in compression from complex shear modulus

The complex compression stiffness $K^*(\omega)$ depends on the geometry of the sample, on the boundary conditions, and on two elastic constants of the material. The complex compression stiffness is a function of $G^*(\omega)$, $\nu^*(\omega)$, geometry, and boundary conditions. Under the assumption that Poisson's ratio of polymeric foams is real and frequency independent on the frequency range used, the dynamic stiffness has the same frequency dependence as the shear modulus. A simple linear relation between the complex compression stiffness and the complex shear (or Young moduli) can then be written:

$$K^*(\omega)=\beta G^*(\omega), \tag{14}$$

where β is a real parameter depending on Poisson's ratio, geometry, and boundary conditions but not on frequency. Writing these complex quantities as the sum of their real and imaginary parts, Eq. (14) takes the following form:

$$\log K'(\omega)=\log G'(\omega)+\log\beta, \tag{15}$$

$$\log K''(\omega)=\log G''(\omega)+\log\beta. \tag{16}$$

These last equations show that, on a logarithm scale, one can deduce the complex compression stiffness from the complex shear moduli by shifting the real and imaginary parts by $\log\beta$. Here, the complex compression stiffness is predicted over a large range of frequency (0.1–3000 Hz) (Fig. 11) using the frequency dependence function of the shear modulus computed previously and a quasistatic value of the complex compression stiffness. The complex compression stiffness has been measured at 10 Hz using the apparatus described in Sahraoui *et al.* (2001): the PU foam sample (diameter: 50 mm, thickness: 10 mm) is tested in compression at low strain rate and the complex stiffness is deduced from measurements of the imposed displacement and of the transmitted force. The shift factor has been calculated as follows:

$$\log\beta=\log K'_{10}-\log G'_{10}, \tag{17}$$

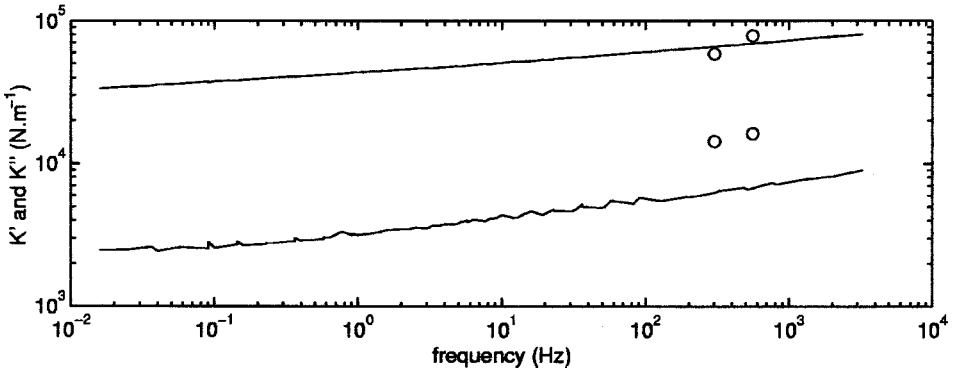


FIG. 13. Real and imaginary parts of predicted complex stiffness at 20 °C. The circles symbolize the measured values at 302 and 568 Hz by the resonant method (system 4).

where K'_{10} and G'_{10} are the measured quasistatic compression stiffness and the shear modulus at 10 Hz.

2. Dynamic stiffness measurement

The objective of the following is to compare the complex stiffness obtained from a measurement of the frequency response of a mass-loaded sample to the one predicted in Sec. IV C 1. The low flow resistivity of the sample (see Table I) leads us to neglect the effects of the air on the response of the system (Dauchez *et al.*, 2002). In addition, since the mass of the sample is much lower than the total loading mass (mass of the plate+mass of the accelerometer+mass of the sample/2) the system is at first considered as an ideal one-degree-of-freedom system with pure masslike and stiffness-like parameters. The natural frequency of the system f_0, defined as the frequency for which the amplitude of the loading mass is maximum, and its transmissibility T are hence given by the following classical relations (Snowdon, 1968):

$$f_0 = \frac{1}{2\pi}\sqrt{\frac{K}{m}}, \quad T = \frac{(1+\eta^2)^{1/2}}{([1-\Omega^2]^2+\eta^2)^{1/2}}, \tag{18}$$

where the damping factor η is defined by Eq. (7) and Ω represents the frequency ratio f/f_0. Figure 12 presents the two experimental transmissibilities and phase differences of the system for the two masses.

The two resonant frequencies are found near 302 and 568 Hz for transmissibility values of 4.22 and 5.53, respectively. The real and imaginary parts of the complex stiffness, calculated with Eqs. (18) and (7), are reported in Fig. 13. A good agreement is found on the real parts while the calculated imaginary parts are overestimated. Such results have already been observed by Pritz (1990) and can be caused by viscous dissipation due to interaction between the air and the skeleton.

V. SUMMARY AND CONCLUSIONS

The method proposed in this paper gives a new approach of the experimental characterization of the viscoelastic constants of polymeric foams on a large frequency range. In this paper we showed that the frequency–temperature superposition principle, usually used in the characterization of mechanical properties of bulk materials, can be applied to polymeric foams. The complex shear modulus computed on a large frequency range by means of the frequency–temperature superposition principle has been validated by quasistatic shear measurements on the frequency range of 1–100 Hz. In addition, we showed that, in the particular case of open-cell foams, the complex compression stiffness can be deduced from the frequency dependence function of a mechanical modulus and a simple quasistatic measurement. This application has been performed under the assumption that Poisson's ratio is real and frequency independent of the frequency range relevant to the application.

Furthermore, this method avoids difficulties associated with the difficulties of the mechanical characterization of porous sound absorbing materials. Indeed, all measurements being performed at low frequencies, neither coupling effects between the skeleton and the saturated fluid leading to complex inversion, nor difficulties due to dynamic effects occur using this technique.

ACKNOWLEDGMENTS

The authors would like to thank V. Gareton, L. Jaouen, and B. Brouard for all the fructuous discussions and stimulating comments during this work.

Allard, J. (**1993**). *Propagation of Sound in Porous Media: Modeling Sound Absorbing Materials* (Chapman et Hall, Londres).

Biot, M. (**1956**). ''The theory of propagation of elastic waves in a fluid-satured porous solid. i. Low frequency range. ii. Higher frequency range,'' J. Acoust. Soc. Am. **28**, 168–191.

Corsaro, R., and Sperling, L. (**1990**). *Sound and Vibration Damping with Polymers* (American Chemical Society).

Dauchez, N., Etchessahar, M., and Sahraoui, S. (**2002**). ''On measurements of mechanical properties of sound absorbing materials,'' in Second Biot Conference on Poromechanics, Grenoble, France.

Ferry, J. (**1961**). *Viscoelastic Properties of Polymers* (Wiley, New York).

Mariez, E., Sahraoui, S., and Allard, J. (**1996**). ''Elastic constants of polyurethane foam's skeleton for biot model,'' in Internoise 96, Liverpool, Great Britain, pp. 951–954.

Oyadiji, S., and Tomlimson, G. (**1985**). ''Determination of the complex moduli of viscoelastic structural elements by resonance and non-resonance techniques,'' J. Sound Vib. **101**, 277–298.

Panneton, R., and Langlois, C. (**2001**). ''Polynomial relation for the mechanical characterization of poroelastic materials,'' in ICA 01, Roma, Italie.

Pritz, T. (**1990**). ''Non-linearity of frame dynamic characteristics of mineral and glass wool materials,'' J. Sound Vib. **136**, 263–274.

Pritz, T. (**1994**). ''Dynamic young's modulus and loss factor of plastic foams for impact sound isolation,'' J. Sound Vib. **178**, 315–322.

Pritz, T. (**1996**). ''Analysis of four-parameter fractional derivate model of real solid materials,'' J. Sound Vib. **195**, 103–115.

Rodriguez-Pérez, M., and Saja, J. D. (**2000**). ''Dynamic mechanical analysis applied to the characterisation of closed cell polyolefin foams,'' Polym. Test. **19**, 831–848.

Sahraoui, S., Mariez, E., and Etchessahar, M. (**2001**). ''Mechanical testing of polymeric foams at low frequency,'' Polym. Test. **20**, 93–96.

Sfaoui, A. (**1995**). ''On the viscoelasticity of the polyurethane foam,'' J. Acoust. Soc. Am. **97**, 1046–1052.

Snowdon, J. (**1968**). *Vibration and Shock in Damped Mechanical Systems* (Wiley, New York).

Warren, W., and Kraynik, A. (**1988**). ''The linear elastic properties of open-cell foams,'' J. Appl. Mech. **55**, 341–346.

Williams, M., Landel, R., and Ferry, J. (**1955**). ''The temperature dependence of relaxation mechanisms in amorphous polymers and other glass-forming liquids,'' J. Am. Chem. Soc. **77**, 3701–3707.

Use of a curved reflector to amplify ultrasonic standing waves in an air-filled channel

Michael J. Anderson,[a] Andrew C. Cluff, E. Clark Lemmon, and Gabriel C. Putnam
Department of Mechanical Engineering, University of Idaho, Moscow, Idaho 83844

(Received 27 April 2004; revised 27 October 2004; accepted 20 December 2004)

The application of a slightly curved reflector to increase the amplitude of an ultrasonic standing wave in a semi-infinite rectangular channel was explored. Air was assumed to be the acoustic medium in the channel. Excitation of the standing wave was assumed to be provided by a square transducer flush-mounted to one wall of the channel. A slight curvature was placed in the reflecting wall of the channel. A finite element analysis was used to predict the amplitude of the standing wave that would be excited in the channel. A perfectly matched layer was used to model the semi-infinite channel geometry. At frequencies near 50 kHz, for source ka ranging from 6.6 to 26.6, and channel depths necessary to excite standing waves at one-half and one wavelength resonance, the computations predicted that an increase in acoustic pressure amplitude from 2 to 11 dB could be achieved with a reflector whose depth of curvature was 16% of the channel depth. Much of this increase could be obtained with curvatures of smaller depth. Experiments with a channel and reflector of representative geometry gave a measured increase in acoustic pressure amplitude of 4.86 dB. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1858171]

PACS numbers: 43.35.Zc, 43.20.Ks [YHB] Pages: 1122–1128

I. INTRODUCTION

An emerging application for ultrasound is the use of radiation pressure to manipulate or sort small particles suspended in moving fluid, either water,[1–3] or air.[4–6] In this application, the fluid flows through a channel. A resonant acoustic standing wave is excited in the channel, propagating perpendicular to the fluid flow direction. Forces exerted on the particles by radiation pressure cause the particles to move to pressure nodes of the standing wave as they are carried along with the mean fluid flow. Of crucial importance to the performance of such a device is the amplitude of the standing acoustic pressure field.

Acoustic pressure amplification in a flow channel is often achieved by choosing the channel depth to excite an acoustic cavity resonance, in spite of the fact that the cavity is not entirely closed. The pressure amplification is limited by the absorption of the flowing medium, thermal conduction and viscous shear at the channel boundaries, and by the loss of acoustic energy that propagates through the channel openings and does not return. For a rectangular channel with parallel walls, it was found that choosing a frequency high enough to collimate the standing wave and minimize acoustic energy passage through the channel openings, although not too high to cause excessive bulk-fluid and boundary absorption, could optimize the pressure amplification in a channel with parallel walls.[7] We show that a significant gain in pressure amplification can be achieved by imposing a slight curvature, significantly less than a wavelength in depth, into the reflecting channel wall.

Some consideration has been given in the past to the design of reflectors used in acoustic levitation. Ideally, it is desirable to account for acoustic energy losses due to absorption, and propagation of acoustic energy away from the standing wave between the source and reflector. Xie and Wei[8] used a boundary element analysis to determine the acoustic standing wave field that can be obtained by using a curved reflector in acoustic levitation. Because their model did not include any mechanism of absorption, their predictions of pressure amplification may not be accurate in an absolute sense. Their analyses and experiments, with air as the working medium, considered sources and reflectors of relative size ka in the range $3.86<ka<7.73$, a reflector radius of curvature R in the range $0<R<5\lambda$, and source-reflector separations extending to $3\lambda/2$, where λ is the acoustic wavelength. Their experiments showed that the relative increase in pressure amplification that would be achieved by changing from one reflector geometry to another was sufficiently explained by the boundary element model. It was apparent, for the relative size ka ranges considered, that a spherical reflector with a radius of curvature near one wavelength would be optimal. Zheng and Apfel[9] employed a finite element model to predict the acoustic standing wave field caused by an acoustic source and reflector used in levitation experiments conducted in water. They used a perfectly matched layer to model the propagation of acoustic waves away from the standing wave between the source and reflector. They did not report whether their model was used to optimize the source-reflector geometry to increase acoustic pressure amplitude in the standing wave field.

The present application, with air as the working fluid, requires a set of circumstances somewhat different from many levitation applications. It is desirable to operate at a high frequency, in excess of 50 kHz so that transduction efficiency is improved, and particle diameter-to-wavelength ratio is favorable for forces caused by radiation pressure. In this frequency range, ka values for the source and reflector

[a]Author to whom correspondence should be addressed; electronic mail: anderson@uidaho.edu

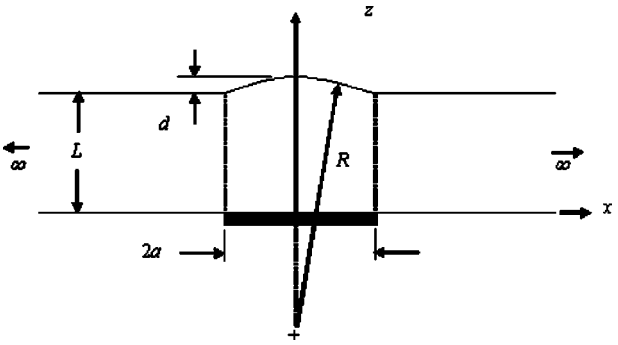


FIG. 1. Geometry of shaped channel.

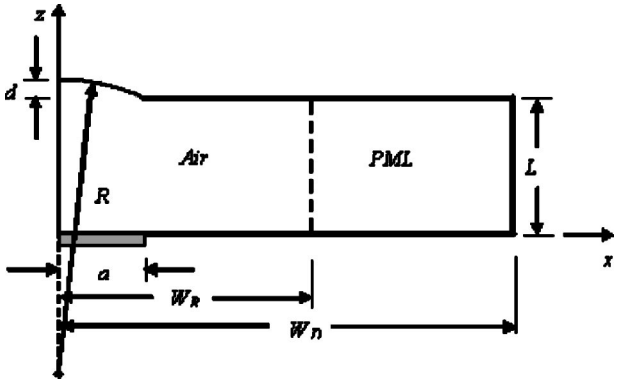


FIG. 2. Geometry used for finite-element analysis.

are somewhat higher than 10. Absorption in the body of the fluid, and that caused by viscous shear and thermal conduction at the boundaries also become important relative to leakage of acoustic energy away from the standing field through the channel openings. Finally, it is desirable to minimize the distortion of the standing wave field from a collimated planar geometry in order that fluid-flow streamlines do not become mixed. This would imply that reflector curvature used to amplify the standing wave should be held as small as possible.

In what follows, we considered the excitation of an acoustic standing wave field in a two-dimensional semi-infinite channel with walls that were nominally parallel. The working medium was air. A finite aperture ultrasonic transducer was flush-mounted into one wall of the channel. Opposite the transducer, a small cylindrical deviation was placed in the reflecting wall. A finite element method was used to calculate the acoustic standing wave field that is generated by the transducer. The model accounted for the body fluid absorption in the fluid, and viscous shear and thermal conduction at the boundaries.

II. PROBLEM DEFINITION

The shaped two-dimensional channel under consideration is shown in Fig. 1. The nominal channel geometry consisted of parallel rigid walls, separated by a distance L. An ultrasonic transducer of width $2a$ was flush-mounted in one wall. The transducer was assumed to move as a plane piston at circular frequency ω and velocity amplitude U. A cylindrical indentation, or reflector, with fixed radius of curvature R was placed in the wall directly opposite the transducer. The depth of the reflector was denoted as d. For a given transducer of width $2a$, channel depth L, and indentation depth d, the radius of curvature R was given by

$$R=\frac{a^2+L^2-(L+d)^2}{2d}+L+d=d\left[\frac{1}{2}\left(\frac{a}{d}\right)^2+2\right]. \tag{1}$$

In the limit of $d\rightarrow 0$, the geometry reverted to a parallel wall channel with no reflector.

A model equation that accounted for body absorption in the fluid, viscous shear and thermal conduction at the boundaries, was used to determine the standing acoustic pressure field in the channel caused by the motion of the transducer. The model equation was

$$\frac{\partial^2 p}{\partial t^2}=c_0^2\left(1+\tau\frac{\partial}{\partial t}\right)\left(\frac{\partial^2 p}{\partial x^2}+\frac{\partial^2 p}{\partial z^2}\right), \tag{2}$$

where p was the acoustic pressure, c_0 was the equilibrium speed of sound in air, t was time, and τ was a relaxation appropriate for body absorption in the fluid, viscous, and thermal dissipation at the boundaries. In a previous work, it was determined that these physical processes could be adequately modeled by an appropriate absorption coefficient α, which is related to the relaxation parameter τ by $\tau=2\alpha c_0/\omega^2$. A harmonic solution $p=P(x,z)e^{j\omega t}$ was sought for the semi-infinite domain. Boundary conditions necessary for a solution were rigid walls, with the exception of the transducer location $-a<x<a$, where a velocity of $Ue^{j\omega t}$ normal to the boundary was assumed.

III. FINITE ELEMENT MODELING

A finite element technique was used to determine the acoustic pressure amplitude $P(x,z)$ that would be encountered in the semi-infinite domain. To apply a finite element technique to the semi-infinite domain specified in Sec. II, some procedure must be applied to model the propagation of acoustic waves in the x direction without reflection. In 1995, an absorbing boundary layer was introduced to eliminate reflections caused by artificial truncation of a domain.[10] This layer is composed of an artificial material. Analytically, it can be shown that such a layer would absorb plane waves without reflection, independent of the angle of incidence or frequency of the incident waves.[10,11] The absorbing layer, called a ''perfectly matched layer'' (PML), was used here to account for the acoustic waves in air propagating out of the domain of interest.

The geometry adopted for the finite element analysis is shown in Fig. 2. For the domain $x<W_R$ marked ''Air'' in Fig. 2, Eq. (2) was solved numerically assuming air was the medium. For the domain $W_R<x<W_D$ marked ''PML'' in Fig. 2, the coupled equations for the PML,

$$\frac{\partial^2 p_u}{\partial t^2}+2\sigma_x\frac{\partial p_u}{\partial t}+\sigma_x^2 p_u=c_0^2\frac{\partial^2 p_u}{\partial x^2}+c_0^2\frac{\partial^2 p_v}{\partial x^2}, \tag{3}$$

$$\frac{\partial^2 p_v}{\partial t^2}+2\sigma_z\frac{\partial p_v}{\partial t}+\sigma_z^2 p_v=c_0^2\frac{\partial^2 p_u}{\partial z^2}+c_0^2\frac{\partial^2 p_v}{\partial z^2}, \tag{4}$$

were solved numerically. In Eqs. (3) and (4), p_u and p_v were pressures associated with fluid velocities in the z and x directions, respectively. The symbols σ_x and σ_y are factors that control the absorption of acoustic waves transmitted into the perfectly matched layer.

Galerkin's formulation of the method of weighted residuals[12] was used to determine the time-harmonic acoustic pressure $P(x,z)$ that would be excited by the transducer in the semi-infinite cavity with a shaped reflector. Symmetry about the z axis was exploited to reduce the size of the computational domain (Fig. 2). A PML domain was extended along the semi-infinite x direction allowing for a finite-sized domain with semi-infinite behavior at the resonator outlet $x = W_R$.

It must be noted that although analytically it is possible to transition instantaneously from a normal domain to a PML domain without error, a step transition in numerical modeling is not possible without causing reflections back into the domain of interest. To avoid these errors, it was necessary to gradually increase the value of the absorption control coefficients σ_x and σ_y from zero at the transition boundary $x = W_R$ to a positive value at the domain boundary $x = W_D$. To determine an appropriate absorption layer depth and absorption coefficient dependence on the x coordinate, one adjusts the PML layer depth W_D-W_R and the absorption coefficient values until the finite element solution agrees with a known solution.

Because a perfectly matched layer was used to absorb all outgoing waves, all boundaries, including the end cap, were assumed to be rigid with the exception of the transducer location, where a normal velocity of $Ue^{j\omega t}$ was enforced. Triangular elements were used with a typical discretization on the order of 144 elements/wavelength. A gradient in element size of 0.8% in front of the transducer and 1.0% elsewhere was used. The mesh density was increased until solution convergence was verified. The finite element solution to a plane surface reflector was then compared to a known solution.[7] With respect to this benchmark, the numerical accuracy of the finite element model was observed to be within 0.24% of the analytical solution.

IV. EXPERIMENTS

Experiments were performed to compare with theoretical predictions of acoustic pressure gains that could be achieved using a curved reflector. A rectangular channel, of width 29 ± 0.125 mm, depth L 6.86 ± 0.05 mm, and length 127.0 ± 0.25 mm was fabricated from aluminum and Lucite. A photograph of this channel is shown in Fig. 3. In the photograph, the channel is dissembled into two parts. Note the position of microphone ports in the reflector. In the photograph, two ports are plugged and one port is open. Relative to the photograph shown in Fig. 3, the reflector was turned over and placed over the assembly containing the transducer to form the channel. A square piezoelectric transducer of side width 28.55 ± 0.05 mm and thickness 4.0 ± 0.025 mm was flush-mounted to one wall of the channel and provided a standing acoustic wave between the transducer wall and the reflector. The piezoelectric material was APC 841 (manufactured by American Piezoceramics), with nominal elastic compliance at constant electric field $s_{33}^E = 15.87(10)^{-12}$ m^2/N, piezoelectric stress constant $d_{33} = 300$ pm/V, and dielectric permittivity at constant stress $\varepsilon_{33}^T = 11.95$ nF/m. Alignment of the piezoelectric transducer with the reflector was provided by a three-point fine-thread adjustable screw mechanism recessed behind the transducer. Two reflectors were used; one was flat, while the second contained a circular indentation corresponding to that described in Sec. II and Fig. 1. The curved reflector was fabricated using a numerically controlled mill. The depth of the circular indentation was measured to be 1.05 ± 0.025 mm, which corresponded to a normalized reflector depth $d^* = d/L \approx 16\%$. This depth was chosen to maximize dimensional control over reflector curvature during fabrication. A 1/8-in.-diam measurement microphone of nominal sensitivity 1.05 mV/Pa and bandwidth 40 Hz–150 kHz was flush-mounted in the reflector directly opposite the piezoelectric transducer to measure the acoustic pressure at the boundary of the channel.

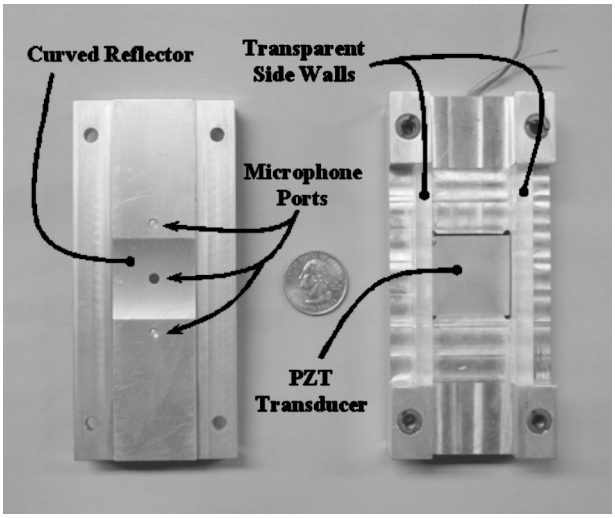


FIG. 3. Rectangular channel used in experiments.

The experiments were conducted with the intent to compare with finite element predictions. In the experiments, a curved reflector was installed in a channel of given dimensions. This caused a change in the effective depth of the channel, and a corresponding change in depth resonance frequency. Consequently, in the experiments the amplitude of the standing wave was affected by the type of reflector and the change in velocity amplitude U of the transducer associated with the change in depth resonance frequency. It was necessary to specify the transducer velocity amplitude U in the finite element computations to correspond to the transducer velocity amplitudes that were encountered in the experimental measurements for a fair comparison between theory and experiment.

A laser-vibrometer was used to measure the velocity amplitude of the piezoelectric transducer. At the setting used to take the velocity amplitude measurements, the laser vibrometer was capable of detecting vibration velocities as small as 0.5 μm/s over a bandwidth 1 Hz–250 kHz. These measurements were performed by removing the reflector side of the channel to expose the transducer. The microphone or

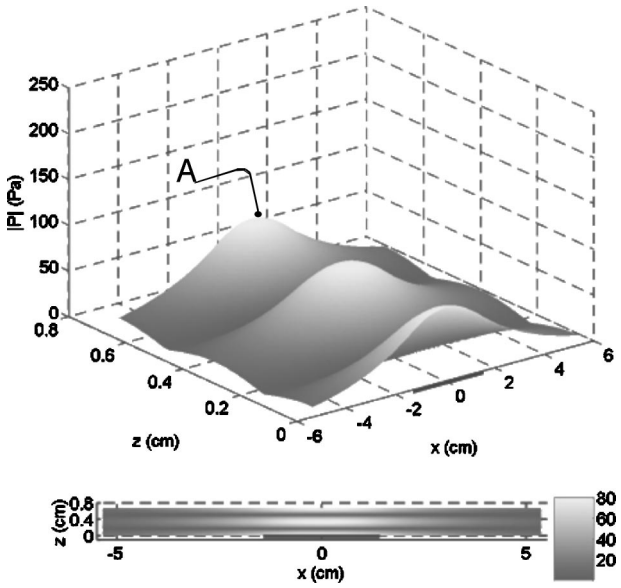


FIG. 4. Acoustic pressure amplitude, $d^*=0$ (no curved reflector). (a) Surface plot, (b) contour plot, actual geometric aspect ratio.

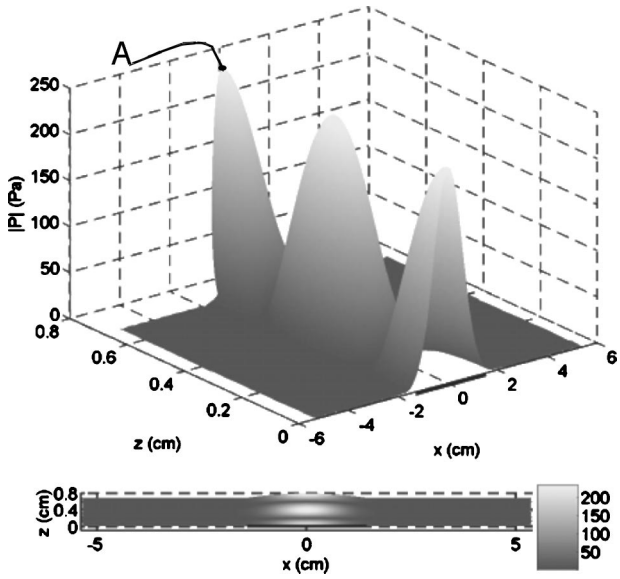


FIG. 5. Acoustic pressure amplitude, $d^*=16$ (with curved reflector). (a) Surface plot, (b) contour plot, actual geometric aspect ratio.

vibrometer signal was recorded with a digital spectrum analyzer. This analyzer had a bandwidth of 102 kHz and resolution of up to 1600 frequency lines.

V. RESULTS AND DISCUSSION

The intent was to use finite element analysis to choose a reflector depth d that would increase pressure amplification between the channel walls while minimizing geometric distortion of the channel shape or acoustic standing wave. We first present the determination of acoustic pressure for a given case, and then proceed to discuss the use of the finite element analysis to explore the effect of reflector depth d on acoustic pressure amplification.

Example computations of acoustic pressure amplitude for a given geometry and frequency are shown in Figs. 4 and 5. In these computations, parameter values of $U=2.95$ mm/s, $2a=28.57$ mm, $L=6.8$ mm, $c_0=343$ m/s, $\tau=8.16$ ns, and a density of $\rho_0=1.21$ kg/m^3 were used. In Fig. 4, the reflector depth d was zero, i.e., there was no curved reflector opposite the transducer. For this situation, the frequency was $f=\omega/2\pi=50$ kHz, such that the channel depth L was exactly one wavelength. A significant amount of acoustic energy propagated away from the standing wave directly in front of the transducer when a curved reflector was not present. The maximum pressure amplitude, found at point A shown in Fig. 4, was 80 Pa. In Fig. 5, the normalized reflector depth $d^*=d/L$ was set to 16%. The radius of curvature for this reflector R was 94 mm, or $13.87L$, many times greater than the nominal wavelength. In this case, the frequency for a nominal one-wavelength depth resonance was determined to be $f=44.084$ kHz. When a reflector of normalized depth $d^*=16\%$ was placed opposite the transducer, most of the acoustic energy was confined to the standing wave directly in front of the transducer and little acoustic energy propagated away from the standing wave. Consequently, the pressure amplitude at point A in Fig. 5 was 234 Pa, about 2.93 times (9.34 dB) greater than the pressure amplitude observed when a curved reflector was not present.

Xie and Wei[8] reported an increase in radiation pressure by a factor of 5431, which would correspond to a pressure increase by a factor of 74, when a reflector with radius of curvature one-wavelength was placed opposite a transducer of relative size $ka=3.86$. Although this amplification seems to greatly exceed that shown in the present examples, it must be remembered that the model used by Xie and Wei did not include any mechanism of absorption. If a perfect reflector was used in their calculations, the pressure amplification would be infinite when the reflector trapped all of the acoustic energy in the standing wave since the only losses in their model were caused by propagation of acoustic energy away from the standing wave between the source and reflector. In the calculations presented here, a realistic model for absorption was used which more closely predicts the absolute amplification that can be achieved in practice. Finally, the relative size of the source used in the calculations shown in Figs. 4 and 5 was $ka=13.2$ at 50 kHz, which was somewhat larger than $ka=3.86$ considered by Xie and Wei. The nearly 9.34 dB in acoustic pressure increase is surprising given a reflector depth of only 16% of the nominal channel depth.

Finite element analysis was used to explore the effect of reflector curvature on resonant amplification. For a given normalized reflector depth d^*, the acoustic pressure field was computed for a range of frequencies. From these data, it was possible to isolate the acoustic pressure amplitude at location A $(x=0,z=L+d)$ and determine the $\lambda/2$ and λ depth resonance frequencies, the maximum pressure amplitude at resonance, and the associated quality factors. The quality factor was extracted from the acoustic pressure amplitude at location A using the half-power point method. Common parameters for the computations were $U=2.95$ mm/s, $\rho_0=1.21$ kg/m^3, $c_0=343$ m/s; $L=6.8$ mm, and $L=3.4$ mm for λ and $\lambda/2$ depth resonances, respectively. To explore the effect of beam collimation on reflector perfor-

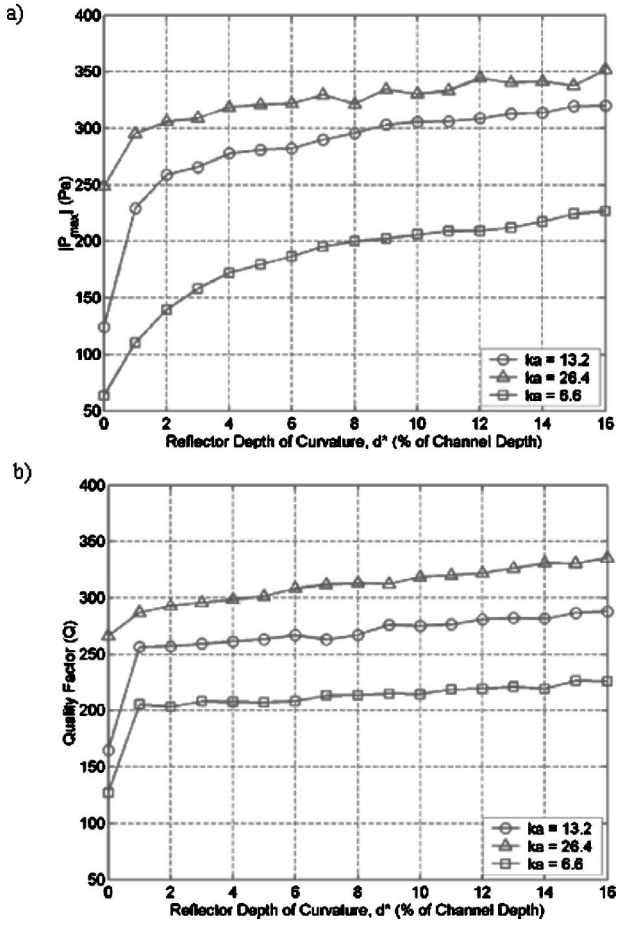


FIG. 6. Maximum acoustic pressure and quality factor for λ/2 resonance, (a) acoustic pressure amplitude, (b) quality factor.

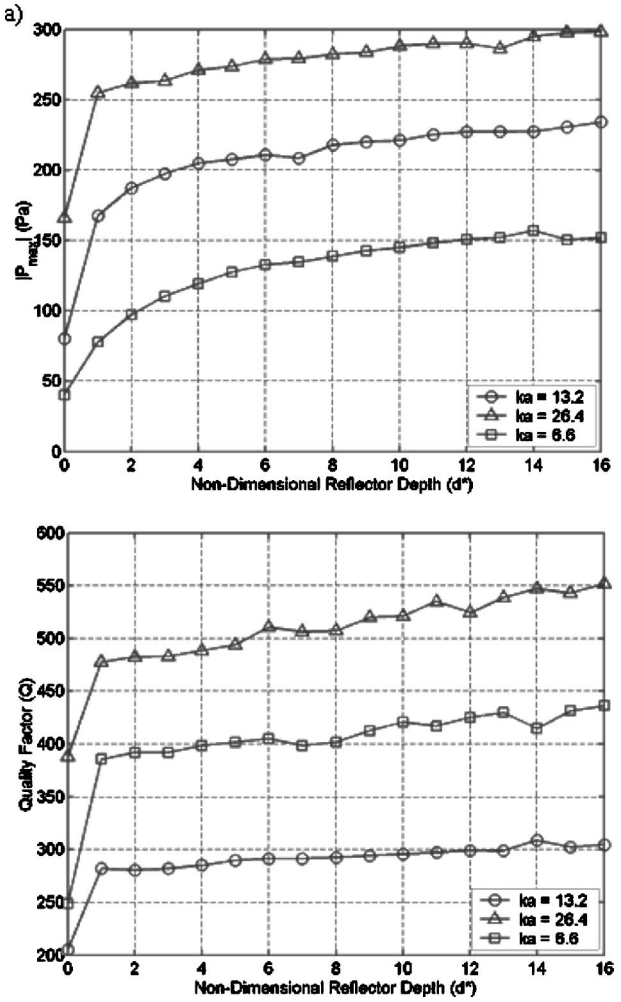


FIG. 7. Maximum acoustic pressure and quality factor for λ resonance, (a) acoustic pressure amplitude, (b) quality factor.

mance, three transducer widths were considered, $2a = 14.29$ mm, $2a = 28.57$ mm, and $2a = 57.14$ mm.

Plots of the maximum acoustic pressure amplitude and quality factor versus normalized reflector depth d^* for a λ/2 depth resonance are shown in Fig. 6. For Figs. 6(a) and 6(b), the horizontal axis is the normalized reflector depth d^*. In Fig. 6(a), the vertical axis is the acoustic pressure amplitude at location A, and in Fig. 6(b), the vertical axis is the quality factor for the λ/2 depth resonance. In Figs. 6(a) and 6(b), three curves are plotted. The curves correspond to the transducer widths $2a = 14.29$ mm, $2a = 28.57$ mm, and $2a = 57.14$ mm. Based upon a nominal resonance frequency of 50 kHz for the λ and λ/2 depth resonances with the flat reflector, the transducer widths correspond to relative source sizes of $ka = 6.6$, 13.2, and 26.4, respectively. A similar graphical summary of the effect of normalized reflector depth d^* on the nominal λ depth resonance is shown in Fig. 7.

In each case, it was observed that the presence of a reflector with curvature of relatively small depth could significantly increase the resonant amplification in the channel. For a λ/2 depth resonance [Fig. 6(a)], and a source ka of $ka = 6.6$, the acoustic pressure amplitude gradually increased from 64 to 226 Pa (11.0 dB) as the normalized reflector depth d^* reached 16%. When the source ka was increased to $ka = 26.4$, the pressure amplitude increased rapidly from 249 Pa at $d^* = 0\%$ to 320 Pa as the normalized reflector depth d^* reached 5%. These observations are consistent with a collimated standing wave requiring a smaller reflector curvature to limit acoustic energy leakage, and that smaller gain in resonant amplification would be achieved. In Fig. 6(b), it was observed that the quality factor for the λ/2 depth resonance increased rapidly with normalized reflector curvature, achieving the majority of its increase when the normalized reflector depth d^* reached 1%.

One might expect the pressure amplitude to be within a constant of the quality factor as the normalized reflector depth d^* changed, but it was observed that the pressure amplitude at location A was additionally affected by spatial geometric distortion of the standing wave caused by reflector curvature. As the normalized reflector depth d^* increased, it was observed that the standing wave tended to concentrate on the axis of the transducer. This effect can be seen in the contour view in Fig. 5(b), where the standing wave has retreated from the edges of the transducer toward the axis.

There is evidence to believe that the quality factors extracted from the finite element analysis tends to be an overestimate. In a previous paper,[7] a numeric Fourier inversion computation of acoustic pressure for a channel containing a flat reflector identical to the $d^* = 0$ case in Fig. 7(b) pre-

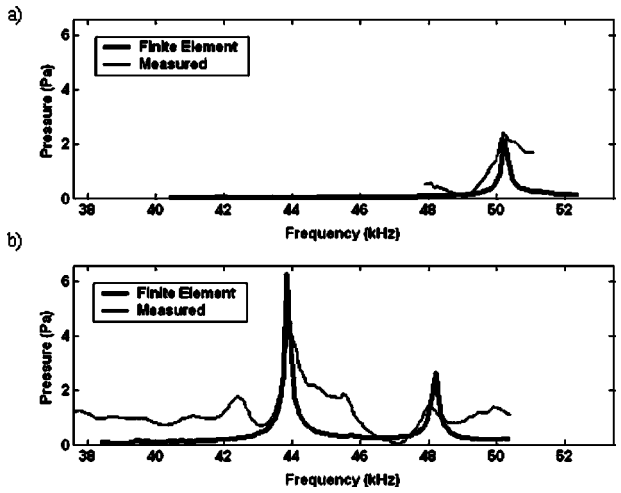


FIG. 8. Measured and predicted acoustic pressure amplitude in channel: (a) with flat reflector, (b) with curved reflector.

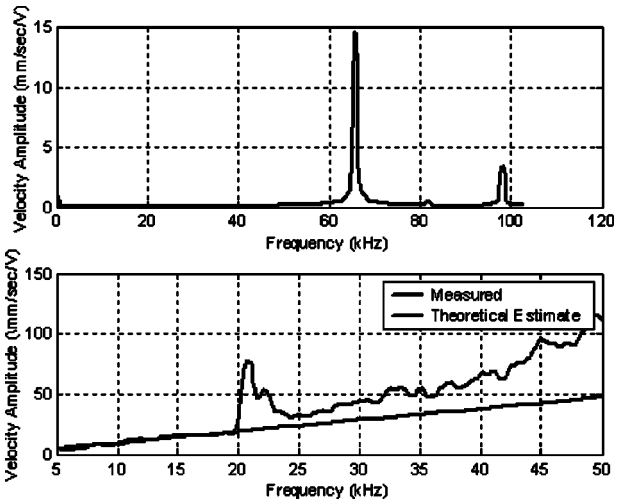


FIG. 9. Measured velocity amplitude of ultrasonic transducer. (a) Wide band view, (b) narrow band view.

dicted a quality factor of 226. This was lower than the quality factor of 249 extracted from the finite element computation at $d^*=0$ shown in Fig. 7(b). The corresponding measurement of quality factor from Ref. 7 was 185, very near the measured quality factor of 175 reported in this paper. From a physical point of view, another source of overestimation would be expected. As curvature is placed in a reflector, it would be expected to find fluid shear taking place. The model for absorption from Ref. 7 only includes thermal conduction at the reflector. For this reason, it was thought that the overestimate of quality factor would get worse with increasing normalized depth of curvature d^* when compared to experimental measurements.

A comparison of theoretically predicted and experimentally measured acoustic pressure spectrum for the channel described in Sec. III is shown in Fig. 8. The experimental measurements show the pressure amplitude spectrum caused by a white noise voltage source of amplitude $V=2.23$ V/frequency-line applied to the transducer when a flat or curved reflector was used. The theoretical predictions of acoustic pressure amplitude were provided by the finite element computations procedure described in this paper. The reflector geometry used in the theoretical predictions corresponded to the experimental apparatus. Computations were performed when using a flat and curved reflector. These computations used a velocity amplitude U for the piezoelectric transducer appropriate for the relevant depth resonance frequency.

A measurement of the velocity amplitude U of the piezoelectric transducer is shown in Fig. 9. Figure 9(a) shows a wide-band scale, which contains transverse resonances at 65.3 and 98.3 kHz. In Fig. 9(b), an expansion of the frequency scale from 0 to 55 kHz is shown. The solid line contains the measured velocity amplitude of the transducer, while the dashed line shows a low frequency theoretical prediction[6]

$$U=\frac{1}{2}\frac{\omega d_{33}}{2(d_{33}^2/\varepsilon_{33}^T)-1}V, \tag{5}$$

of the transducer velocity amplitude U based upon its previously mentioned electro-mechanical properties. The departure between the predicted performance near 25 kHz was presumed to be caused by the adjustable screw alignment mechanism behind the transducer, and the gradual departure at frequencies greater than 25 kHz was believed to be caused by the growth of the transverse mode. For interpretation of acoustic pressure measurements described in the following, the predicted performance with the dashed line was taken as the velocity amplitude of the piezoelectric transducer at a given frequency. It was reasoned that the motions of the transducer caused by the transverse mode would not contribute to the measured acoustic pressure at a depth resonance, $L\approx\lambda/2, L\approx\lambda$.

When the flat reflector was installed in the channel, numerical values extracted from Fig. 8(a) revealed that a depth resonance of $L\approx\lambda$ was observed at a frequency of 50.188 Hz, which for the size of the transducer corresponded to a source ka of $ka=13.2$. It was apparent that the experimentally measured spectrum was corrupted by the presence of length, width, and combination modes other than the depth modes predicted by the finite element analysis. Excluding the length modes that would be caused by reflection from the open ends of the channel, we calculated that there would be nine depth and width modes in the frequency range 38–53 kHz. In particular, we calculated a frequency of 50.5 kHz for the (1,2) mode, where the mode indices correspond to the depth and width dimensions, respectively. This mode presumably caused the resonance peak at 50.188 kHz to be asymmetric. The pressure amplitude at 50.188 Hz was 2.4 Pa, and a quality factor of 175 was estimated for the resonance peak. Estimation of the quality factor was performed using just the left side of the resonance peak, since the peak was corrupted on the right side by the extraneous mode. A finite-element prediction of the acoustic pressure amplitude based upon the velocity amplitude $U=105\ \mu$m/s of the transducer at 50.188 Hz is also shown in Fig. 8(a). The theoretical prediction showed a maximum pressure amplitude of 2.15 Pa at a frequency of 50.000 Hz and a quality factor of 249. The theoretically predicted and experimentally measured acoustic pressure amplitudes were in reasonable agree-

ment, while the experimentally measured quality factor was 30% smaller than the finite element prediction. We believe that the discrepancy was caused by the fact that the finite element computations tend to overestimate the quality factor.

When the curved reflector was installed in the channel, numerical values extracted from the data shown in Fig. 8(b) revealed that a λ depth resonance frequency of 43.904 kHz was observed. At this frequency, a pressure amplitude of 4.21 Pa and quality factor of 175 were observed. The finite element computation showed a λ depth resonance frequency of 43.855 kHz, and was based upon the measured velocity amplitude of $U=92$ μm/s for the piezoelectric transducer at this frequency. At 43.855 kHz, a pressure amplitude of 6.28 Pa and a quality factor of 437 were observed for the theoretically predicted spectrum. In this circumstance, the theoretically predicted pressure amplitude exceeded the measured pressure amplitude by approximately 50%, and the quality factors differed by a factor greater than 2. The difference in pressure amplitudes can be partially explained by the fact that the finite element computation tended to overestimate the quality factor. The overestimate in quality factor for a curved reflector would be greater than would be the case for a flat reflector, since the absorption model from Ref. 7 neglected fluid shear at the boundary. The resonance peak at a frequency of 43.904 kHz measured with the curved reflector was asymmetric, which likely indicated the presence of an extraneous mode, artificially lowering the quality factor estimate from the experimentally measured acoustic pressure spectrum.

The measurements do give a check on the gain that can be achieved in practice by using a slightly curved reflector. After compensating for frequency-dependent changes in transducer velocity amplitude, the finite element computations predicted that the acoustic pressure amplitude would change from 2.15 to 6.28 Pa when the flat reflector was changed to a curved one, a gain of 2.91 (9.28 dB). The experiments showed a corresponding gain in pressure amplitude from 2.4 to 4.21 Pa, a gain of 1.75 (4.86 dB) when using a curved reflector, somewhat less than would be theoretically expected but worth the effort if maximizing acoustic pressure amplitude in the channel is of importance.

VI. CONCLUSIONS

Finite element computations and experiments were used to determine the increase in amplitude of a standing acoustic wave propagating in a rectangular channel due to the use of a curved reflector. The determination applied to channel depths of one-half and one wavelength, at frequencies near 50 kHz, using air as the acoustic medium in the channel. Finite element computations predicted an increase in pressure amplitude ranging from 2 to 11 dB when a reflector with a curvature of 16% of the channel depth was used. An appreciable amount of this increase could be obtained with curvatures of much smaller depths. These figures apply to transducers sized to fall in the ka range 6.6–26.4. Experiments show that about one-half of the decibel increase in pressure gain can be achieved in practice.

ACKNOWLEDGMENTS

This material is based upon work supported by, or in part by, the U. S. Army Research Laboratory and the U. S. Army Research Office under Contract/Grant No. DAAD19-00-1-0150.

[1] M. Groschl, ''Ultrasonic separation of suspended particles. I. Fundamentals,'' Acustica **84**, 432–447 (1998).
[2] M. Groschl, ''Ultrasonic separation of suspended particles. II. Design and operation of separation devices,'' Acustica **84**, 632–642 (1998).
[3] M. Groschl, W. Burger, and B. Handl, ''Ultrasonic separation of suspended particles. III. Application in biotechnology,'' Acustica **84**, 815–822 (1998).
[4] M. J. Anderson, R. S. Budwig, K. S. Line, and J. Frankel, ''Use of acoustic radiation pressure to concentrate small particles in an air flow,'' 2002 IEEE Ultrasonics Symposium Proceedings, pp. 464–467.
[5] G. Kaduchak, D. N. Sinha, and D. C. Lizon, ''Novel cylindrical, air-coupled acoustic levitation devices,'' Rev. Sci. Instrum. **73**, 1332–1336 (2002).
[6] M. J. Anderson, R. S. Budwig, A. C. Cluff, E. C. Lemmon, and G. C. Putnam, ''The physics and technology of ultrasonic particle separation in air,'' Proceedings of the 2003 World Congress on Ultrasonics, pp. 1615–1621.
[7] M. J. Anderson, R. S. Budwig, and K. S. Line, ''The effect of large openings on cavity amplification at ultrasonic frequencies,'' J. Acoust. Soc. Am. **112**, 2771–2778 (2002).
[8] W. J. Xie and B. Wei, ''Dependence of acoustic levitation capabilities on geometric parameters,'' Phys. Rev. E **66**, 026605–1206605-11 (2002).
[9] R. Apfel and Y. Zheng, ''Characterization of acousto-electric cluster and array levitation and its application to evaporation,'' Proceedings of the Fifth Microgravity Fluid Physics and Transport Phenomena Conference, 2000, pp. 1766–1767.
[10] J. P. Berenger, ''A perfectly matched layer for the absorption of electromagnetic waves,'' J. Comput. Phys. **114**, 185–200 (1994).
[11] F. Q. Hu, ''Absorbing boundary conditions for linearized Euler equations by a perfectly matched layer,'' J. Comput. Phys. **129**, 201–219 (1996).
[12] L. J. Segerlind, *Applied Finite Element Analysis*, 2nd ed. (Wiley, New York, 1994).

The effect of flow-induced coupling on sound radiation from convected fluid loaded plates

Kenneth D. Frampton[a)]
Department of Mechanical Engineering, Vanderbilt University, VU Station 351592, Nashville, Tennessee 37235-1592

(Received 28 April 2004; revised 15 September 2004; accepted 6 December 2004)

Recent investigations concerning the effects of fluid convection on structural acoustic radiation have noted that radiation efficiency increases with increasing flow speeds. However, most of these studies based this conclusion on simulations that neglected flow-induced structural mode coupling. Yet, flow-induced coupling is know to have dramatic effects on structural dynamics including static and dynamic instabilities, and should therefore be expected to significantly affect sound radiation. The purpose of this investigation is to quantify the effects that fluid flow has on the sound radiated from rectangular vibrating plates when flow-induced structural modal coupling is considered. The discussion includes a description of the fundamental physics associated with a simply supported, vibrating, rectangular plate imbedded in an infinite baffle and radiating into a semi-infinite, convected fluid. This is followed by a discussion of the effect that flow-induced coupling has on the structural dynamic behavior. Finally, numerical results are presented that demonstrate the effect that such coupling has on the sound power radiated from a plate. [DOI: 10.1121/1.1852894]

PACS numbers: 43.40.Rj [EGW] Pages: 1129–1137

I. INTRODUCTION

Acoustic radiation from simple vibrating structures, such as plates, shells, and cylinders, into stationary fluids is well understood.[1–3] However, little has been done to date concerning structural acoustic radiation into a fluid that is in moving with a steady, constant motion. Of the few available references that do address radiation into a convected fluid only a very few include flow-induced structural modal coupling. Yet, such flow-induced coupling is well understood in the field of aeroelasticity[4] and is known to cause substantial changes in dynamic behavior culminating in static or dynamic instabilities. Such instabilities may not be of specific concern in structural acoustics. However, these flow-induced changes in dynamic response become significant well away from the stability boundaries and are important to structural acoustic analysis.

As stated, a few investigations have been published that address sound radiation from convected fluid-loaded plates. Work by Abrahams provided a solution to the convected wave equation using asymptotic techniques.[5] Atalla and Nicolas[6] calculated the radiation impedance for a piston in the presence of flow and noted that the radiation resistance increased and the radiation reactance decreased with increasing flow. Similar conclusions were drawn for a simply-supported plate vibrating in the (1,1) and (2,2) modes. Graham[7] noted a decrease in the critical frequency of plate modes in the presence of flow. Lastly, the author published analysis of the effect of convected fluid loading on radiation efficiency.[8] The basic conclusion was that as flow speed increases the radiation efficiency increases as well. The cause of this increase in radiation efficiency was found to be an effective increase in plate phase velocity for structural waves traveling upstream relative to the flow. Furthermore, the supersonic wave number region broadens as flow velocity increases, reaching an infinite width as the Mach number approaches 1. While each of the studies mentioned above noted increases in radiated sound as a result of flow, none of them considered the effects of flow-induced coupling.

A few publications are available that consider the effects of coupling. Sgard, Attalla, and Nicolas[9] used a finite element approach along with the extended Kirchhoff integral to calculate the power radiated from a vibrating plate into a convected fluid. They, too, noted an increase in radiated sound power with increasing flow speed. Wu and Maestrello[10] also included flow-induced coupling when calculating the radiated acoustic pressure. However, radiated sound power was not specifically discussed since the focus was on stability analysis.

The objective of this work is to quantify the sound power radiated from a rectangular, simply supported plate subject to a convected fluid on one surface and a vacuum on the other. Of particular importance to this effort is the inclusion of flow-induced structural modal coupling. The discussion begins with the dynamic modeling of the plate followed by a summary of the solution to the convected wave equation subject to the boundary conditions of a vibrating plate in an infinite baffle. This solution is the result of a combined transform/numerical technique that was detailed previously.[8] Next, the method for calculating the radiated sound power is described. Numerical results are presented that indicate the effect of convected fluid loading on the structural response and the effect on radiated sound power.

[a)]Assistant Professor of Mechanical Engineering, Associate Member of the Acoustical Society of America and corresponding author. Electronic mail: ken.frampton@vanderbilt.edu

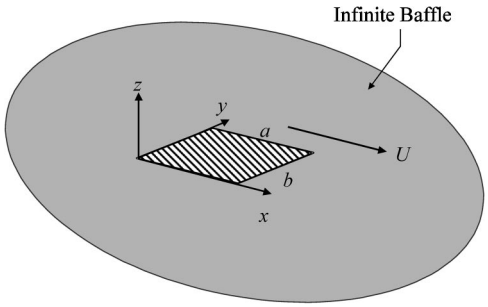


FIG. 1. Schematic of the plate in an infinite baffle.

II. THEORETICAL DEVELOPMENT

The development begins by considering a finite, elastic, rectangular plate in an infinite baffle as depicted in Fig. 1. On one side of the plate is a semi-infinite fluid, flowing parallel the plane of the plate in the positive x-axis direction. The other side of the plate is exposed to a vacuum. As the plate vibrates a pressure disturbance is created in the convected fluid. A mathematical model of this system begins with a modal description of the plate motion and the relationship of this motion to the dynamic behavior of the fluid. Next, the dynamics of the convected fluid are described by the convected wave equation. Given these coupled dynamic equations, a method for solving them is presented as is a method for computing the radiated sound power.

A. Plate dynamics

The plate is modeled by employing Galerkin's technique to discretize the linear equations of motion.[11] The partial differential equation of motion for a thin, uniform plate is

$$0 = D\nabla^4 w(x,y,t) + \rho_s h_s \frac{\partial^2 w(x,t)}{\partial t^2} + p(x,y,t) + p_d(x,y,t), \tag{1}$$

where $w(x,y,t)$, D, ρ, and h are the panel displacement, stiffness, density, and thickness, respectively. The plate is forced by the fluid and disturbance pressures, $p(x,y,t)$ and $p_d(x,y,t)$, respectively.

A separable solution is assumed using the *in vacuo* panel eigenfunctions and generalized coordinates of the form

$$w(x,y,t) = \sum_{n=1}^{N} \psi_n(x,y) q_n(t), \tag{2}$$

where Ψ_n is the mode shape and $q_n(t)$ is the generalized coordinate. The approach presented here can accommodate any plate boundary condition, however, results will be presented for a simply supported plate thus having the following mode shapes:

$$\psi_n(x,y) = \sin\left(\frac{j\pi}{a}x\right)\sin\left(\frac{k\pi}{b}y\right), \tag{3}$$

where the nth plate mode has (x,y) directional indices (j,k).

Substituting Eq. (2) into Eq. (1), multiplying by an arbitrary expansion function, $\Psi_m(x,y)$, and integrating over the domain yields a set of ordinary differential equations of the form

$$0 = M_n[\ddot{q}_n(t) + 2\xi\omega_n\dot{q}_n(t) + \omega_n^2 q_n(t)] + Q_n(t) + Q_n^d(t), \tag{4}$$

where M_n is the modal mass and ω_n is the modal frequency. Proportional structural damping has been included *ad hoc* at this point, as represented by the damping ratio ξ.[11] The disturbance generalized forces are

$$Q_n^d(t) = \int_0^b \int_0^a p_d(x,y,z=0,t)\psi_n(x,y)\,dx\,dy, \tag{5}$$

and the fluid generalized forces are defined by

$$Q_n(t) = \int_0^b \int_0^a p(x,y,z=0,t)\psi_n(x,y)\,dx\,dy, \tag{6}$$

where $p_a(x,y,z=0,t)$ is the fluid pressure acting on the plate due to plate motion and $Q_n(t)$ is the resulting generalized force. This generalized force is due entirely to the fluid response created by plate motion and is therefore the mechanism through which modal coupling is created.

Equation (4) actually represents N simultaneous ordinary differential equations, where N is the number of structural modes used in the solution. In order to facilitate integration of Eq. (4) with the fluid dynamics it is cast in state variable form such that[12]

$$\dot{\mathbf{x}}_s = \mathbf{A}_s\mathbf{x}_s + \mathbf{B}_s\mathbf{u}_s$$
$$\mathbf{y}_s = \mathbf{C}_s\mathbf{x}_s. \tag{7}$$

The $\mathbf{A}_s$ and $\mathbf{B}_s$ matrices are populated with the coefficients of Eq. (4), $\mathbf{C}_s$ is the identity matrix, and the state, input, and output vectors $\mathbf{x}_s$, $\mathbf{u}_s$, and $\mathbf{y}_s$ are defined as

$$\mathbf{x}_s = \{q_1(t)\ \cdots\ q_N(t)\ \dot{q}_1(t)\ \cdots\ \dot{q}_N(t)\}^T, \tag{8}$$

$$\mathbf{u}_s = \{Q_1(t)\ \cdots\ Q_1(t)\ Q_1^d(t)\ \cdots\ Q_1^d(t)\}^T, \tag{9}$$

and

$$\mathbf{y}_s = \mathbf{x}_s = \{q_1(t)\ \cdots\ q_N(t)\ \dot{q}_1(t)\ \cdots\ \dot{q}_N(t)\}^T. \tag{10}$$

The main challenge in this work is developing an expression for the fluid generalized force that is the next step toward studying the effects of flow on sound radiation.

B. Convected fluid dynamics

Previously, the author published results concerning the effect of a convected fluid on radiation efficiency while neglecting flow-induced coupling.[8] Since the solution approach used here is identical to that used previously, the development is summarized. Details can be obtained in Refs. 4 and 8.

The generalized forces of Eq. (6) are governed by the convected wave equation (or the linearized potential flow equations) describing the velocity potential of an inviscid, irrotational fluid flowing parallel to the x axis,

$$\nabla^2\phi - \frac{1}{c_0^2}\left[\frac{\partial^2\phi}{\partial t^2} + 2U\frac{\partial^2\phi}{\partial t\,\partial x} + U^2\frac{\partial^2\phi}{\partial x^2}\right] = 0, \tag{11}$$

and subject to the boundary conditions for a plate embedded in an infinite baffle,

$$\left.\frac{\partial\phi}{\partial z}\right|_{z=0} = \begin{cases} U\dfrac{\partial w}{\partial x} + \dfrac{\partial w}{\partial t} = f(x,y,z,t), & \text{on the plate,} \\ 0, & \text{off the plate,} \end{cases} \tag{12}$$

as well as a finiteness condition as z approaches infinity. Here, ϕ, c_0, and U are the velocity potential, speed of sound, and flow velocity, respectively. These equations are solved by a combination of the transform solution of Dowell[4] combined with the approximation employed by Frampton.[8] First, the convected wave equation [Eq. (11)] is transformed in space and time by means of a double Fourier transform with respect to the x and y spatial coordinates and a Laplace transform with respect to time according to the following definitions:

$$\tilde{\Phi}(\alpha,\gamma,z,s) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-i(\alpha x+\gamma y)}\Phi(x,y,z,s)\,dx\,dy, \tag{13}$$

$$\Phi(x,y,z,s) = \int_0^{\infty} e^{-st}\phi(x,y,z,t)\,dt. \tag{14}$$

The result is

$$\frac{d^2\tilde{\Phi}}{dz^2} = \tilde{\Phi}\mu^2, \tag{15}$$

where

$$\mu = \left[\frac{s^2}{c_0^2} + \frac{2Msi\alpha}{c_0} - \alpha^2(M^2-1) + \gamma^2\right]^{1/2}, \tag{16}$$

and where s is the Laplace variable, $M = U/c_0$ is the Mach number, and α and γ are the Fourier transform variables in x and y, respectively. The boundary condition of Eq. (12) is similarly transformed, resulting in

$$\frac{d\tilde{\Phi}}{dz} = \tilde{F}. \tag{17}$$

Next, Eq. (15) (the transformed version of the convected wave equation) is solved subject to Eq. (17) (the transformed boundary condition), resulting in

$$\tilde{\Phi}|_{z=0} = \frac{-\tilde{F}}{\mu}, \tag{18}$$

which is the transformed velocity potential on the plate surface. Now Eq. (18) is inverted with respect to the Laplace variable resulting in the velocity potential on the plate surface in the Fourier/time domain,

$$\tilde{\phi}(\alpha,\gamma,z=0,t) = -c_0\int_0^t \tilde{f}(\alpha,\gamma,\tau)\exp[-iMc_0\alpha(t-\tau)] \times J_0[(\alpha^2+\gamma^2)^{1/2}c_0(t-\tau)]\,d\tau, \tag{19}$$

where $J_k[\cdot]$ is a Bessel function of the first kind and of order k and τ is a dummy variable of integration. The goal of this calculation is the generalized force of Eq. (6) that requires the pressure on the plate surface. The pressure can be obtained from the velocity potential by means of Bernoulli's equation,

$$p = -\rho\left(\frac{\partial\Phi}{\partial t} + U\frac{\partial\Phi}{\partial x}\right), \tag{20}$$

where ρ is the fluid density. Substituting Eq. (19) into (20) yields

$$\tilde{p}(\alpha,\gamma,z=0,t) = \rho c_0\tilde{f}(t) - \rho c_0^2(\alpha^2+\gamma^2)^{1/2}\int_0^t \tilde{f}(\tau) \times\exp[-ic_0M\alpha(t-\tau)] \times J_1[c_0(\alpha^2+\gamma^2)^{1/2}(t-\tau)]\,d\tau. \tag{21}$$

Equation (21) is now substituted into Eq. (6) and the integration with respect to x and y is performed first. This is followed by the inverse Fourier integrals with respect to α and γ. The result is an expression for the fluid generalized forces on the nth panel mode due to motion of the mth panel mode, such that

$$Q_n(t) = \sum_{m=1}^{N}[q_m(t)S_{mn} + \dot{q}_m(t)D_{mn}] + \sum_{m=1}^{N}\left[\int_0^t q_m(\tau)H_{mn}(t-\tau)\,d\tau\right] + \sum_{m=1}^{N}\left[\int_0^t \dot{q}_m(\tau)I_{mn}(t-\tau)\,d\tau\right], \tag{22}$$

where

$$S_{mn} = \frac{1}{M}\int_0^a\int_0^b \frac{\partial\psi_m}{\partial x}\psi_n\,dx\,dy \tag{23}$$

$$D_{mn} = \frac{1}{MU}\int_0^a\int_0^b \psi_m\psi_n\,dx\,dy \tag{24}$$

$$H_{mn} = \frac{U}{4\pi^2M^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} G_{mn}i\alpha\sqrt{\alpha^2+\gamma^2}e^{-i\alpha Ut}J_1 \times[c_0t\sqrt{\alpha^2+\gamma^2}]\,d\alpha\,d\gamma \tag{25}$$

$$I_{mn} = \frac{1}{4\pi^2M^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} G_{mn}\sqrt{\alpha^2+\gamma^2}e^{-i\alpha Ut}J_1 \times[c_0t\sqrt{\alpha^2+\gamma^2}]\,d\alpha\,d\gamma, \tag{26}$$

$$G_{mn} = \int_0^a\int_0^b \psi_m(x,y)e^{-i(\alpha x+\gamma y)}\,dx\,dy \times\int_0^a\int_0^b \psi_n(x,y)e^{i(\alpha x+\gamma y)}\,dx\,dy. \tag{27}$$

The variables S_{mn} and D_{mn} are referred to as aerodynamic influence coefficients and are defined by the integrals in Eqs. (23) and (24), which can be performed analytically for most

plate eigenfunctions. Results of these integrals for simply supported plates are provided in Ref. 8. The aerodynamic influence functions, $H_{mn}(t)$ and $I_{mn}(t)$, are defined by integrals with no closed form solution. These may be performed numerically permitting a discrete time-marching solution to the system equations. Here, however, a solution is obtained by the means of a singular value decomposition technique. The result is a state variable representation of the fluid dynamics where the plate generalized coordinates and generalized velocities are the inputs and the fluid generalized forces are the outputs. This approach is not the most efficient means of obtaining a solution to the system equations; particularly in the case where the solution is limited to harmonic motion. However, it is convenient conceptually, since the radiated power can be expressed as a system transfer function (as shown subsequently).

C. Approximation of the fluid dynamic generalized forces

A singular value decomposition (SVD) technique is used to approximate the fluid generalized forces.[13] This technique is well suited to this problem since it requires system impulse responses from which a state variable realization is created. Note that the fluid dynamic response defined in Eq. (22) is characterized by influence functions, $H_{mn}(t)$ and $I_{mn}(t)$, which are essentially impulse responses relating the plate motion to the generalized force.

The objective is to create a state variable representation of the relationship between the plate motion and the fluid generalized forces, as expressed by Eq. (22). This mapping can be represented in state variable form as

$$\begin{aligned} \dot{\mathbf{x}}_f &= \mathbf{A}_f \mathbf{x}_f + \mathbf{B}_f \mathbf{u}_f, \\ \mathbf{y}_f &= \mathbf{C}_f \mathbf{x}_f + \mathbf{D}_f \mathbf{u}_f. \end{aligned} \tag{28}$$

The system $(\mathbf{A}_f, \mathbf{B}_f, \mathbf{C}_f, \mathbf{D}_f)$ has the plate generalized coordinates and their derivatives as inputs and provides the fluid dynamic generalized forces as outputs [per Eq. (22)] so that

$$\mathbf{u}_f = \{q_1(t) \ \cdots \ q_N(t) \ \dot{q}_1(t) \ \cdots \ \dot{q}_N(t)\}^T \tag{29}$$

and

$$\mathbf{y}_f = \{Q_1(t) \ \cdots \ Q_N(t)\}^T. \tag{30}$$

The fluid dynamic state variables are of mathematical construct and have no physical significance. The derivation details for Eq. (28) are not provided here for brevity. However, details can be found in Ref. 8. The primary advantage of using this SVD technique is that it dramatically reduces the computational requirements once the system model is in place. This is because, once the influence functions are numerically integrated, the equations are substantially overdetermined. The SVD technique allows the equations to be reduced in a numerically accurate manner.

D. System coupling and the radiated sound power

The plate and fluid systems in state variable form can be assembled into a coupled fluid/structure system, as shown in Fig. 2. Note that this fluid/structure coupling is created by the dependence of the fluid generalized forces on the plate motion. So, motion in the plate creates a response in the fluid, which in turn creates a force back on the plate. Furthermore, when cast in the form shown in Fig. 2, the input/output relationships for the system can be easily manipulated and transfer functions between inputs and system variables can be obtained. This feature of state variable modeling is important to for the calculation of radiated sound power.

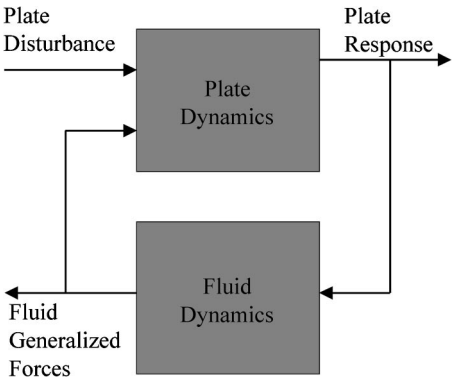


FIG. 2. Schematic of the coupled fluid/structure system.

The instantaneous sound power radiated from the plate can be expressed as[1]

$$\Pi(t) = \int p(x,y,z=0,t)\dot{w}(x,y,t)dS, \tag{31}$$

where $\dot{w}(x,y,t)$ is the velocity of the vibrating plate. By substituting the modal representation of Eq. (2) into Eq. (31), and making use of Eq. (6), the radiated power can then be expressed as

$$\begin{aligned} \Pi(t) &= \int_S p(x,y,z=0,t) \sum_{n=1}^{N} \psi_n(x,y)\dot{q}_n(t)dS, \\ &= \sum_{n=1}^{N} \dot{q}_n(t) \int_S p(x,y,z=0,t)\psi_n(x,y)dS, \\ &= \sum_{n=1}^{N} \dot{q}_n(t) Q_n(t). \end{aligned} \tag{32}$$

Now, the time-averaged radiated power can be expressed as

$$\langle \Pi \rangle = \frac{1}{T} \int_0^T \sum_{n=1}^{N} \dot{q}_n(t) Q_n(t) dt, \tag{33}$$

where T is the period. If harmonic motion is now assumed, this expression becomes

$$\begin{aligned} \langle \Pi \rangle &= \sum_{n=1}^{N} \frac{1}{T} \int_0^{2\pi/\omega} \mathbb{R}\{\tilde{Q}_n e^{i\omega t}\} \mathbb{R}\{\tilde{q}_n e^{i\omega t}\} dt \\ &= \frac{1}{2} \sum_{n=1}^{N} \mathbb{R}\{\tilde{Q}_n \tilde{q}_n^*\}, \end{aligned} \tag{34}$$

where $\tilde{\dot{q}}_n$ and $\tilde{Q}_n$ are the complex magnitudes of the nth modal velocity and nth fluid generalized force respectively; the asterisk superscript indicates the complex conjugate; $\mathbb{R}$ indicates the real component of the argument; and ω is the excitation frequency. Now, making use of system transfer

function relationships, the time-averaged radiated sound power can be expressed as

$$\langle \Pi \rangle = \frac{1}{2} \sum_{n=1}^{N} \mathbb{R}\{I_n(\omega) H_n^*(\omega)\} |d(\omega)|^2, \tag{35}$$

where $I_n(\omega)$ is the system transfer function between the disturbance input, $d(\omega)$, and the fluid generalized force acting on the nth plate mode. The transfer function $H_n(\omega)$ is the system transfer function between the disturbance input and the nth modal velocity. Given the coupled system of Fig. 2, these transfer functions can be readily obtained. So the radiated sound power can be calculated by finding the transfer functions of the coupled system, specifying the disturbance spectrum, and carrying out the calculation of Eq. (35).

III. COMPUTATIONAL RESULTS

A computational investigation was undertaken based on the theory presented in Sec. II. The results presented focus on two areas: the effects of flow-induced coupling on the structural response and on the radiated sound power. The physical parameters used were for an aluminum plate (density of 2780 kg/m^3, thickness of 4 mm, length of 1 m, width of 0.8 m, modulus of elasticity of 72.4 GPa, and a damping ratio of 0.02) exposed to air at sea level (density of 1.16 kg/m^3, speed of sound of 350 m/s). The accuracy of this modeling approach was established in Ref. 8, where results for the no-flow case were compared to previously published works. In order to ensure accuracy in these results, the plate model of Eq. (7) included 20 modes while the fluid dynamic system of Eq. (28) included 600 degrees of freedom. With this configuration, the plate response and radiated sound power demonstrated sufficient convergence over the first 15 plate modes.

A. Flow-induced coupling of structural modes

The effect of flow on plate dynamics is well understood in the field of aeroelasticity.[4] However, this branch of engineering is typically concerned with coupling-induced instabilities: static divergence in subsonic flows and dynamic flutter in supersonic flows. These instabilities are not usually of interest in structural acoustics since these kinds of structures are designed to avoid them. However, the effects on plate dynamics can occur well away from instability points, therefore affecting the structural acoustic response. In structural acoustics literature Sgard, Attalla, and Nicolas[9] quantified modal coupling with a novel approach based on the coherence between modes. The author avoided the specifics of coupling previously by considering the total radiation efficiency (including all modes simultaneously.[14]

In order to quantify coupling, it is important to clarify the language used to describe the structural response. The term "*in vacuo* mode" will be used to describe the plate eigenfunctions of Eq. (3). (The *in vacuo* modes could be more accurately called admissible functions[11] since, in the fully coupled system, they are really just mathematically convenient functions with which to solve the differential equation.) The modes of the fully coupled system shown in Fig. 2 will be referred to as *coupled modes*. These are the structural mode shapes that result from solving the eigenvalue problem of the coupled fluid/structure system. Each coupled mode includes degrees of freedom associated with the plate and with the fluid. It should also be noted that the resulting coupled system eigenfrequencies will be complex. This implies that the *in vacuo* modes contributing to each coupled mode need not be in phase.

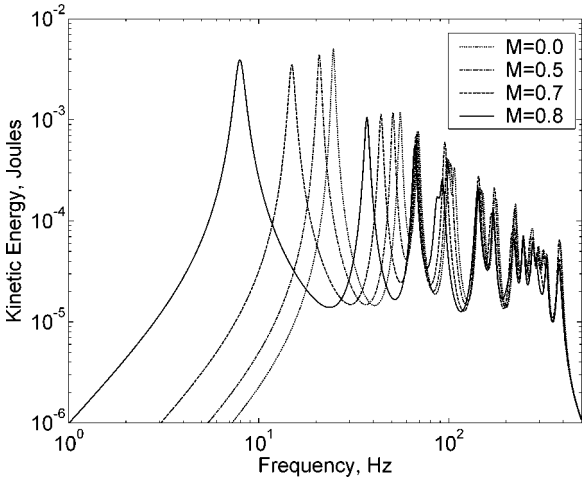


FIG. 3. Plate kinetic energy as a function of frequency for various flow speeds.

The effect of flow on the structural response is demonstrated in Fig. 3, which shows the kinetic energy of the plate for various flow velocities. In this analysis, as well as all subsequent analyses, the plate is acted upon by a disturbance input that produces generalized forces of equal magnitude and uniform spectrum. Although this is not a physically realistic disturbance, it does represent a "worst case" input. As shown in Fig. 3, the effects of flow are greatest on the first two coupled modes. The resonant frequency of these modes changes considerably as the Mach number increases. The fundamental coupled mode has a resonant frequency of about 25 Hz when $M=0.0$, which decreases to about 8 Hz when $M=0.8$. If the flow speed were increased further, the plate would become statically unstable (diverge) at $M=0.93$. This behavior is well understood and has received considerable attention in the aeroelasticity literature.[4] The second coupled mode experiences a shift in frequency, beginning at 57 Hz and decreasing to 38 Hz. A further effect to note in Fig. 3 is that, while the resonant frequencies of the coupled modes change with changing flow speed, the amount of energy in each coupled mode remains nearly constant.

The effects of flow-induced coupling are further highlighted in Fig. 4, which shows the plate kinetic energy at Mach numbers of 0, 0.5 and 0.8. Shown are the kinetic energies for the fully coupled model, a model that includes only intramodal coupling and a model with no coupling. The fully coupled model includes the aerodynamics forces on all modes created by the motion of all other modes [i.e., the full implementation of Eq. (22)]. The model including only intramodal coupling uses aerodynamic generalized forces that only contain the effects of each mode on itself [i.e., Q_n of Eq. (22) is limited to the case where $m=n$ only]. The uncoupled model was created by disconnecting the feedback

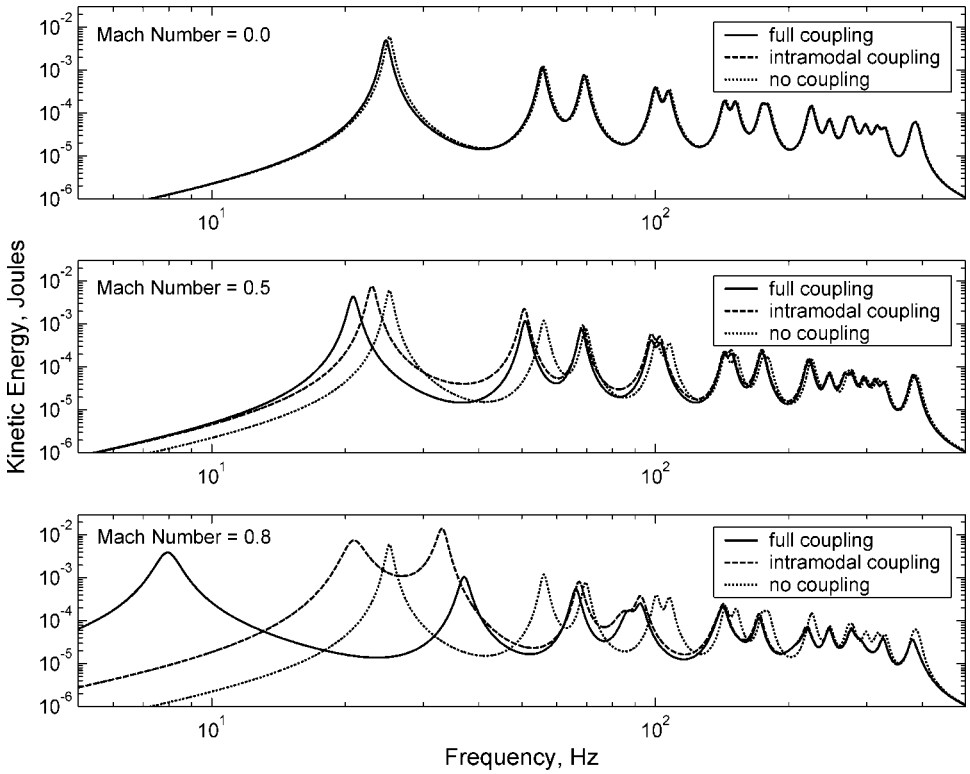


FIG. 4. A comparison of plate kinetic energy with full intermodal coupling, with intramodal coupling and without coupling for various flow speeds.

mechanism of Fig. 2, which is equivalent to setting the fluid generalized forces acting on the plate to zero. Note that when coupling is not included in the model, the plate kinetic energy does not change as flow conditions change. However, when full intermodal coupling is included, significant changes in the plate dynamics occur. Finally, the model limited to intramodal coupling accurately captures the effects of coupling for very low flow velocities. This is demonstrated for the $M=0.0$ case in Fig. 4, where the fully coupled curve and the intramodal coupling curves are indistinguishable. But, as demonstrated by the case in Fig. 4 with $M=0.5$, limiting the model to only intramodal coupling is inaccurate around the first mode. These inaccuracies only increase as the flow velocity increases.

While the changes in resonant frequencies are important, a more interesting effect of flow is in the modal coupling. It is common in structural acoustics to quantify radiation from individual *in vacuo* modes. However, as the flow speed increases, the structural mode shapes of the coupled system become combinations of these *in vacuo* modes. This is demonstrated in Fig. 5, which shows the contributions of the *in vacuo* modes to the first six coupled modes when the Mach number is 0.8. Note that each of these coupled modes has significant contributions from more than one other *in vacuo* mode. However, the coupled modes are dominated by one *in vacuo* mode.

As stated previously, these modal coupling effects are well understood in the aeroelasticity community. In this community the nondimensional dynamic pressure, $\lambda = \rho U^2 a^3/D$, is known to be a better indicator of coupling than the Mach number.[4] A reasonable rule of thumb is that intermodal coupling will become significant when $\lambda \cong 50$ for simply supported rectangular plates. Divergent instabilities will occur when $\lambda \cong 200$. These rules apply to the plate investigated in this study when $M=0.4$ and $M=0.93$, respectively.

B. Effects of flow on radiated sound power

The overall effect of flow on the radiated sound power is shown in Fig. 6, which depicts the total radiated sound power as a function of frequency for a range of Mach numbers. As with the kinetic energy, the most notable effect is the shift in resonant frequency of the lower modes. The lowest coupled mode, while shifting frequency considerably, does not exhibit a significant change in radiated power. However, all of the coupled modes above the fundamental show significant increases in radiated power. This increase in radiated power is not due to increases in coupled mode vibrational amplitudes. This is demonstrated by the fact that the plate kinetic energy remains nearly constant as the flow speed increases. The increase in radiated power is due to two other effects: increases in radiation efficiency of *in vacuo* shapes and the redistribution of energy due to modal coupling.

Increases in radiation efficiency of *in vacuo* modes without coupling were dealt with previously by the author.[8] The cause of this increase in radiation efficiency was found to be an effective increase in plate phase velocity for plate waves traveling upstream relative to the flow. Furthermore, the supersonic wave number region broadens as the flow velocity increases, reaching an infinite width as Mach number approaches 1. In the simplified case of a 1-dimensional plate the supersonic wave number region is defined as[8]

$$\frac{k}{M-1} < k_x < \frac{k}{M+1}. \tag{36}$$

As demonstrated in Eq. (36), the supersonic wave number region shifts and expands as the Mach number increases. The

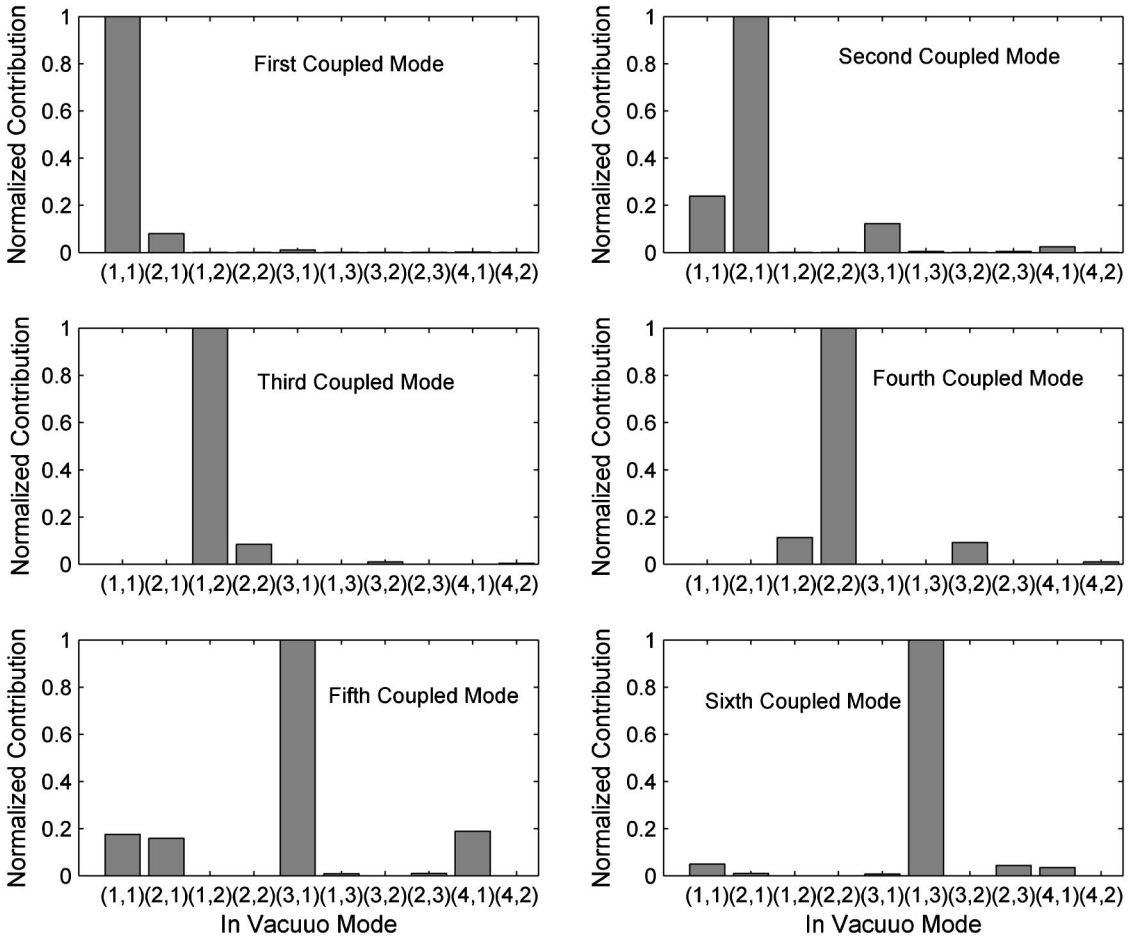


FIG. 5. The contributions of *in vacuo* modes to the first six coupled structural modes when the Mach number is 0.8.

overall effect is that, for a fixed wave number spectrum in the plate response, as the flow velocity increases more of the plate wave number spectrum is enveloped by the supersonic region, thus resulting in increased radiation.

Of specific interest in this work is the effect on radiated sound power due to the redistribution of energy through flow-induced modal coupling. As noted before in Fig. 5, when the flow velocity increases the coupled modes become combinations of the *in vacuo* shapes. The path through which energy moves from one coupled structural mode to another is the fluid. This conveyance of energy between modes is demonstrated in Figs. 7 and 8, which show the real and imaginary sound power due to motion of a single coupled mode when all modes are excited [i.e., one instance of Eq. (35)]. Figure 7 is for the first coupled mode and Fig. 8 is for the second coupled mode when $M=0.8$. Note that negative power (energy entering the plate) is indicated by a dotted line while positive power (energy exiting the plate) is shown by a solid line. As shown in Figs. 7 and 8, energy enters the plate at some frequencies and exits the plate at other frequencies. Energy tends to enter the plate at or near frequencies associated with *in vacuo* modes that are secondary contributors to the coupled mode in question. For example, the second coupled mode radiation shown in Fig. 8 has energy entering the plate at frequencies between the fundamental coupled mode up to the second coupled mode. Energy also enters the plate at frequencies in the 90–100 Hz range, which is near the coupled mode with a dominant (3,1) *in vacuo* component as well as the 250–300 Hz range, which is near the coupled mode dominated by the (4,1) *in vacuo* mode.

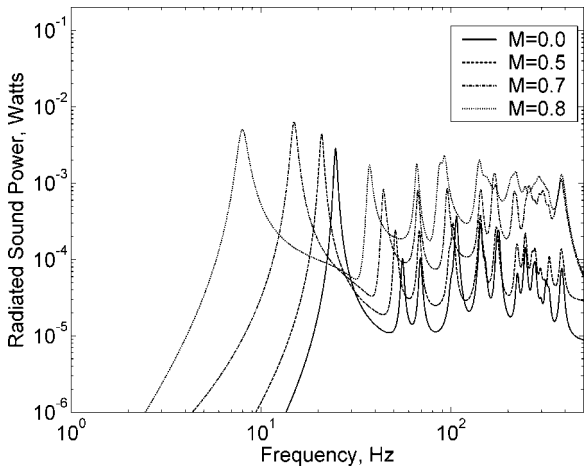


FIG. 6. Radiated sound power as a function of frequency for various flow speeds.

Clearly, the exchange of energy among modes that occurs in the presence of modal coupling alters not only the dynamic response of the plate, but the radiated sound as well. The extent to which coupling affects radiation is demonstrated in Fig. 9, which shows the radiated power with full intermodal coupling, with intramodal coupling only, and

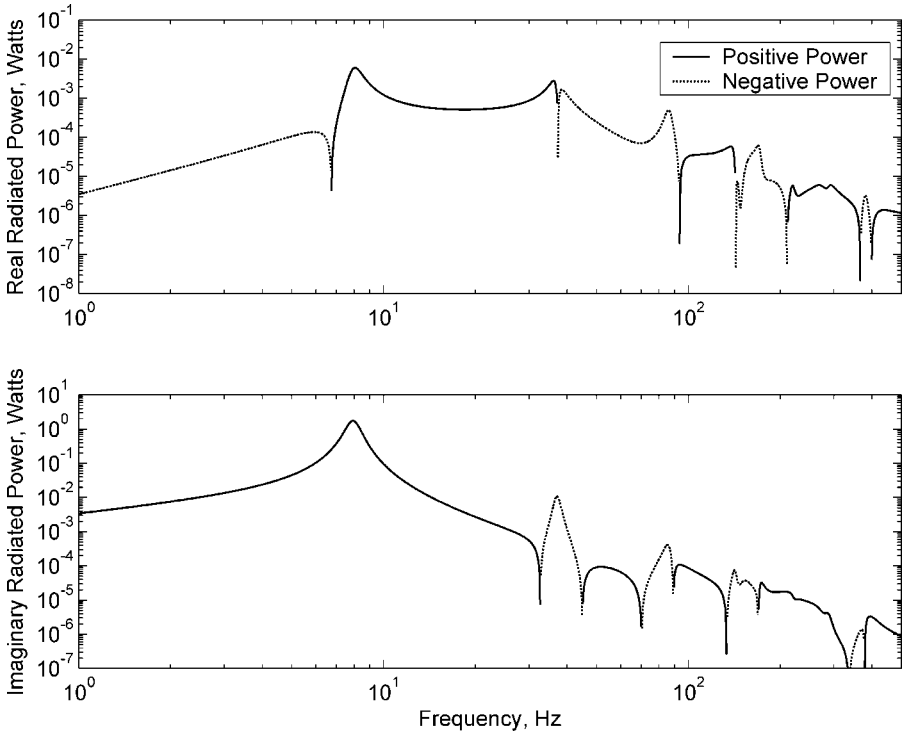


FIG. 7. Real and imaginary sound power radiated by the first coupled mode when all modes are subject to the disturbance when $M=0.8$.

without coupling and for various Mach numbers. As can be seen in Fig. 9, the effects of coupling on radiated power are significant when the flow speed is greater than zero. As noted earlier, limiting the model to intramodal coupling is quite inaccurate. Also noted previously, intermodal coupling becomes significant when $\lambda \cong 50$ or when $M \cong 0.4$ in this example. As can be seen in Fig. 9 for $M=0.5$ the predicted radiated power differs from the uncoupled case by a small, but significant, amount. The difference at higher flow speeds is much more significant.

IV. CONCLUSIONS

The radiation effect of flow-induced coupling on the radiated sound power from plates has been presented. The modeling of the plate and fluid has been summarized and the effects of flow-induced coupling on the plate response have been described. The most notable effect is that the resonant frequencies of the lowest plate modes shift downward as flow velocities increase. However, the kinetic energy in the plate remains fairly constant. It was further noted that the

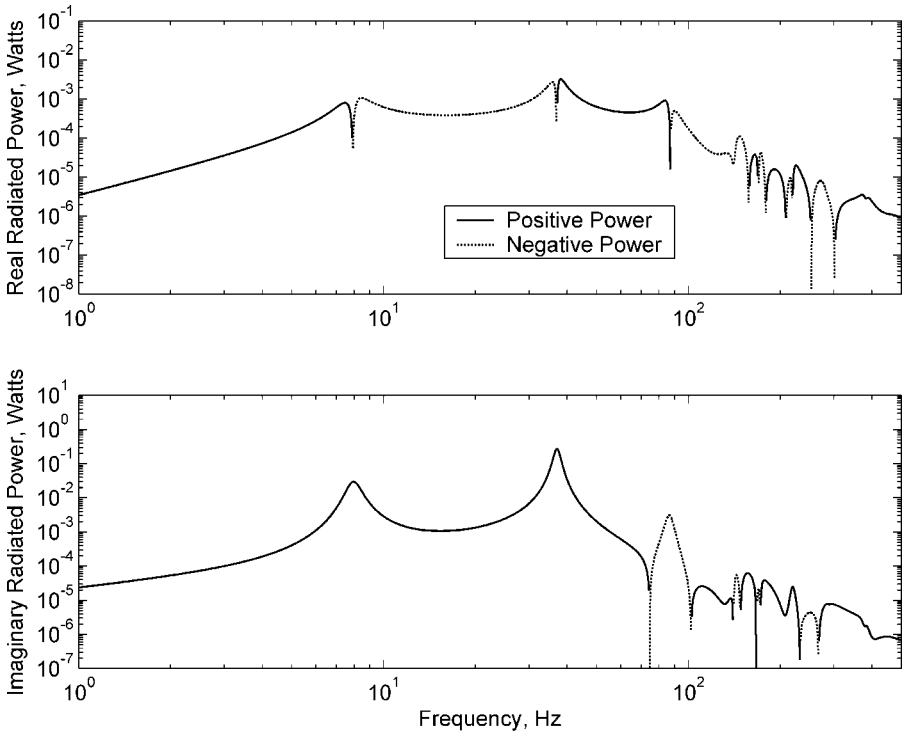


FIG. 8. Real and imaginary sound power radiated by the second coupled mode when all modes are subject to the disturbance when $M=0.8$.

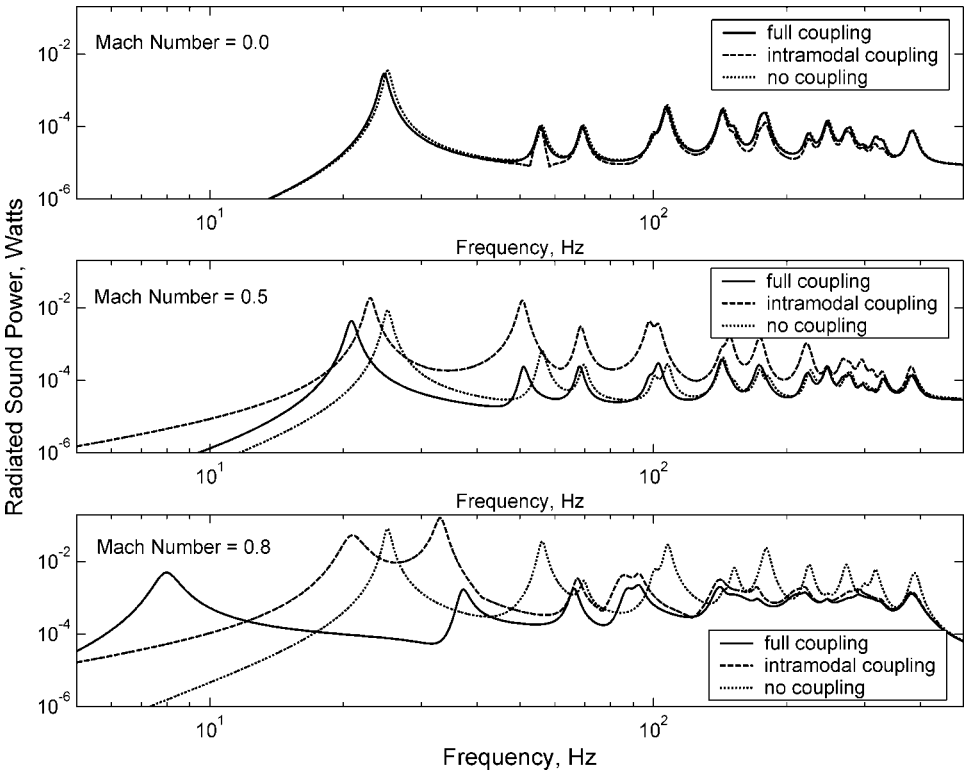


FIG. 9. A comparison of radiated sound power with full intermodal coupling, with intramodal coupling and without coupling for various flow speeds.

radiated sound power increases as the flow velocity increases. This was shown to be due to two factors. First was the effect of increased radiation efficiency due to an effective decrease in the critical frequency. Second, the flow-induced coupling was shown to allow energy to be redistributed among plate modes, resulting in an increase in radiated sound power.

[1] F. Fahy, *Sound and Structural Vibration* (Academic, New York, 1985).
[2] M. C. Junger and D. Feit, *Sound, Structures and Their Interaction*, 2nd ed. (MIT Press, Cambridge, MA, 1986).
[3] M. S. Howe, *Acoustics of Fluid–Structure Interactions* (Cambridge University Press, Cambridge, UK, 1998).
[4] E. H. Dowell, *Aeroelasticity of Plates and Shells* (Noordhoff International Publishing, The Netherlands, 1975).
[5] I. D. Abrahams, ''Scattering of sound by an elastic plate with flow,'' J. Sound Vib. **89**, 213–231 (1983).
[6] N. Atalla and J. Nicolas, ''A formulation for mean flow effects on sound radiation from rectangular baffled plates with arbitrary boundary conditions,'' J. Vibr. Acoust. **117**, 22–29 (1995).
[7] W. R. Graham, ''The effect of mean flow on the radiation efficiency of rectangular plates,'' Proc. R. Soc. London **454**, 111–137 (1998).
[8] K. D. Frampton, ''Radiation efficiency of convected fluid loaded plates,'' J. Acoust. Soc. Am. **113**, 2663–2673 (2003).
[9] F. Sgard, N. Atalla, and J. Nicolas, ''Coupled FEM-BEM approach for mean flow effects on vibro-acoustic behavior of planar structures,'' AIAA J. **32**, 2351–2358 (1994).
[10] S. F. Wu and L. Maestrello, ''Responses of finite baffled plate to turbulent flow excitations,'' AIAA J. **33**, 13–19 (1995).
[11] L. Meirovitch, *Analytical Methods in Vibrations* (Macmillan, New York, 1967).
[12] R. Clark, W. Saunders, and G. Gibbs, *Adaptive Structures: Dynamics and Controls* (Wiley, New York, 1998).
[13] S. Kung, ''A new identification and model reduction algorithm via singular value decomposition,'' *IEEE 12th Asilomar Conference on Circuits, Systems and Computers*, 1978, pp. 705–714.
[14] K. D. Frampton, ''Radiation efficiency of convected fluid loaded plates,'' *Proceedings of the Symposium on Flow-Induced Vibration and Noise of Thin Materials*, IMECE (ASME, Nashville, TN, 1999).

Full-scale measurements for noise transmission in tunnels

K. M. Li[a] and K. K. Iu
Department of Mechanical Engineering, The Hong Kong Polytechnic University, Hung Hom, Hong Kong

(Received 21 February 2004; revised 27 December 2004; accepted 27 December 2004)

In many previous studies, energy-based methods are used to predict the attenuation of sound in long tunnels. However, these models do not address the interference effects of the sound fields generated by all image sources. A numerical model has been developed, in which the total sound field is computed by summing contributions from all image sources coherently. This numerical model also incorporates a correction term for calculating the atmospheric absorption of sound in air. To validate the numerical models in practical situations, two road traffic tunnels have been chosen for extensive measurements. The levels of the transmitted noise have been recorded in one-third octave band frequencies at various separations up to a maximum of 400 m. The predictions using the coherent model agree reasonably well with the measured data at all frequencies. The agreements between the field data and the theoretical predictions using the energy-based model are tolerable at high frequency, but less so at low frequency. In most cases, the predictions of the coherent model give the best results, with an accuracy to within 3 dB. On the other hand, the energy-based models are not able to predict the peaks and dips across the frequency spectra, the variance with the measurement results being up to 7 dB at low-frequency bands. [DOI: 10.1121/1.1859251]

PACS numbers: 43.50.Rq, 43.55.Gx, 43.55.Ka Pages: 1138–1145

I. INTRODUCTION

The current study is motivated by the need for a better understanding of sound propagation in tunnels. This can provide useful information for improving the acoustic environments of road traffic tunnels. Despite considerable improvements in computer performance and the accelerating development of many numerical achemes, full-scale measurements are still popular for studying sound fields in long ducts, such as corridors and tunnels. The present work is a continuation of our earlier study[1] that investigates the theoretical models for predicting the propagation of sound in long enclosures. A comprehensive literature review of the problem can be found in Ref. 1 and its cited references. In this paper, we wish to show the robustness of the complex image source model[2] (also known as the coherent model) for the prediction of sound fields in long ducts with geometrically reflecting boundaries. Extensive full-scale field measurements in two tunnels for road traffic are conducted and to the results compared with numerical predictions. We endeavor to study the importance of the interference effects caused by the direct and reflected waves for the accurate prediction of sound fields in long tunnels. To demonstrate that the coherent model is valid for a long propagation distance, measurements are taken in the tunnels up to a distance of several hundred meters from the source.

The problems are simplified, somewhat, by modeling the propagation of sound in long tunnels as an analogous situation to the determination of sound fields in a long enclosure bounded by two parallel vertical inwalls of infinite extent and two parallel horizontal planes. Although impedance boundary conditions can be implemented onto all vertical and horizontal planes in our numerical model, the analyses here are restricted to perfectly reflecting walls, ceiling, and ground. This simplification is justifiable because the boundary walls of most tunnels are constructed with acoustically hard materials. Two different tunnels, which are normally used for road traffic, are chosen for field measurements in the present study. The experimental results will be used to compare with different numerical models.

The structure of the paper is as follows. In Sec. II we outline the numerical models used in the present study. The effect of atmospheric absorption of sound has been incorporated in the numerical analyzes. The issue on the number of image sources required for the calculation has also been addressed. In Sec. III, we describe experimental studies for noise transmitted in two tunnels. Experimental results are compared with numerical predictions. Finally, the outcomes of the current study are summarized in Sec. IV.

II. NUMERICAL SIMULATIONS

A. The effect of atmospheric absorption of sound in air on the prediction models

The attenuation of sound due to atmospheric absorption is well known and has been studied for decades.[3] A practical method for calculating the attenuation of sound in air is to multiply an exponential term, $\exp(-ad)$, to account for the reduction of sound energy caused by air absorption. This term is also known as the air absorption factor where a is the attenuation of sound in air and d is the distance of the receiver from the source.

Recently, the Acoustical Society of America has published an American National Standard, ANSI S1.26-1995,[4] to provide a means of calculating the atmospheric attenuation of sound in air from any source, moving or stationary, for a wide range of meteorological conditions. In view of a recent

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

study,[1] the modified expressions for sound propagation in a long tunnel with geometrically reflecting boundaries may be written as

$$P(N)=\frac{1}{4\pi}\sum_{N}\frac{\exp(ik'd_N)}{d_N}, \tag{1}$$

where $k'(=k+0.115ia)$ is the modified wave number and a is the coefficient of sound attenuation that has units of $\mathrm{dB\,m^{-1}}$. The real part of the modified wave number, k $(=\omega/c)$, is the ratio of the angular frequency of the source to the speed of sound in air. The imaginary part of the modified wave number accounts for the attenuation of sound in air due to the effect of atmospheric absorption. The total sound field is obtained by summing contributions from all image sources to the receiver with d_N as the distance from the Nth image source to the receiver. The term N may also be regarded as the order reflections from the boundary surfaces. We refer to the formulation given in Eq. (1) as the coherent model. It is noteworthy that the use of Eq. (1) implies that specularly reflections occur at the boundaries of the enclosure. This assumption is justifiable because there are usually few diffusely reflecting surfaces employed in the construction of road traffic tunnels. The use of a model for diffuse reflection of boundary surfaces will not be considered in the present study.

The third formulation used in our study is a simple energy-based model developed by the Acoustical Society of Japan published.[1,5] The formula, which may be found in Eq. (9) of Ref. 1, is referred to as the ASJ model. The attenuation of sound due to atmospheric absorption can easily be incorporated in the ASJ model by multiplying the existing formula with the same exponential factor, $\exp(-ad)$, as discussed earlier in the incoherent model.

B. Noise reduction

The levels of sound energy are usually calculated in the incoherent and ASJ models, but the sound pressure levels are predicted by the coherent model. For ease of comparison of different theoretical models and experimental data, we use a term known as the noise reduction (NR), which is defined as the ratio of the total sound field, $P(N)$, measured at various receiver locations to the total field, $P_{10}(N)$, measured at 10 m in front of the source:

$$\mathrm{NR}=20\log[P(N)/P_{10}(N)]. \tag{2}$$

For the coherent model, the total sound pressure levels can be calculated from Eq. (1). On the other hand, the sound energy is usually used in the incoherent and ASJ models. We may modify Eq. (2) to obtain the corresponding noise reduction spectrum as follows:

$$\mathrm{NR}=10\log(I/I_{10}), \tag{3}$$

where I is the total sound energy at various receiver locations and I_{10} is the total sound energy at 10 m from the source. We note that 10 m is chosen as a reference point instead of 1 or 5 m, as used in other studies.[1,6,7] The choice of 10 m ensures that all measurements are conducted outside a hydrodynamic near-field region[8] for a frequency range down to 50 Hz of which the wavelength is 6.8 m. Consequently, the loudspeaker used in our experimental studies can be assumed to be a point source.

C. Number of image sources

The use of ray models to compute the sound field requires the summation of a series of infinite terms. Each ''virtual'' image source is represented by a term from the series. The contributions of the higher-order image sources, which become weaker in strength when they are located farther apart from the receiver, can be neglected. As a result, only a finite number of image sources contribute significantly to the total sound pressure levels. Dance and Shield[9] described a computer-based image-source model for the prediction of sound distribution in a nondiffuse fitted enclosed space, such as a factory. In their study, the number of ''allowable'' reflections was determined by using a percentage of energy discontinuity. The percentage is set at a level such that the contribution from a given image source is negligibly small relative to the overall sound level. The relationship between the order of reflection, N, and the energy discontinuity percentage, E_p, is defined by

$$N=\frac{\ln(1-E_p/100)}{\ln(1-\alpha)-al}, \tag{4}$$

where α is the average absorption coefficient of the boundaries, a is the absorption attenuation of air, and l is the mean pathlength from the image sources to the receiver. A typical value of E_p, which allows accurate predictions of the sound field in a range of enclosed spaces, lies between 90% and 99%.[10] Obviously, a smaller value E_p leads to a shorter computing time for ray tracing models because the allowable reflection order, N, is smaller.

When the coherent model is used, we anticipate that a finite number of terms are required in the prediction of the sound field in long enclosures. However, Eq. (4) cannot be used for determining the number of the required terms in the ray series. This is because the coherent model is not an energy-based scheme. In our study, the relative error at the lowest one-third octave band of interest in the total sound field is used to determine the number of required reflections. Here, we define the relative error in total sound field, ΔP, by

$$\Delta P=\left|\frac{P(m+n+2)-P(m+n)}{P(m+n)}\right|, \tag{5}$$

where $P(N)$ given in Eq. (1). It is the total sound field at a particular receiver location due to an image source with N as the order of reflection from the boundary surfaces. The order of reflection can be split into m and n where they are the respective orders of reflections at the vertical and horizontal boundaries. The error in truncating the ray series to Nth order of reflections is bounded by ΔP if N is sufficiently large. The smaller the relative error, the more the terms (i.e., more computational time) are required in the ray series given in Eq. (1).

Figures 1 and 2 show the predicted relative errors for the propagation of sound in tunnels with different conditions. In Fig. 1, we display the numerical simulations for a tunnel that

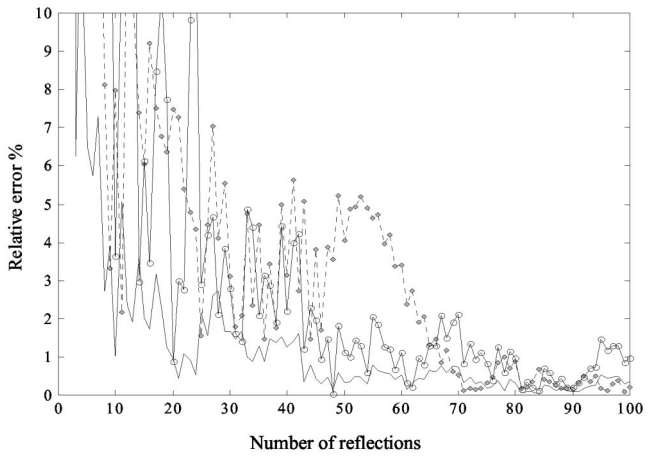


FIG. 1. Comparisons of predicted relative errors versus numbers of reflections at 50 Hz for sound propagation in a realistic tunnel with a cross-sectional area of 12.5 m×5.8 m. Three source/receiver separations are shown. (Dashed line with diamonds: 400 m; solid line with diamonds: 50 m; solid line: 10 m.)

has a rectangular cross-section of 12.5 m×5.8 m. Predictions of the relative error, ΔP, is calculated for a receiver locating at different horizontal distances of 10, 50, and 400 m from the source. In the simulations, the source and receiver are situated at 0.9 and 1.3 m above the ground, respectively. Both the source and receiver are placed at a distance of 2.65 m from one of the vertical walls. Figure 1(a) shows that the direct distance between the source and receiver affects the required order of reflection, i.e., the number of term in the ray series. More image sources are generally required for greater separations between the source and receiver separation.

Next, we show that the size of the tunnel also affects the number of terms required for the ray series. This can be achieved by comparing ΔP for two tunnels with different cross-sectional areas. The first tunnel (Western Cross Harbor Tunnel) has a cross-sectional area of 12.5 m×5.8 m, which has the same dimension as the tunnel used in Fig. 1. The second tunnel (Tai Lam Tunnel), which has a dimension of

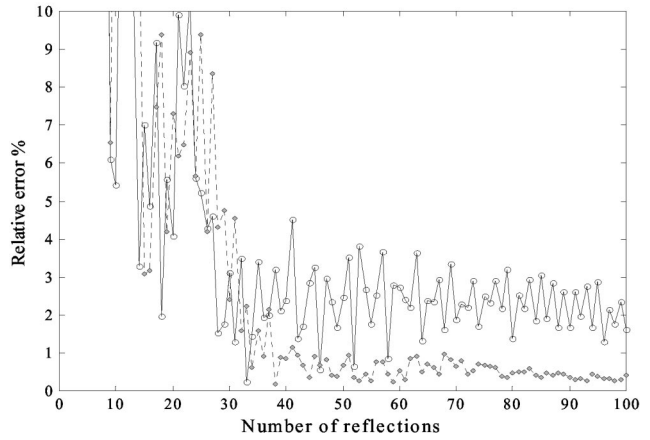


FIG. 2. Comparisons of predicted relative errors versus numbers of reflections at 50 Hz for sound propagation in two realistic tunnels with the source/receiver separation of 200 m. Two cross-sectional areas of the tunnel are shown. (Solid line with diamonds: 12.5 m×5.8 m; dashed line with diamonds: 14.2 m×6.0 m.)

14.2 m×6.0 m, is larger than Western Cross Harbor Tunnel. The choice of dimensions used in numerical simulations reflects the actual sizes of the two tunnels employed for field measurements. More details about the tunnels will be described in Sec. III.

In the numerical simulations, the source is placed at the centerline and at 1 m above the ground in Tai Lam Tunnel. The receiver, which is located at a horizontal distance of 200 m from the source, is placed at an offset position of 3.5 m from one of the vertical wall and at 2.33 m above the ground. For Western Cross Harbor Tunnel, the source is placed at 6.25 m from one of the vertical walls and 0.9 m above the ground. The receiver is placed at 2.65 m from the same vertical wall and at 2.38 m above the ground. Again, the separation between the source and receiver is set at 200 m for Western Cross Harbor Tunnel in the numerical simulations. Figure 2 displays the predicted relative errors versus the orders of reflection for the tunnels with different dimensions. To achieve a comparable relative error in predicting the total sound field, more terms are required for a tunnel with a smaller cross-sectional area. This conclusion agrees with one of the finding in an earlier study for the sound propagation in a narrow city street[6] where more higher-order rays are required for a narrower street.

In the current study, the 50 Hz frequency band is used as the basis to determine the required terms needed in Eq. (1). This is because 50 Hz is the lowest frequency considered in the present study. We find that use of a maximum of 60 reflections from the tunnel walls is sufficiently accurate for the geometrical configuration used in our experimental studies. Generally speaking, fewer terms are required for a source with higher frequencies and for shorter separations between the source and receiver. Although it is possible to optimize the number of terms required in the ray series, no attempt has been made in the numerical simulations shown in this paper. Setting N equal to 60 is adequate to ensure that the resulting ΔP is less than 5% in all calculations.

III. FIELD MEASUREMENTS IN TWO TUNNELS

In a recent study,[1] we have demonstrated that predictions by the coherent model agree well with the measurement data in a scale model experiment. To further validate the coherent model, measurements were conducted at nighttime in two realistic long tunnels, namely Tai Lam Tunnel and Western Harbor Crossing Tunnel in Hong Kong. These tunnels were designed for use by automobiles but closed for general maintenance purposes in the nighttime. Hence, it was a relatively ''quiet'' environment with a typical background noise level of 60 dB(A) because of the lack of traffic flow during the measurement period. However, occasional noise was generated as a result of maintenance activities.

When conducting the measurements at Tai Lam Tunnel, a subwoofer, the Tannoy B475, was chosen as the noise source. This generates a high definition sound at low and ultra-low frequencies. The frequency response was ±3 dB in the range from 28 to 240 Hz. While the measurements in Western Harbor Crossing Tunnel are intended to extend to high-frequency regions, a general-purpose speaker, the Tannoy T300, was chosen as the noise source. The Tannoy T300

TABLE I. Absorption coefficients of the tunnel boundaries at 1/3-octave bands.

Frequency (Hz)	50	63	80	100	125	160	200	250
Absorption coefficient	0.015	0.018	0.02	0.022	0.025	0.027	0.029	0.03
	315	400	500	630	800	1 k	1.25 k	1.6 k
	0.032	0.034	0.035	0.037	0.039	0.04	0.044	0.047
	2 k	2.5 k	3.15 k	4 k	5 k	6.3 k	8 k	
	0.05	0.057	0.063	0.07	0.075	0.08	0.09	

comprises one 12 in. dual concentric driver, in which the low-frequency and high-frequency sources were coincidentally aligned to a point source, resulting in a smooth uniform response (± 3 dB) over a wide frequency range from 55 to 22 kHz.

As the wall surfaces along the length of the tunnels were made of concrete, we assume that all boundaries have the same absorption coefficient in the predictions based on the incoherent model. See Table I for the list of the absorption coefficients at each octave band used by Yang and Shield.[10] In predictions according to the ASJ model, Takagi *et al.*[11] have selected an absorption parameter of a value 0.04 for tunnels with concrete wall surfaces. This value is used in our study of the sound propagation in the two tunnels mentioned above.

Both tunnels were built for the use of automobiles. To minimize the influence of background noise, the measurements were taken in the nighttime with a loudspeaker generating high levels of random noise. A precision type sound level meter was used as a signal receiver to measure sound pressure levels in one-third octave bands. In the following sections, we shall show comparisons of the theoretical predictions with experimental measurements.

To facilitate the comparison the theoretical predictions with field measurements, each one-third band is divided into a number of smaller sub-bands of a constant bandwidth. The total sound field is then computed by summing the noise levels of all sub-bands. Since the number of predicted data for each one-third octave band increases with the increasing center frequency because the one-third octave bands are spaced logarithmically. Consequently, the computational effort increases with the increase of the octave-band center frequencies. To reduce the computational time without affecting the accuracy of predictions, the sub-bands are also spaced logarithmically. A preliminary study suggests that use of the logarithmic sub-bands is sufficient for the present numerical analyses.

A. Measurements in Tai Lam Tunnel

Tai Lam Tunnel was 3.7 km long with a rectangular cross-section of a nominal width and height of 14.2 and 6.0 m, respectively. The two sides of the tunnel walls were furnished with smooth concrete panels that were slightly curve for decorative purposes. These vertical walls were assumed to be reflective flat surfaces. The ground was made of concrete with raised walkways formed adjacent to the vertical walls. A slightly curved ceiling made of concrete slab was regarded as a reflective flat surface in parallel with the ground. The lighting equipment and signal boxes were hung from the ceiling of the tunnel. It is remarkable that the boundaries of the tunnel, i.e., the two vertical walls, the ground and the ceiling are considered to form a long rectangular enclosure in our prediction model. As shown in the following comparisons, this simplification does not cause significant errors in predicting the overall noise levels in the tunnel.

The experiments were conducted at the location such that the tunnel exits at either end do not have significant effects on the measurements. A Tannoy Superdual B475 loudspeaker with external dimensions of 0.55 m$\times$0.58 m high was used as a sound source generating white noise in the field measurements. The sound source was located 1 m above the concrete ground and either 3.5 or 7.1 m from one of the vertical walls. These source locations were chosen to simulate the approximate locations of ground-based sources, such as engine noise, emitted from heavy vehicles or elevated noise sources, such as jet fans, installed under the ceiling of the tunnel. Measurements of sound pressure levels in one-third octave bands were recorded using an Ono Sokki Precision Sound Level Meter type LA-5110 at various receiver locations from 10 to 200 m. The receivers were located at the centerline of the tunnel and on the side at 3.5 m from one of the vertical walls. These two locations are referred to as the centerline and offset line, respectively, in this section. The height of the receiver was set at 1.25 and 2.33 m above the ground. They are referred to as the low and high positions, respectively.

In the first set of measurements, the source was located at the centerline, 7.1 m from one of the vertical walls and 1 m above the concrete ground. To obtain the values of noise reduction (NR), as defined above, the sound pressure level measured at a receiver location at a horizontal distance of 10 m in front of the source and at a height 1.25 m above the ground was used as the reference. This receiver location was referred to as the reference point.

Due to the interference effect of the source and its images, the sound levels at octave bands vary considerably, even when the receivers were located close to the source. Figure 3 shows the experimental results of a frequency spectrum for the octave bands from 50 to 500 Hz. A typical geometry was used in the measurements with a receiver located at a horizontal distance of 10 m from the source. The measurements were taken at the same vertical plane as the reference point, but the receiver was moved from the reference point to the high position at the offset line. As illustrated, the coherent model can be used to predict the interference dips at the low-frequency region considerably well, while the predictions according to the incoherent and ASJ models are rather insensitive to the change of receiver loca-

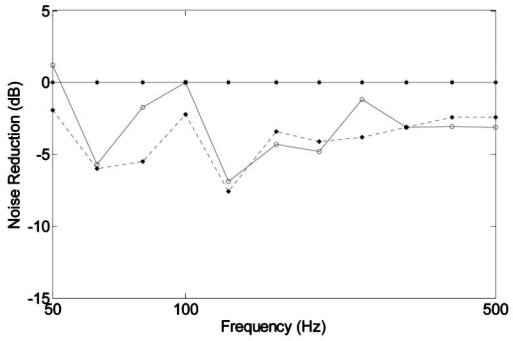


FIG. 3. A comparison of noise reduction spectra among the predictions by the coherent model, incoherent model, ASJ model, and measurements in Tai Lam Tunnel. The source and receiver were placed, respectively, at 1 and 2.33 m above the ground and at respective distances of 7.1 and 10.7 m from one side of the tunnel wall. They were separated by a horizontal distance of 10 m. The predictions according to the incoherent and ASJ models coincide at all frequencies. (Solid line with open circles: Measured results; dashed line with open circle: predictions according to the coherent model; solid line with crosses: predictions according to the incoherent model; dot–dashed line with dark circles: predictions according to the ASJ model.)

tions. The comparison shows that the discrepancy between the measurement result and the prediction by the incoherent and ASJ models can be up to 7 dB at 125 Hz. On the other hand, the variation in the noise reduction (NR) gradually arrives at a steady level of about −3 dB when the octave-band frequency exceeds 315 Hz. The incoherent and ASJ models predict that there is no noise reduction for sound transmitted over a source/receiver separation of 10 m.

To demonstrate the usefulness of the coherent model in the prediction of sound propagation in a long-range situation, we move the receiver farther from the source. We show in Fig. 4 only the results for the measurements at a horizontal distance of 150 m from the source. Again the spectrum frequency is shown with the octave bands varying from 50 to 500 Hz. In this set of measurements, the source was located at the centerline and the receiver at the high position of the offset line. The figure illustrates that the coherent model can be used to predict the fluctuations of noise reduction across the frequency spectrum with accuracy. The incoherent and ASJ models are less accurate prediction schemes for the source frequency below 200 Hz. At the frequency range of interest from 50 to 500 Hz, the ASJ model generally underestimates the sound attenuations along the tunnel.

In the next set of measurements, the source was set at 1

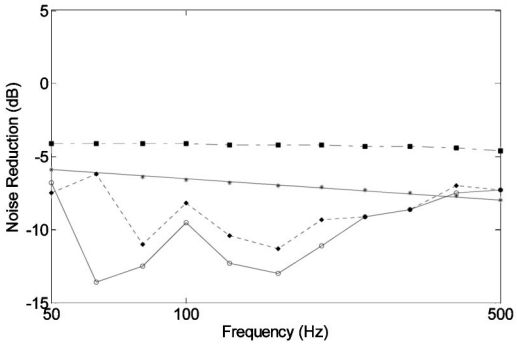


FIG. 4. The same as Fig. 3, except that the source and receiver were separated by a horizontal distance of 150 m.

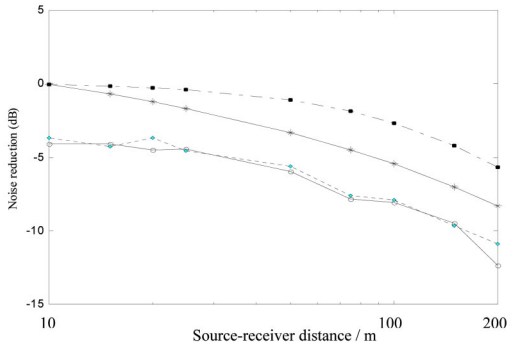


FIG. 5. A comparison of sound attenuation at different source/receiver separations along Tai Lam Tunnel with the noise source located 1 m above the ground and 7.1 m from the tunnel wall. Measurements were conducted for the source frequency of 160 Hz and the receiver locating at the offset line along the tunnel and 1.25 m above ground. The keys for the lines show in the figure are the same as in Fig. 3.

m above the ground and placed either at the centerline or offset at 3.5 m from one of the vertical walls. Noise measurements were conducted along the tunnel from a horizontal distance of 10 to 200 m at the frequency range from 50 to 500 Hz. Two typical sets of data are selected for presentation. In Fig. 5, the receiver was placed at the low position at the offset line with the octave-band frequency of the source at 160 Hz. On the other hand, in Fig. 6, the receiver was placed at the high position at the centerline with the octave band frequency at 315 Hz. These two plots display typical results of the predicted noise reduction by various models with the measured noise reduction plotted against the horizontal distance from the source. These two figures show that higher sound attenuation along the tunnel was achieved when the receiver was located at the low position at the offset line. We also note that there was no appreciable sound attenuation at the receiver locations between 20 and 50 m from the source. Generally speaking, the coherent model gives a more accurate prediction in all source/receiver configurations. As shown in both figures, the predictions according to the coherent model agree to within 3 dB of the experimental measurements. However, there are noticeable discrepancies between the measurement results and predictions according to the incoherent and ASJ models. The differences between the ASJ model and the measurements can

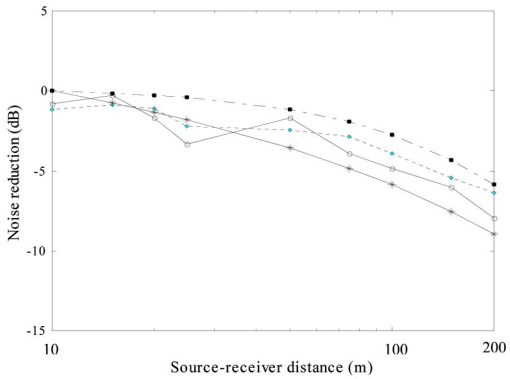


FIG. 6. The same as Fig. 5, except that the source frequency was 315 Hz and the receiver was located at the centerline along the tunnel and at 2.33 aboveground.

be as high as 7 dB. For the source located at the offset position and the receiver located at other distances, comparisons of the experimental data with theoretical predictions have rather similar results. These results are not shown here for brevity.

The usefulness of using the coherent model, i.e., considering the interference, in predicting sound propagation at low frequencies, is generally expected. However, it is of interest to determine the critical frequency where the incoherent models should not be used. The critical frequency may be determined by considering the pathlength difference of the direct ray and the reflected ray of the first order. Suppose that the cross-sectional area of the tunnel is A and the horizontal distance between the source and receiver is L. Then, the typical length scale of the tunnel with a rectangular cross-section is $\sqrt{A}$. The paths of the direct and the reflected ray of the first order have the respective length scales of $\sqrt{L^2+A}$ and $\sqrt{L^2+2A}$. Hence, the pathlength difference has an approximate order of $AL/2(L^2+A)$ which is derived by assuming $D \gg 0$. The effect of interference between the contributory rays is significant if the wavelength λ of the transmitted sound is much less than the pathlength difference, i.e., $\lambda \ll AL/2(L^2+A)$. The critical frequency is then determined by

$$f_c \sim 2c(L^2+A)/AL, \tag{6}$$

where c is the speed of sound in air. The incoherent model should be adequate in predicting the average levels of the transmitted noise if the source has a frequency much greater than the critical frequency.

We remark that the frequency used in the measurements at Tai Lam Tunnel was not high enough to allow an assessment of the critical frequency and source/receiver separation in which the incoherent model may be used to predict the average level of the transmitted noise. This issue will be addressed in the next section.

B. Measurements in Western Harbor Crossing Tunnel

Western Harbor Crossing Tunnel was a 2 km three-lane road tunnel in Victoria Harbor of Hong Kong. This tunnel, which has a rectangular cross-section of a nominal width of 12.5 m and a height of 5.8 m, shares a similar design to the Tai Lam Tunnel described earlier. The tunnel walls and ground were made of concrete with flat and smooth surfaces. Raised walkways adjacent to the vertical walls and other scattering surfaces hung from the reflective ceiling were found. To obtain a full spectrum of measurement results from 50 to 8 kHz, the Tannoy loudspeaker model T300 with external dimensions of 0.37 m×0.59 m in height was used as a sound source generating pink noise. The sound source was located at 0.9 m above the concrete ground and at either 2.65 or 6.25 m from one of the vertical walls. An Ono Sokki Precision Sound Level Meter type LA-5110 was used to take measurements of sound pressure levels in one-third octave bands at various receiver locations from 10 to 400 m. The receivers were located at the centerline of the tunnel and on the side at 2.65 m from one of the vertical walls. This is referred to as the offset line in this section. The height of the receiver was set at 1.3 and 2.38 m above the ground. These are referred to as the low and high positions, respectively.

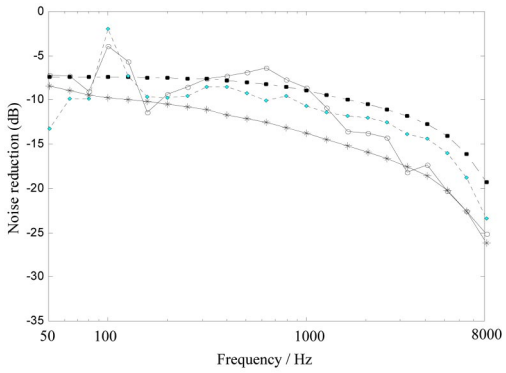


FIG. 7. A comparison of noise reduction spectra among the predictions by the coherent model, incoherent model, ASJ model, and measurements in Western Harbor Crossing Tunnel. The source and receiver were placed, respectively, at 0.9 and 2.38 m above the ground and at respective distances of 9.85 and 6.25 m from one side of the tunnel wall. They were separated by a horizontal distance of 250 m. The keys for the lines show in the figure are the same as in Fig. 3.

An initial measurement of the sound pressure levels at a horizontal distance of 10 m in front of the source was conducted in the tunnel. The result was used as the reference data for the deduction of the noise reduction spectra in our subsequent field measurements. At the reference location, the receiver was placed 1.3 m above the ground. A comprehensive set of measurements was then conducted for various source and receiver locations. These experimental data were used for a comparison with different numerical models. When the horizontal distance between the source and receiver was extended beyond 75 m, say, the noise levels dropped significantly due to the high absorption of sound energy in air, especially at frequency bands over 1 kHz. It is important to include the air absorption factor in the prediction models in order to accurately predict sound propagation in tunnels. To illustrate this point, selected noise reduction spectra are presented for the octave bands varying between 50 and 8000 Hz. Figures 7 and 8 display the results for receivers at a horizontal distance of 250 and 350 m from the source, respectively. In these two examples, the receiver was located at the centerline and at the high position. The figures reveal that the coherent model is able to predict peaks and

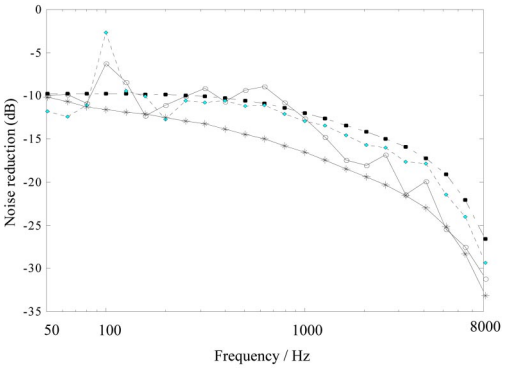


FIG. 8. The same as Fig. 7, except that the source and receiver were separated by a distance of 350 m.

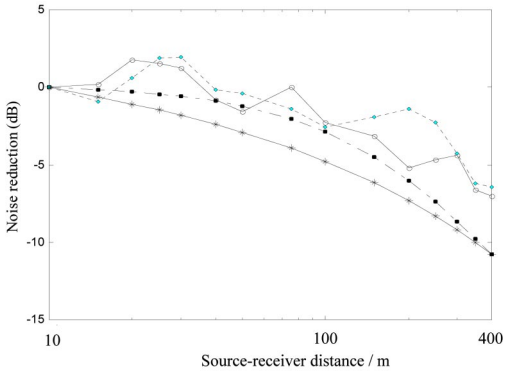


FIG. 9. A comparison of sound attenuation at different source/receiver separations along the Western Harbor Crossing Tunnel. The center-band frequency of the source was 50 Hz. The source and receiver were placed, respectively, at 0.9 and 1.3 m above the ground and both at a distance of 6.25 m from one side of the tunnel wall. The keys for the lines shown in the figure are the same as in Fig. 3.

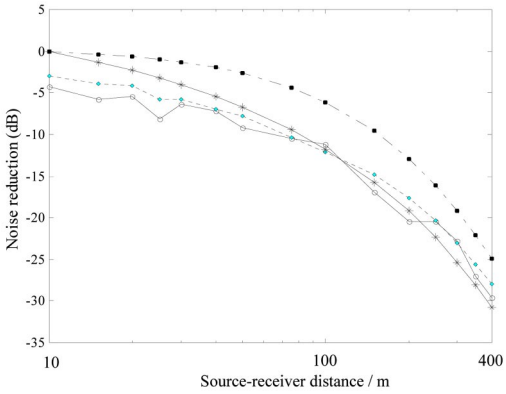


FIG. 10. The same as in Fig. 9, except that the receiver was located at a distance of 6.25 m from the same side of the tunnel wall and the center-band frequency of source was 6300 Hz.

dips occurring at the low-frequency region and gives the highest accuracy across the whole frequency range of interest as compared with the ASJ and incoherent models. In this geometrical configuration, the ASJ model underestimates the noise reduction of up to 6 dB at 3150 Hz, where destructive interference occurs. On the other hand, the incoherent model overestimates the noise reduction of up to 6 dB at 100 and 630 Hz, where constructive interference occurs.

As discussed in the last section, we can roughly estimate the critical frequencies, f_c, for these two examples in Figs. 7 and 8. According to Eq. (6), they are about 2 and 3 kHz for the horizontal distances of 250 and 350 m, respectively. We expect that the incoherent model should be adequate to estimate the average levels of the transmitted noise if the source frequency is higher than the critical frequency. This is supported by the experimental results shown in Figs. 7 and 8 in which the incoherent model gives a fair estimation of the Noise Reduction in the tunnel. Nevertheless, the incoherent model can only provide the average noise levels and are unable to predict the variation of NR in the spectrum as shown in the figures.

To prove the robustness of the coherent model in predicting sound attenuation along the tunnel in a long range propagation and at all frequencies of interest, noise measurements were conducted from horizontal distances of 10 to 400 m from the source, with the mid-band frequency ranging from 50 to 8000 Hz. To present the results, we have chosen only two representative sets of data for the purposes of illustration in the following paragraphs. Similar results comparing the experimental data with theoretical predictions can be found elsewhere.[12]

In the first set of experimental data, the noise source was set at the centerline and at 0.9 m above the concrete ground. The receiver was located at the centerline, and the center-band source frequency was 50 Hz. The noise reduction along the tunnel at various one-third octave bands are plotted to compare the measurement results with the predictions by various models. Figure 9 illustrates that the predictions by the coherent model give the best agreement with the measurement results for sound propagation up to 400 m. It also shows that a strong constructive interference effect occurs at receiver locations at a distance of between 20 and 30 m. Redmore[13] referred to this phenomenon as a ''plateau'' in his experimental data. In our case, both the incoherent model and the ASJ model overestimate the noise reduction, and the discrepancies become relatively large at distances when destructive interference occurs.

In the next set of data, the receiver was shifted to the offset line and at 2.38 m above the ground. The center-band frequency of the source is 6.3 kHz. Despite the fact that the interference effect becomes less significant at high frequencies, Fig. 10 shows that prediction results by the coherent model agree very well with the measurements at this set of data for the separation between the source and receiver extending to 400 m. The results predicted by the ASJ model appear to be the least accurate in this example. Similar conclusions can be drawn based on the comparison of measured and predicted results for other geometrical configurations but these results are not shown here for brevity.

IV. SUMMARY

The experimental data for the full-scale field experiments at Tai Lam Tunnel and Western Harbor Crossing Tunnel confirms a finding in a recent study on scale model experiments. The coherent model gives the best agreement with experimental results among all three prediction schemes, particularly at the low- and mid-frequency regions. The variation of the noise reduction at different frequencies is a feature that can only be predicted by the coherent model with accuracy to within 3 dB. The effect of mutual interference caused by the direct and reflected rays is an important factor for the accurate prediction of sound fields in long enclosures.

The incoherent model generally gives satisfactory predictions when the total sound field shows a less significant effect of interference between contributory rays. This will occur when the source frequency is sufficiently high and when the separation between source and receiver is small. For instance, as shown in a typical set of experimental measurements, the agreements are good for a source frequency of 6.3 kHz and a horizontal separation of 30 m. In this case, the total sound field can be approximated by summing all contributory rays incoherently.

Although the coherent model gives reasonably accurate results in most cases, some discrepancies between the field data and theoretical predictions are found. This is probably due to the scattering of sound from the raised walkway adjacent to the vertical walls and other scattering surfaces hung from the ceiling. The fact that the tunnel section is not a true rectangular shape also affects the results.

ACKNOWLEDGMENTS

One of the authors (KKI) was supported by a Teaching Company Scheme jointly sponsored by the Industry Department of the HKSAR Government and NAP Acoustics (Far East) Ltd. The research described in this paper was supported in part by the Research Grants Council of the HKSAR Government and The Hong Kong Polytechnic University. The authors are grateful to the Route 3 (CP3) Co. Ltd. and the Western Harbour Tunnel Co. Ltd. for granting access to their tunnels.

[1] K. M. Li and K. K. Iu, ''Propagation of sound in long enclosures,'' J. Acoust. Soc. Am. **116**, 2759–2770 (2004).
[2] G. Lemire and J. Nicolas, ''Aerial propagation of spherical sound waves in bounded spaces,'' J. Acoust. Soc. Am. **86**, 1845–1853 (1989).
[3] H. E. Bass, L. C. Sutherland, A. J. Zuckerwar, D. T. Blackstock, and D. M. Hester, ''Atmospheric absorption of sound: Further developments,'' J. Acoust. Soc. Am. **97**, 680–683 (1974).
[4] ANSI S1.26, ''Method for calculation of the absorption of sound by the atmosphere,'' 1995.
[5] Research Committee of Road Traffic Noise in Acoustical Society of Japan, ''ASJ Prediction Model 1998 for Road Traffic Noise,'' J. Acoust. Soc. Jpn. **55**, 281–324 (1999).
[6] K. K. Iu and K. M. Li, ''The propagation of sound in narrow street canyons,'' J. Acoust. Soc. Am. **112**, 537–550 (2002).
[7] J. Kang, ''Reverberation in rectangular long enclosures with geometrically reflecting boundaries,'' Acustica **82**, 509–516 (1996).
[8] R. G. White and J. G. Walker, *Noise and Vibration* (Wiley, New York, Year), p. 54.
[9] S. M. Dance and B. M. Shield, ''The complete image-source method for the prediction of sound distribution in non-diffuse enclosed spaces,'' J. Sound Vib. **201**, 473–489 (1997).
[10] L. Yang and B. M. Shield, ''The prediction of speech intelligibility in underground stations of rectangular cross section,'' J. Acoust. Soc. Am. **109**, 266–273 (2001).
[11] K. Takagi, T. Miyake, K. Yamamoto, and H. Tachibana, ''Prediction of road traffic noise around tunnel mouth,'' *Proc. InterNoise 2000*, pp. 3099–3104.
[12] K. K. Iu, ''The prediction of noise propagation in street canyons and tunnels,'' M. Phil. thesis, The Hong Kong Polytechnic University, 2003.
[13] T. L. Redmore, ''A theoretical analysis and experimental study of the behaviour of sound in corridors,'' Appl. Acoust. **15**, 161–170 (1982).

Effective impedance of surfaces with porous roughness: Models and data

Patrice Boulanger, Keith Attenborough,[a] and Qin Qin
Department of Engineering, The University of Hull, Hull HU6 7RX, United Kingdom

(Received 28 June 2004; revised 11 November 2004; accepted 23 November 2004)

A ''boss'' formulation by Twersky [J. Acoust. Soc. Am. **73**, 85–94 (1983)] enables prediction of the plane wave reflection coefficient from a surface composed of *rigid-porous* roughness elements embedded in an *acoustically hard* plane where the roughness elements and their mean spacing are small compared with the incident wavelengths. Predictions for air-filled porous roughness elements on a hard ground plane are compared with effective impedance spectra obtained from laboratory measurements over random distributions of polystyrene hemi-spheres, polyurethane pyramids, and sand hemispheroids on glass plates. Overall the predictions agree well with these data. To enable prediction of the effective admittance of rough porous surfaces, Twersky's original formulation is extended heuristically. The resulting theory is compared with a previous model [J. Acoust. Soc. Am. **108**, 949–956 (2000)], which is a heuristic extension of Tolstoy's theory [J. Acoust. Soc. Am. **72**, 960–972 (1982)] to include nonspecular scattering. The theories are found to give different predictions for relatively large bosses. The modified Twersky theory gives relatively good predictions of the effective impedance spectra obtained from complex sound pressure level measurements over sand surfaces containing semielliptical roughness elements and over uncultivated soil. [DOI: 10.1121/1.1850211]

PACS numbers: 43.50.Vt, 43.28.En [RR] Pages: 1146–1156

I. INTRODUCTION

Surface roughness is known to have significant influence on near-grazing sound. One approach to modeling long wavelength sound reflection from randomly rough surfaces considers scattering from idealized roughness elements or ''bosses.'' Several measurements have been made of sound pressure level (SPL) or relative SPL spectra above rough surfaces, where the roughness height and spacing are small compared to the wavelengths of interest.[1–5] These data have been compared with predictions of models derived by Attenborough and Taherzadeh[1] from a boss theory by Tolstoy.[6,7] It was necessary to adjust the impedance of the scatterers and imbedding plane to obtain good agreement between predictions based on the Tolstoy boss theory and the data. Tolstoy's effective admittance models[6,7] predict that a surface wave is generated at grazing-incidence above a hard rough boundary and that the effective admittance above a hard rough boundary is purely imaginary. However, comparison with data indicates that Tolstoy-based predictions overestimate the surface wave component, especially at grazing incidence[2] and that it is necessary to include an attenuation term due to incoherent scatter to obtain a good fit with these data.[4] In other comparisons of predictions and data,[5] the assumed location of the effective admittance plane has been adjusted to improve agreement with data at higher frequencies. Poor agreement between laboratory measurements of propagation over rough convex surfaces and predictions based on the Tolstoy effective admittance formulation has been found[4] and it has been suggested that the absence of a real part of admittance corresponding to nonspecular scattering effects might be responsible for this. Nonspecular scattering is particularly important for random roughness.[1] The theory of Lucas and Twersky[8] incorporates nonspecular scattering and resulting expressions have been used to model two-dimensional (2D) periodic and random hard roughness elements, giving reasonable agreement with measured short-range ground effect.[9] A heuristic extension of this model to sound reflection from rough surfaces of finite impedance, named the Biot/Tolstoy/Howe/Twersky (BTHT) model, has given predictions in good agreement with ground effect measured over rough surfaces in the laboratory and outdoors.[10] However, the real part of the effective impedance obtained from Lucas and Twersky's theory for *hard* rough surfaces does not have the low frequency limit that is expected from physical considerations,[11] i.e., the real part should be much greater than the imaginary part.

Twersky's general theory[12] models reflection from porous roughness elements on a hard plane. It introduces a term proportional to wave-number k in the real part of effective admittance ($\mathrm{Re}(\beta)$). To the authors' knowledge, this theory has not been applied to an air/solid interface with air-filled porous roughness elements, and has not been validated by comparison with data. To this end, measurements of complex relative SPL over porous roughness on a flat hard surface have been made. Effective impedance spectra have been obtained from these measurements by using Muller's numerical method to solve the complex admittance equation obtained from the classical expression for a point source over an impedance boundary.[13,14] The complex compressibility and mass density of fluid in the pores of the roughness elements have been calculated assuming a rigid matrix of tortuous slit-like pores.[15,16] Predictions of Twersky's theory[12] have been compared with these data.

[a]Electronic mail: k.attenborough@hull.ac.uk

For predicting outdoor sound, propagation over completely rough and porous surfaces is of interest. It is found that predictions of Twersky's theory[12] do not agree at low frequencies with measured data for rough porous surfaces. This model has been extended, heuristically, to predict the effective admittance of a rough porous boundary and compared to the heuristic BTHT model used in previous work.[10] The effective impedance spectra obtained from these theories are compared with those deduced from measurements of complex relative sound pressure level over rough sand surfaces and uncultivated soil.

Section II A presents Twersky's effective admittance model for surfaces formed by arbitrarily shaped porous roughness elements on a hard plane. Section II B summarizes the theory used to model the complex compressibility and thermal conductivity for the rigid-porous roughness elements considered. Section II C derives the admittance expressions in the particular cases of semicylindrical and semielliptical porous roughness elements from Twersky's results presented in Sec. II A. Section II D derives the admittance expressions for hemispherical and hemispheroidal porous roughness elements. Section II E recalls the heuristic BTHT model and introduces a heuristic extension of Twersky's model to predict the effective admittance of a porous surface with porous roughness elements. Results from both models are compared for identical parameters. Section II F reprises briefly how complex relative SPL spectra are used in effective impedance calculations. Section III describes the experimental procedures. To validate Twersky's original theory,[12] Sec. IV A compares the model's predictions with effective impedance spectra obtained from laboratory measurements of complex relative sound pressure level over random distributions of polystyrene hemispheres, polyurethane pyramids, and sand hemispheroids on glass plates. Section IV B compares impedance predictions of the heuristic theories for rough porous boundaries with those deduced from measurements of complex relative sound pressure level over corrugated sand surfaces and rough uncultivated soil. Conclusions are drawn in Sec. V.

II. THEORETICAL FORMULATIONS

A. Twersky's theory for porous roughness elements distributed randomly on a hard plane

The main results of Twersky's general theory[12] are summarized here using notation consistent with previous publications.[9,10] The direction cosines of the incident wave vector $\mathbf{k}$ are γ_i where $i=1$, 2, or 3 corresponding to the coordinate axes z, x, and y, respectively. If φ is the azimuthal angle and α is the angle between $\mathbf{k}$ and the vertical axis z then $\gamma_1=\cos(\alpha)$, $\gamma_2=\sin(\alpha)\cos(\varphi)$, $\gamma_3=\sin(\alpha)\sin(\varphi)$. The acoustical properties of the porous roughness elements are specified by the relative compressibility κ and the inverse mass density ratio B_i. Specifically, complex coefficients B_i' and C are introduced such that

$$C=\kappa-1 \tag{1}$$

and

$$B_i'=-(B_i-1)/(1+(B_i-1)A_i), \tag{2}$$

where

$$B_i'=B_{i(1)}'-iB_{i(2)}' \tag{3}$$

and

$$C=C_1+iC_2. \tag{4}$$

C_2 and $B_{i(2)}$ are positive and small compared to C_1 and $B_{i(1)}$. The anisotropy factors $A_i=q_i+\epsilon_i$ $(i=1,2,3)$, where q_i are related to roughness element shape through the principal diameters of the roughness elements and ϵ_i are given by integral equations involving the packing density of the random roughness distribution. The number of roughness elements per unit area is denoted by n, the wave number magnitude by k, and the cross-sectional area of the roughness element by U. The imaginary part of the effective surface admittance β is given by

$$\mathrm{Im}(\beta_{\pm})=-\frac{nU}{2}k\Phi_1^{\pm}, \tag{5}$$

where

$$\Phi_1^{+}=C_{(1)}+B_{2(1)}'\gamma_2^2-B_{3(1)}'\gamma_3^2 \tag{6}$$

and

$$\Phi_1^{-}=-B_{1(1)}'\gamma_1^2. \tag{7}$$

The $\pm$ subscripts indicate hard/pressure release surfaces, respectively. The general formulation introduces a dyadic parameter $\boldsymbol{B}$ for inverse mass density ratio, and incorporates the shape coupling integrals q_i in a dyadic $\boldsymbol{q}$ but, here, it is assumed that the principal axes of all dyadics are aligned with the Cartesian axes and vectors B_i are used instead.

The real part of the effective surface admittance, based on Twersky's[12] relationship (95) and involving absorption and scattering terms, is given by

$$\mathrm{Re}(\beta_{\pm})=-\frac{nU}{2}k\Phi_a^{\pm}-\frac{nW_sU^2}{8}\left(k^3,\frac{k^4}{\pi}\right)\Phi_S^{\pm}, \tag{8}$$

$$\Phi_a^{+}=C_{(2)}+B_{2(2)}'\gamma_2^2+B_{3(2)}'\gamma_3^2, \tag{9}$$

$$\Phi_a^{-}=B_{1(2)}'\gamma_1^2, \tag{10}$$

$$\Phi_s^{+}=C_{(1)}^2+\frac{B_{2(1)}'^2\gamma_2^2}{m}+\frac{B_{3(1)}'^2\gamma_3^2}{m}, \tag{11}$$

$$\Phi_s^{-}=\frac{B_{1(1)}'^2\gamma_1^2}{m}. \tag{12}$$

In Eqs. (8) and (11), the values $(k^3,m=2)$ and $(k^4,m=3)$ are used for 2D and three-dimensional (3D) roughness, respectively. The packing factor W_s depends on the packing density W of the corresponding distribution of roughness elements. $W=nb^*=b^*/b$, where b is the average roughness separation distance and b^* is the size of the exclusion region around a scatterer, i.e., is the minimum distance between the centers of adjacent roughness elements. Since the work reported here excludes pressure release surfaces, only expressions corresponding to the superscript $+$ are used. Note that Twersky's theory assumes that the product of roughness size and wave number is *very much* less than unity and that the

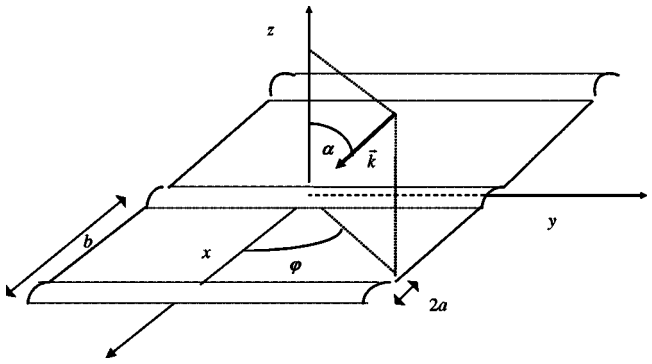


FIG. 1. A plane wave incident on a surface containing a grating of semicylinders.

product of roughness separation distance and wave number is less than unity.

B. The tortuous slit-pore model for rigid-porous materials

For simplicity, although more sophisticated models for the acoustical properties of arbitrary rigid-porous materials are available,[15] the pores in the roughness elements are considered to be identical tortuous slits. It should be noted that, for a given flow resistivity and porosity, pore shape has little influence on the surface impedance.[16] In a medium with flow resistivity R_s, porosity Ω, fluid density ρ_f, and tortuosity T, and assuming time dependence $\exp(-i\omega t)$, the effective relative compressibility κ is given by[16]

$$\kappa(\omega)=\gamma-(\gamma-1)\left[1-\frac{\tanh(\lambda\sqrt{-iN_{PR}})}{\lambda\sqrt{-iN_{PR}}}\right], \tag{13}$$

where the dimensionless parameter $\lambda=(3\rho_0\omega T/\Omega R_S)^{1/2}$, $\gamma=C_p/C_v$ is the ratio of specific heat at constant pressure of the gas to the specific heat at constant volume, ρ_0 is the density of air, and $\omega=2\pi f$ is the angular frequency. The relative inverse mass density ratio (the density of air divided by the complex density of air in the pores) is given by

$$B(\omega)=\left[1-\frac{\tanh(\lambda\sqrt{-i})}{\lambda\sqrt{-i}}\right]. \tag{14}$$

The relative impedance for a smooth surface with slit-like pores is given by

$$Z_{\mathrm{rel}}=\frac{1}{\beta_{\mathrm{smooth}}}=\sqrt{\frac{T}{\Omega^2 B(\omega)\kappa(\omega)}}. \tag{15}$$

C. Twersky's theory for semicylindrical porous roughness embedded in a flat hard surface

Consider circular semicylindrical roughness elements with axes along the y direction (see Fig. 1). When the azimuthal angle φ is zero and the plane-wave incidence angle α is measured from the vertical axis z, the direction cosines are $\gamma_1=\cos\alpha$, $\gamma_2=\sin\alpha$, and $\gamma_3=0$. Equations (6), (9), (11) imply that the factors with index $i=3$ are not relevant in 2D. The restriction to nonpressure release surfaces means that only $i=2$ is of interest. For semicylindrical roughness elements, the packing factor[12] $W_s=(1-W)^2$. For circular cylinders, $q_2=1/2$ and $\epsilon_2=-UI_2/2\pi b^{*2}$. The raised cross-sectional area per unit length is $V=n\pi a^2/2=nU/2$, where a is the radius of the circular semicylindrical roughness and n the number of semicylinders per unit length ($=1/b$ in 2D). Hence, for circular semicylinders,

$$A_2=q_2+\epsilon_2=\frac{1}{2}-\frac{I_2U}{2\pi b^{*2}}$$

with[8]

$$I_2\cong 2W(1+0.307W+0.137W^2) \quad \text{for } W<0.8,$$

$$I_2\cong\frac{\pi^2}{3}\left[1-\frac{2(1-W)}{W}\right]+6\frac{(1-W)^2}{W^2}\times\left[\frac{\pi^2}{6}+1.202\right] \quad \text{for } W\geqslant 0.8,$$

$$I\cong\frac{(\pi a)^2}{3b^2} \quad \text{for } W=1 \text{ (periodic spacing)}.$$

The effective admittance β relative to air of a hard plane containing randomly spaced 2D porous circular semicylinders is deduced from Eqs. (5) and (8) as

$$\beta_{2\mathrm{D}}=\eta_{2\mathrm{D}}-i\xi_{2\mathrm{D}}, \tag{16}$$

where

$$\xi_{2\mathrm{D}}=kV\left\{C_1-\mathrm{Re}\left[\frac{B-1}{1+(B-1)A_2}\right]\sin^2\alpha\right\} \tag{17}$$

and

$$\eta_{2\mathrm{D}}=kV\left\{C_2-\mathrm{Im}\left[\frac{B-1}{1+(B-1)A_2}\right]\sin^2\alpha\right\}+k^3(1-W)^2\times\frac{V^2}{2n}\left\{C_1^2+\frac{1}{2}\mathrm{Re}\left[\frac{B-1}{1+(B-1)A_2}\right]^2\sin^2\alpha\right\}. \tag{18}$$

Note again that Twersky's theory[12] is based on the approximation $ka\ll 1$ and $kb^*<1$.

This is a generalization of the theory for nonporous rigid roughness elements presented in previous work.[9] If the roughness elements are nonporous, rigid, and on a rigid base plane, the inverse mass density ratio B tends to *zero* and $\kappa_r\ll\kappa_a$ implying $C_1=-1$ and $C_2=0$. As a result, for azimuthal angle $\varphi=0$, the effective impedance relationship (2) and (3) in Ref. 9 are recovered. If the roughness elements are air, i.e., there is no roughness, the roughness compressibility is that of air, so both $C_1=\mathrm{Re}(\kappa_r/\kappa_a-1)$ and $C_2=\mathrm{Im}(\kappa_r/\kappa_a)$ are zero and the relative inverse mass density ratio $B=1$. This means that the effective admittance of the surface is zero, as it should be for a smooth hard surface.

Lucas and Twersky's theory[8] for hard rough surfaces predicts an imaginary part of impedance that dominates at low frequencies. This is unphysical since the real part should dominate at the rough interface between two fluids when the wavelength is much larger than the roughness size. However, in expression (18) for porous roughness elements on a hard

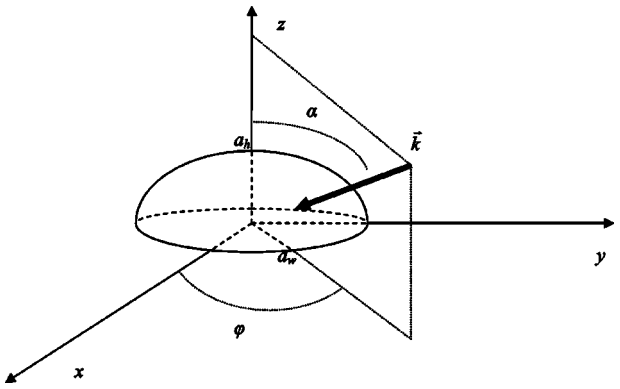


FIG. 2. A plane wave incident on a single hemispheroidal boss.

plane, the first power of wave number in the real part of admittance η_{2D} depends on the imaginary part of the effective compressibility and on the imaginary part of the inverse effective density ratio of the roughness elements. These give the correct behavior at low frequency. Predictions are compared with measured data for surfaces with porous roughness elements in Sec. IV A so testing whether the combination of such small imaginary terms allows $\mathrm{Re}(Z) \gg \mathrm{Im}(Z)$ at low frequency.

If the roughness elements are semielliptical cylinders[12] with semi-axes a_h and a_w,

$$A_2 = \frac{a_h}{a_h + a_w} + \frac{I_2 U}{2\pi b_*^2},$$

and the roughness cross section per unit length becomes $V = n(\pi a_h a_w/2)$. For elliptic semicylindrical roughness elements, these values of V and A_2 are the only modifications required to the effective admittance expressions (17) and (18) for circular semicylinders.

D. Twersky's theory for hemispherical and hemispheroidal porous roughness embedded in a flat hard surface

For n hemi-spherical roughness elements per unit area, with symmetry axis along the z direction (see Fig. 2), the direction cosines are $\gamma_1 = \cos(\alpha)$, $\gamma_2 = \sin(\alpha)\cos(\varphi)$, and $\gamma_3 = \sin(\alpha)\sin(\varphi)$. For any 3D roughness elements, the packing factor[12] $W_s = (1-W)^3/(1+W)$. If the exclusion region around a roughness element is assumed to be isotropic, $b_3^* = b_2^* = b^*$, $W = n\pi(b^*/2)^2$, $q_1 = q_2 = q_3 = 1/3$, and $\epsilon_3 = \epsilon_2 = -UI_2/8\pi b^{*3}$. In Eqs. (6), (9), and (11) only $i = 2$ or 3 is of interest since pressure-release surfaces are not considered. If the radius of the hemispherical roughness elements is a, the raised cross-sectional area per unit length is $V = 2n\pi a^3/3 = nU/2$ and

$$A_3 = A_2 = q_2 + \epsilon_2 = \frac{1}{3} - \frac{UI_2}{8\pi b^{*3}}.$$

The expression for the integral I_2 is based on a 2D packing of impenetrable circular disks[12] such that $I_2 = 8W(1 + 0.4157W)$ for $0 < W < 0.84$. The upper bound on W is an experimental value for the densest random packing of identical circular disks. Note that the packing density W for disks (and hemispheres) can never reach the value unity as is possible with 2D roughness elements. From Eqs. (5) and (8), the effective admittance relative to air of a rough surface containing randomly spaced 3D porous hemispheres embedded in a smooth hard surface is given by

$$\beta_{3D} = \eta_{3D} - i\xi_{3D}, \tag{19}$$

where

$$\xi_{3D} = kV\left\{ C_1 - \mathrm{Re}\left[\frac{B-1}{1+(B-1)A_2}\right]\sin^2\alpha \right\} \tag{20}$$

and

$$\eta_{3D} = kV\left\{ C_2 - \mathrm{Im}\left[\frac{B-1}{1+(B-1)A_2}\right]\sin^2\alpha \right\} + k^4 \frac{(1-W)^3}{1+W}\frac{V^2}{2\pi n}\left\{ C_1^2 + \frac{1}{3}\mathrm{Re}^2\left[\frac{B-1}{1+(B-1)A_2}\right]\sin^2\alpha \right\}. \tag{21}$$

If the roughness elements are prolate or oblate hemispheroids and the z axis is assumed to be an axis of symmetry then $q_2 = q_3$. If the exclusion region around a roughness element is assumed to be isotropic, $b_3^* = b_2^* = b^*$. Therefore, $\epsilon_3 = \epsilon_2 = -UI_2/8\pi b^{*3}$ where $U = 4\pi a_w^2 a_h/3$ where a_h is the semiheight and a_w is the semiwidth of the spheroid. For prolate spheroids, $r = a_h/a_w > 1$ and $q_2 = q_3 = (1-q_1)/2$ where[12]

$$q_1 = \frac{1}{r^2-1}\left(\frac{r\cosh^{-1} r}{(r^2-1)^{1/2}} - 1\right),$$

$$\cosh^{-1} r = \ln\{r + (r^2-1)^{0.5}\}.$$

In the case of oblate spheroids,

$$r < 1, \quad q_1 = \frac{1}{1-r^2}\left(1 - \frac{r\cos^{-1} r}{(1-r^2)^{1/2}}\right).$$

Note that the choices of roughness shape, such that $a_3 = a_2 = a_w$, and exclusion region, such that $b_3^* = b_2^* = b^*$, imply an isotropic roughness distribution. The theory may be used for anisotropic roughness element distributions but this is not pursued here.

E. Heuristic models for rough porous surfaces

The Biot/Tolstoy/Howe/Twersky model. A model, introduced elsewhere[10] and tested against measured data over sand, grassland, and subsoiled ground, is based on a heuristic extension of Biot/Tolstoy/Howe theory to include a real part of admittance $\eta_{2D\,\mathrm{Twer}}$ for nonspecular scattering based on Lucas and Twersky's work concerning 2D rough hard surfaces. Hence, when the azimuthal angle φ is not zero

$$\beta^* = \eta - ik\cos^2(\varphi)\epsilon_{01}^* + \beta, \tag{22}$$

where

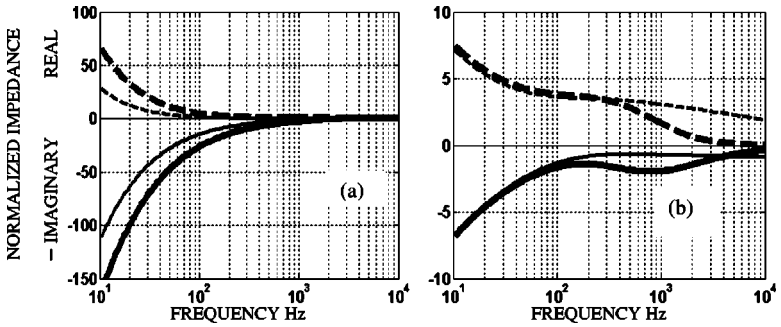


FIG. 3. Comparison of the real (dashed lines) and imaginary parts (solid lines) of the effective impedance spectra predicted by the heuristic BTHT (thin lines) and heuristic Twersky (thick lines) models for rough porous surfaces containing semicylindrical roughness elements ($d=2$ m, $h=0.46$ m, $a=0.05$ m, $b=b^*=0.2$ m and $\Omega=0.4$) (a) $R_s=5000$ kPa s m^{-2}, (b) $R_s=5$ kPa s m^{-2}.

$$\eta(\alpha,\varphi)\approx\frac{k^3bV^2}{2}(1-W)^2\left\{(1-\sin^2\alpha\sin^2\varphi)\times\left[1+\left(\frac{\delta^2}{2}\cos^2\varphi-\sin^2\varphi\right)\sin^2\alpha\right]\right\}+O(k^5),$$

$$\delta=2s_{2D}/\nu_{2D},\quad \epsilon_{01}^*=a_{01}^*-b_{01}^*,$$

$$a_{01}^*=2V\left[\frac{\rho_1-\rho_0}{\rho_1+\rho_0}\right]\frac{s_{2D}}{\nu_{2D}},\quad b_{01}^*=V(1-m_1/m_0),$$

$$s_{2D}=\frac{\rho_0+\rho_1}{\rho_1+K\rho_0}s_2,\quad \nu_{2D}=1+\frac{2\pi V}{3Nb^2}\left[\frac{\rho_1-\rho_0}{\rho_1+\rho_0}\right]s_{2D},$$

$$s_2=\frac{1}{2}(1+K),\quad m_j=\frac{1}{\rho_j c_j^2},\quad j=0,1.$$

ρ_1 and c_1 represent the complex density and sound speed of the porous medium and K represents a hydrodynamic shape factor[17] ($=1$ for semicylindrical roughness elements).

For $\varphi=0$ and porous elements with sufficiently high flow resistivity, Eq. (22) may be expressed as

$$\beta_{\text{Heur Biot}}=\eta_{\text{2D TwerHard}}-ikV[\delta\sin^2\alpha-1+\Omega\gamma]+\beta_{\text{smooth}},\tag{23}$$

where δ is the dipole coupling term.

For hard semicircular cylindrical roughness elements, $\delta=2/(1+I)$ and $I=(a^2/b^{*2})I_2$.

Equation (22) represents the previously published[10] heuristic extension of Tolstoy's theory. Another possibility is to extend Twersky's model for porous roughness elements on a hard plane. The new heuristic model is based on the assumption that the admittance of rough surface containing porous roughness elements embedded in a smooth porous ground is the sum of the admittance of a smooth porous surface and that due to porous roughness elements embedded in a smooth hard surface. Hence the admittance is defined by

$$\beta_{\text{2D,3D-porous}}+\beta_{\text{smooth}},\tag{24}$$

where the first term is computed from either Eqs. (16) to (18) for 2D roughness or from Eqs. (19) to (21) for 3D roughness. Support for this heuristic assumption is provided by Howe's equations (2.28) and (2.31).[18] The β_{smooth} term proves to be important for agreement with measured data at low frequencies. Note that, according to Eqs. (22)–(24), even when the admittance of the smooth porous surface does not depend on incident angle, i.e., the smooth surface is locally reacting, the admittance of the porous rough surface is angle-dependent. Predictions of Eqs. (22) and (24) are compared in Fig. 3. There are significant differences in the predictions for comparatively large (0.05 m radius) semicylindrical roughness elements with both high and low flow resistivity.

F. Computation of relative SPL and deduction of effective impedance

The relative sound pressure level from a point source over an impedance plane is obtained from

$$\text{rel SPL}=20\log\left|\frac{P}{P_1}\right|,\tag{25}$$

where P is the total (complex) pressure at the receiver due to a point source above an homogeneous impedance plane and $P_1=P_0(e^{ikR_1}/R_1)$ is the pressure due to the direct wave from the source [assuming time dependence $\exp(-i\omega t)$]. To estimate the effective impedance relative to air of a rough surface we consider that the measured relative SPL is produced by a smooth impedance surface. In the classical theory for the sound field due to a point source above an impedance boundary,[19]

$$P=P_0\frac{e^{ikR_1}}{R_1}+QP_0\frac{e^{ikR_2}}{R_2}$$

where Q is the spherical wave reflection coefficient defined by $Q=R+[1-R]F(w)$ where $F(w)=1+i\sqrt{\pi}we^{-w^2}\operatorname{erfc}(-iw)$, R_1 is the direct path length from source to receiver, R_2 is the path length through the specular reflection point, and P_0 is a constant. For local reaction, the plane wave reflection coefficient is defined by $R=(\cos\alpha-\beta)/(\cos\alpha+\beta)$. $w=\sqrt{\frac{1}{2}ikR_2}(\beta+\cos\alpha)$ is often called the "numerical distance." The problem of deducing surface impedance from data becomes a search for complex admittance roots, β. Although various numerical root-search methods have been used in the literature, Muller's method[13] is used in this work. Twersky's theory is based on incident plane waves, so the predicted effective admittance β is used in the expressions for R and w to obtain the predicted relative SPL spectra. Equation (25) is used also to obtain relative SPL from measured P and P_1.

III. EXPERIMENTS

In a series of (anechoic) laboratory and outdoor measurements over rough surfaces, a Tannoy driver, fitted with a 1-m-long tube having 3 cm internal diameter was used as a point source. The receiver was a Bruel & Kjaer type 4311

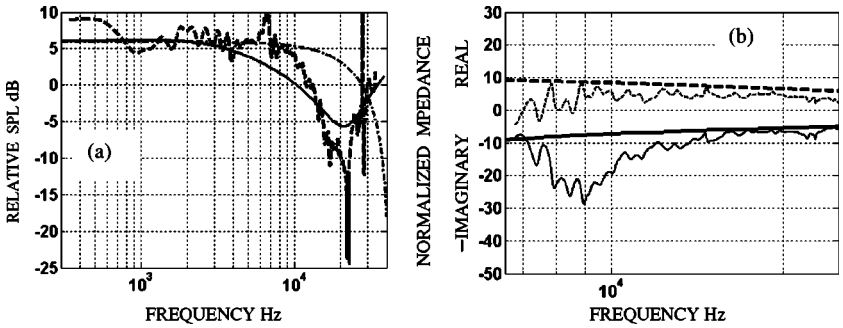


FIG. 4. (a) Averaged measurements (dashed line) of the sound pressure level relative to free field over a rough surface consisting of 5000 m^{-2} hemispherical polystyrene roughness elements on a glass plate ($d=0.5$ m, $h=0.032$ m, $a=0.004$ m, $b^*=2a+0.001$ m) compared to predictions of 3D Twersky theory—Ref. 12 (solid line) for hemispherical porous bosses (all parameters have nominal values except $n=6000\ m^{-2}$, $R_s=500$ kPa s m^{-2}, $\Omega=0.3$, and $T=1/\Omega$) and to the exact solution for a flat hard surface (dot-dash line). (b) Real (dashed lines) and imaginary (solid lines) parts of deduced effective impedance spectra from measurements (thin lines) and predictions (thick lines).

''half-inch'' diameter condenser microphone fitted with a preamplifier. Signal processing and signal generation were carried out using a maximum length sequence system analyzer. This enabled background noise effects to be reduced by analyzing the microphone signal with respect to the known output sequence. To calculate the complex relative SPL, a reference measurement was made with source and receiver raised to a significant height above the surface. In the first series of measurements, 2D semicylindrical and semielliptical roughness elements were created in a smooth slightly wetted sand surface using a comb-like device. The resulting porous rough surfaces consisted of periodic semicylindrical ($a=0.01$ m) and semielliptical ($a_h=0.008$ m and $a_w=0.02$ m) sand roughness elements, respectively. The sand was retained in a 2 m square 30-cm-deep box. Additional measurements have been performed outdoors over an uncultivated porous soil. Another series of (nonanechoic) laboratory measurements were devised to test the original Twersky theory.[12] More or less identical porous roughness elements were placed randomly on thick glass base plates (0.76 m square and 6 mm thick). These experiments have used smaller source–receiver separations and higher frequencies than used in the first series to avoid the influence of unwanted reflections (in the nonanechoic environment) and to reduce the areas of rough surfaces required. The roughness elements consisted, respectively, of (i) hemispheroidal sand ($a_h=0.0153$ m, $a_w=0.0125$ m), (ii) hemispherical polystyrene ($a=0.004$ m), and (iii) open cell rounded-top pyramidal polyurethane foam. The polystyrene and polyurethane roughness elements were glued firmly to the rigid glass base plates.

IV. COMPARISONS BETWEEN PREDICTIONS AND DATA

A. Porous roughness elements on a hard surface

1. Hemispherical polystyrene roughness on a glass plate

Figure 4(a) compares measured and predicted excess attenuation spectra over 5000 m^{-2} 4 mm radius polystyrene hemispheres glued to a glass plate with equal source and receiver heights (h) of 0.032 m and separation $d=0.5$ m. The roughness size is such that the small ka approximation implicit in Eqs. (20) and (21) should be valid up to 13 000 Hz. Comparison of the data with the prediction for a flat hard surface in Fig. 4(a) confirms that a clear ground effect shift is caused by the roughness. Since independent data for the impedance of the polystyrene hemispheres were not available, parameter values have been adjusted for best fit. In this case best-fit values are $R_s=500$ kPa s m^{-2}, $\Omega=0.3$, $T=1/\Omega$. With these values, a minor variation from the nominal geometrical parameters $a=0.004$ m, $b^*=2a+0.001$ m in the 3D Twersky theory [Eqs. (20) and (21)] gives a good fit to the relative SPL averaged over three measured sets. An area density of $n=6000\ m^{-2}$ gives better agreement than the measured value of $n=5000\ m^{-2}$. Figure 4(b) compares the effective impedance spectra for the rough surface obtained from the measured relative SPL with the Twersky predictions and the agreement is tolerable.

2. Pyramidal polyurethane roughness elements on a glass plate

Impedance tube measurements (Fig. 5) of the polyurethane material used to create the pyramidal roughness elements are fitted using Eq. (15) with $R_s=10$ kPa s m^{-2}, $\Omega=0.8$, and $T=1/\Omega$. The roughness elements had a roughly pyramidal/conical shape with average roughness base radius of $a=0.0125$ m. Prolate hemispheroids (see Sec. II D 2) have

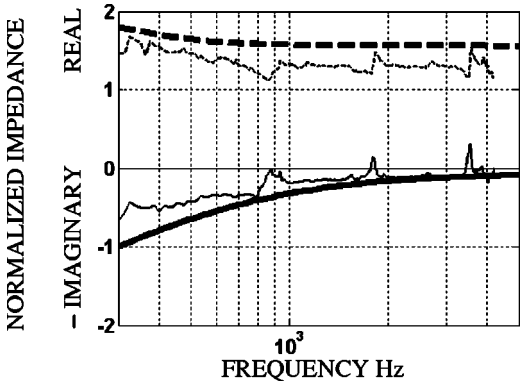


FIG. 5. Predictions (thick lines) of the tortuous slit-pore rigid frame model ($R_s=10$ kPa s m^{-2}, $\Omega=0.8$, and $T=1/\Omega$) compared to impedance spectra deduced from impedance tube measurements (thin lines) of the polyurethane material used to shape the pyramidal roughness. Real and imaginary parts are the dashed and solid lines, respectively.

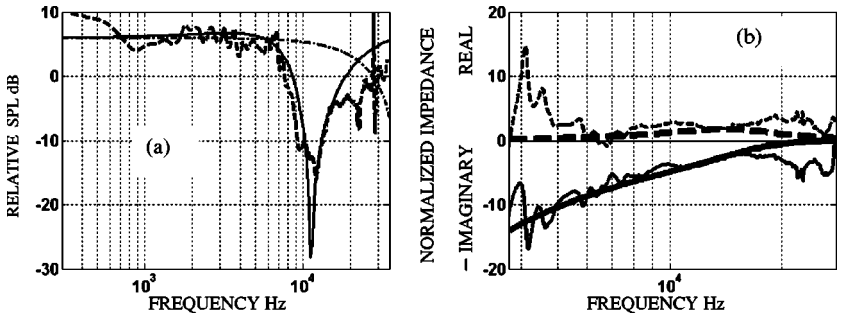


FIG. 6. (a) Averaged measurements (dashed line) of the sound pressure level relative to free field over a rough surface consisting of 450 m^{-2} pyramidal polyurethane porous roughness on a glass plate ($d=0.5$ m, $h=0.032$ m, pyramid base radius $a=0.0125$ m, 5 cm^3 roughness element volume, $b^*=2a+0.02$ m), compared to predictions from 3D Twersky theory (solid line) for prolate hemispheroidal porous bosses (all parameters nominal except $a_h=0.0153$ m, $a_w=0.0125$ m, $R_s=10$ kPa s m^{-2}, $\Omega=0.8$, and $T=1/\Omega$) and the exact solution for a flat hard surface (dot-dash line). (b) Real (dashed lines) and imaginary (solid lines) parts of deduced effective impedance spectra from measurements (thin lines) and predictions (thick lines).

been used to model the conical/pyramidal shapes as closely as possible. The average volume of the roughness element has been estimated to be 5 cm^3. This is equivalent to a hemispheroidal volume with semiaxes height and width $a_h=0.0153$ m and $a_w=0.0125$ m, respectively. These various parameter values have been used in the Twersky theory for 3D roughness [Eqs. (20) and (21)] together with the nominal measured parameters $b^*=2a+0.02$ m and $n=450\ m^{-2}$ to give the predictions shown in Fig. 6(a). These predictions compare well with the average of three series of relative SPL measurements obtained after locating the source receiver axis in three places over the rough surface. For the roughness sizes used in this section, the small ka approximation is valid up to approximately 5000 Hz. The source-receiver geometry is given by $d=0.5$ m, $h=0.032$ m. Comparison of data with the prediction for a flat hard surface confirms that the roughness causes a clear shift in the ground effect. Figure 6(b) shows that there is tolerable agreement between the effective impedance spectra obtained from the measured relative SPL and the predictions.

3. Hemispheroidal sand roughness elements on a glass plate

Impedance tube measurements (Fig. 7) of the sand used to create the hemispheroidal roughness elements are fitted using Eq. (15) with $R_s=300$ kPa s m^{-2}, $\Omega=0.23$, and $T=1/\Omega$. These parameters are used in the 3D Twersky theory to model the complex compressibility and density of the porous bosses. The roughness elements were approximately oblate hemispheroids. The measured semi-axes were $a_h=0.017$ m and $a_w=0.019$ m. The predictions (Sec. II D 2) use the nominal measured parameters $b^*=2a+0.008$ m and $n=329\ m^{-2}$ and compare well with the average measured relative SPL [see Fig. 8(a)]. The source-receiver geometry is given by $d=0.5$ m, $h=0.045$ m. For the roughness sizes used in this section, the small ka approximation is valid up to roughly 3000 Hz. Figure 8(b) shows the tolerable agreement between the effective impedance spectra obtained from the measured relative SPL and the predictions of Twersky's theory. It is interesting that the values of both the measured and predicted effective impedance near grazing incidence of the rough surface formed from 329 m^{-2}, 0.017 m high sand hemispheroids on a glass plate (i.e., 37% covered) are not only significantly less than that of a smooth hard surface but less also than those measured at normal impedance for the sand. Moreover in this case there are no adjusted parameters in the predictions. The measurement-deduced and predicted effective impedance spectra for porous roughness shown in Figs. 4(b), 6(b), and 8(b) do not show whether $\mathrm{Re}(Z)>\mathrm{Im}(Z)$ at low frequency since, at low frequency, the effective impedance spectra deduced from measurements are not reliable and display large oscillations. Whether such oscillations and enhancements are caused by surface waves is discussed elsewhere.[11] However, it should be noted that the predicted effective impedance spectra in Figs. 4(b), 6(b), and 8(b) show $\mathrm{Im}(Z)>\mathrm{Re}(Z)$ at lower frequencies. Predictions at even lower frequencies, not reported here, show even larger values of Im(Z) which is unphysical.

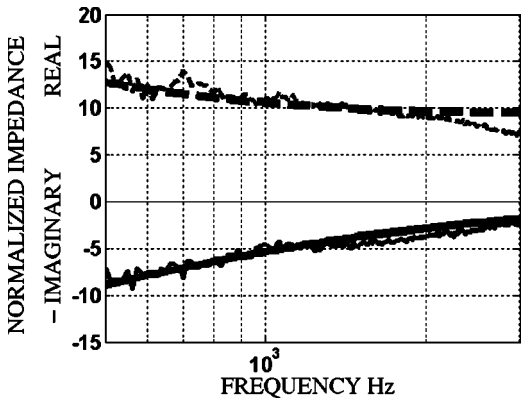


FIG. 7. Predictions (thick lines) of the tortuous slit-pore rigid frame model ($R_s=300$ kPa s m^{-2}, $\Omega=0.23$, and $T=1/\Omega$) compared to the impedance spectra deduced from impedance tube measurements (thin lines) of the sand used to create the hemispheroidal roughness. The real and imaginary parts of the impedance are the dashed and solid lines, respectively.

B. Porous rough surfaces

1. Periodic semicylindrical rough sand surface

The 2D semicylindrical roughness elements of radius $a=0.01$ m were periodically spaced with $b=0.05$ m. Figures 9(a) and (b) presents impedance deduced from data for smooth and rough sand surfaces, respectively. For both of the spectra shown in Fig. 9(a), the source/receiver separation d is 1 m and the source/receiver heights h are 0.1 and 0.2 m.

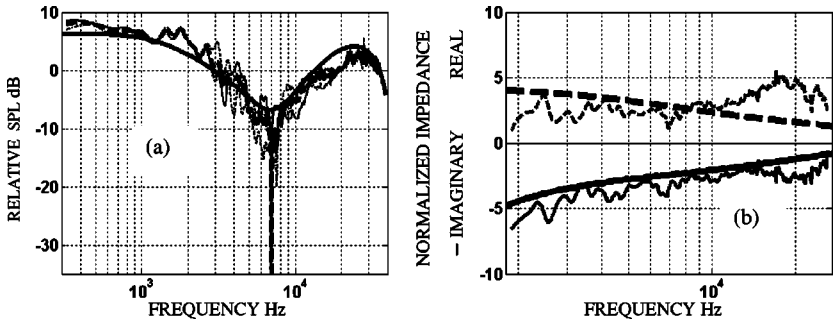


FIG. 8. (a) Three measurements of relative SPL spectra (thin lines) and their average (dashed thick line) over a rough surface consisting of 329 m^{-2} hemispheroidal sand roughness elements on a glass plate ($d=0.5$ m, $h=0.045$ m, $a_h=0.017$ m, $a_w=0.019$ m, $b^*=2a_w+0.008$ m), compared to Twersky's theory predictions (thick solid line) for oblate hemispheroidal porous bosses (all parameters have nominal values except $R_s=300$ kPa s m^{-2}, $\Omega=0.23$, and $T=1/\Omega$) on a hard plane. (b) Real (dashed lines) and imaginary parts (solid lines) of the effective impedance spectra deduced from measurements (thin lines) and model (thick lines).

Except below 500 Hz the spectra are similar and do not show any variation with source/receiver height. As noted elsewhere,[11] the difference at low frequency may result from the sensitivity of the root search to small oscillations in complex relative SPL rather than to the angle dependence of the effective impedance. Figure 9 and additional measured data not reported here show that the effective impedance spectra for rough or smooth sand surfaces deduced from relative SPL are found to be independent of grazing incidence angle in the range 0.1–0.6 rad. Figure 9(a) shows also the best visual fit of the impedance spectra deduced from measurements using Eq. (15). The assumed slit-pore medium has $R_s=600$ kPa s m^{-2}, $\Omega=0.4$, and $T=1/\Omega$. These values are used in subsequent predictions for the effective admittance of the roughened sand also, on the assumption that wetting the sand slightly did not change the porosity-related parameters. For all spectra shown in Fig. 9(b), the measurement geometry is given by $d=1$ m and $h=0.1$ m. As expected, the impedance fitted from data obtained above the smooth sand surface does not match the measurement-deduced impedance for the rough sand surface. Roughening the sand surface decreases its impedance; in particular Re(Z) is reduced. Note that Twersky's theory predicts that Re(Z) is modified by the presence of periodic porous roughness elements [through the first term in Eq. (18)]. For a periodic *hard* roughness element distribution, both terms of Eq. (18) would vanish. So this effect is not caused by nonspecular scattering. Thermal diffusion and friction in the pore roughness result in nonzero values of C_2 and

$$\mathrm{Im}\left[\frac{B-1}{1+(B-1)A_2}\right].$$

However, these factors are negligible compared to the specular scattering term. Using the nominal parameters for semicylindrical roughness in Eqs. (16)–(18) does not give correct predictions of the impedance spectra deduced from measurements. This is not surprising since, without modification, Twersky's model for porous roughness elements assumes that the portions of the sand surface between the roughness elements are acoustically hard. Even uncertainty in the form of the semicylindrical roughness elements and their spacing due to crumbling resulting, for example, in a small deviation from periodicity such that $b^*=2a+0.01$ m, does not improve the results. The tolerable fit shown in Fig. 10(a) requires that the semicylinder radius is increased to $a=0.025$ m, that R_s is decreased to 14 kPa s m^{-2} and the porosity is decreased to $\Omega=0.3$. These values are not consistent with the nominal values or with the expected values for sand. The predictions of the heuristic BTHT model [Eq. (23)] with nominal measured parameters are shown in Fig. 10(b). The agreement is excellent for Im(Z) but less good for Re(Z). Figure 11(a) shows predictions using the nominal values of

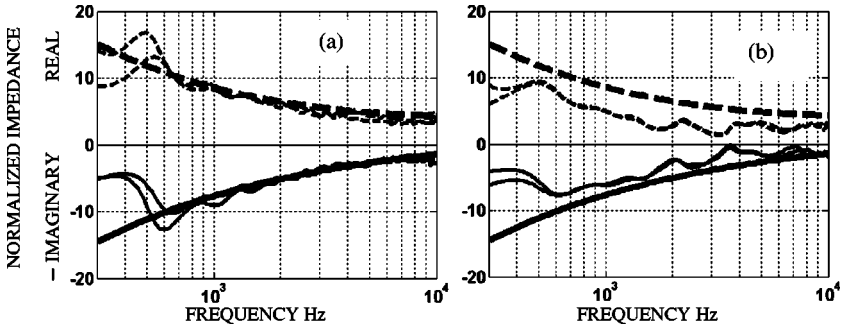


FIG. 9. Predictions of the tortuous slit-pore rigid frame model (thick lines) obtained with $R_s=600$ kPa s m^{-2}, $\Omega=0.4$, and $T=1/\Omega$ compared to the effective impedance spectra deduced from two relative SPL spectra (thin lines) ($d=1$ m) measured over (a) smooth sand (two measured sets for $h=0.1$ m and $h=0.2$ m) and (b) semicylindrical close-packed sand roughness elements (two measured sets for $h=0.1$ m, $b=0.05$ m, $W=1$, and $a=0.01$ m). The real and imaginary parts are the dashed and solid lines, respectively.

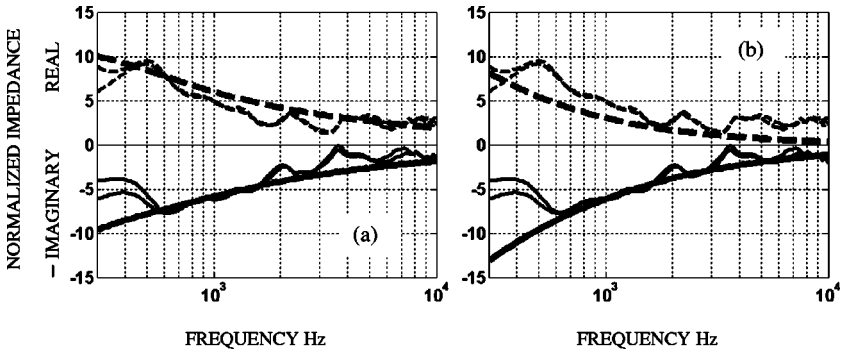


FIG. 10. Real (dashed lines) and imaginary parts (solid lines) of the effective impedance spectra obtained from measurements (thin lines) corresponding to Fig. 9(b) compared to predictions (thick lines) over semicylindrical porous roughness elements of circular cross section from (a) the 2D Twersky model—Ref. 12—for $a=0.025$ m, $b=2a$, $W=1$, $R_s=14$ kPa s m^{-2}, $\Omega=0.3$, and $T=1/\Omega$ and (b) the heuristic BTHT model with $W=1$, $a=0.01$ m, $b=0.05$ m, $R_s=600$ kPa s m^{-2}, $\Omega=0.4$, and $T=1/\Omega$.

the parameters in the heuristically extended Twersky theory for 2D porous roughness elements [Eqs. (23), (17), and (18)]. The agreement with measurements is not as good as that obtained with the heuristic BTHT model at low frequency, but it is slightly better at high frequency. Note that, for this periodic distribution of porous roughness elements and at low frequency, the real part of the admittance is predicted to be much smaller than that for a smooth porous surface (two orders of magnitude at 100 Hz). This is consistent with the wavelength ($\lambda=3.4$ m at 100 Hz) being much larger than the roughness size ($a=0.01$ m) at these low frequencies, so that the interaction of the acoustic waves with the rough surface is more like that with a smooth surface. Since the predicted real part of the admittance due to roughness is negligible, nonspecular scattering is not responsible for the decrease in Re(Z) at low frequency when the porous smooth surface is roughened. It is the specular scattering term in Eq. (17) that is mainly responsible for this measured and predicted effect. This explains the similar low frequency results obtained with the two heuristic models, despite the fact that the heuristic Twersky model for 2D porous roughness includes the nonspecular scattering term due to porous roughness elements. Nevertheless, the heuristic BTHT model predicts higher Re(Z) values at low frequency and these are closer to the measured values. This result reflects the differences in the predicted imaginary parts of admittance. Note also that the real part of admittance of the smooth porous surface [real part of inverse of Eq. (15)] and that predicted for the rough porous surface [the first term in Eq. (18)] become equal near 8000 Hz ($\lambda=0.042$ m). Thus, inclusion of the admittance $\beta_{\text{smooth}}=1/Z_{\text{rel}}$ [Eq. (15)] of the smooth porous surface in Eq. (23) avoids the unphysical (infinite imaginary admittance) low frequency limit of the Lucas and Twersky theory.

Given that the roughness shapes used in the measurements were unavoidably irregular as a result of crumbling, it seems reasonable to model the roughness elements as semielliptical rather than semicylindrical. Figure 11(b) shows predictions obtained from Eqs. (23), (16)–(18), and the expressions for semielliptical roughness elements (Sec. II C). A semiheight $a_h=0.006$ m and a semiwidth $a_w=0.009$ m gives a good fit to the measurement-deduced effective impedance. The sensitivity of both the relative SPL magnitude and effective impedance spectra to variation in the roughness parameters is similar so either relative SPL or effective impedance may be fitted first. However, it is easier to fit the measured relative SPL spectra visually because of the sharpness of the ground effect dip. Note that for the roughness sizes used in this section, the small ka approximation is valid up to roughly 5500 Hz.

2. Periodic semielliptical sand roughness

Figure 12 shows effective impedance spectra deduced from complex relative SPL measured over close-packed

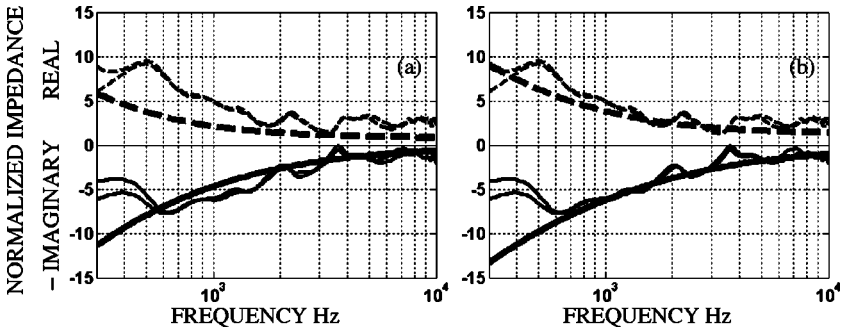


FIG. 11. Real (dashed lines) and imaginary parts (solid lines) of the effective impedance spectra obtained from measurements (thin lines) [as shown in Fig. 9(b)] compared to predictions (thick lines) from (a) the heuristic 2D Twersky model for semicylindrical porous roughness elements of circular cross section (nominal measured parameters in addition to $R_s=600$ kPa s m^{-2}, $\Omega=0.4$, and $T=1/\Omega$) and (b) the heuristic 2D Twersky model for semicylindrical porous roughness of elliptical cross section [same parameters as in (a) except for $a_h=0.006$ m and $a_w=0.009$ m].

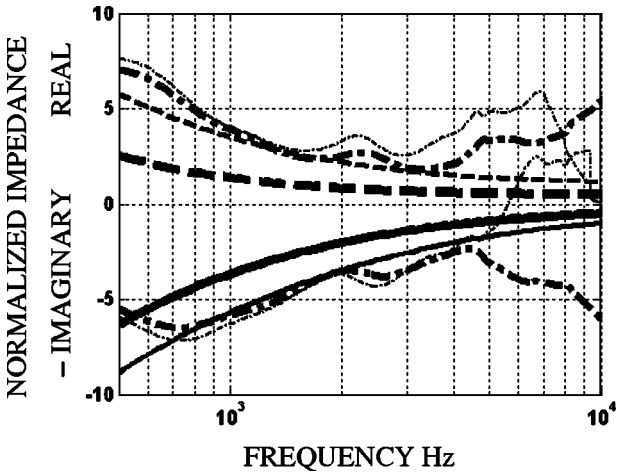


FIG. 12. Real and imaginary parts of the effective impedance spectra deduced from two measured relative SPL spectra with geometry $a_h = 0.008$ m, $a_w = 0.02$ m, $b = 2a_w$, $d = 1$ m, $h = 0.05$ m (thick dot-dash line) or $h = 0.1$ m (thin dot-dash line). Predictions of effective real (dashed lines) and imaginary parts (solid lines) of impedance predicted by the heuristic 2D Twersky model with $h = 0.05$ m and all parameters with nominal values are shown as the thick lines. Predictions for $a_h = 0.003$ m are shown as the thin lines. Measurements and predictions are for periodic semicylindrical sand roughness elements of elliptical cross section.

semielliptical sand roughness with semiheight $a_h = 0.008$ m, semiwidth $a_w = 0.02$ m, and separation $b = 2a_w$. The two geometries used in these measurements ($d = 1$ m, $h = 0.05$ and 0.1 m) display identical effective impedance spectra that seem reliable up to 4000 Hz. The variability beyond this frequency value is probably due to the impedance deduction method that shows unexplained layer-like resonance in cases reported elsewhere.[11] Predictions using Eqs. (23), (16)–(18), and the expressions for semielliptical roughness elements (Sec. II C) are presented also in Fig. 12. Note that the semiheight of the elliptical semicylinders has to be decreased from the measured 8 to 3 mm to obtain a good fit of the measured data. All other parameters are the measured nominal parameters and a prediction using nominal values is shown for comparison. Assuming that the roughness elements were not perfectly packed and that the average separation was slightly greater does not improve the fit significantly. Note that for the roughness sizes used in this section, the small ka approximation is valid up to roughly 3000 Hz.

3. Outdoor uncultivated porous soil

The effective impedance spectra deduced from measurements with source–receiver geometry $d = 1$ m, $h = 0.1$ m over an uncultivated porous soil with no vegetation are shown in Fig. 13(a) together with predictions using Eqs. (19)–(22). The best-fit parameters are $a = 0.05$ m for hemispherical roughness, $n = 15\,\mathrm{m}^{-2}$, $b^* = (2a + 0.1\,\mathrm{m})$, $R_s = 70\,\mathrm{kPa\,s\,m}^{-2}$, $\Omega = 0.4$, $T = 1/\Omega$. The R_s obtained for best fit with the original Twersky model may be smaller than the true value as a result of the fact that smooth areas between the porous roughness elements are modeled as hard (even with 3D packed distributions). Consequently, the parameter R_s has to be decreased to compensate for these hard patches. The low frequency Re(Z) limit predicted by the original Twersky model is unphysical. The k term in Re(β) is smaller than that in Im(β) as a consequence of the effects of pore-friction and thermal conductivity being smaller on Re(β) than on Im(β). However, the heuristically extended model for a 3D rough porous surface [Eq. (23)] results in a good fit to impedance spectra deduced from measurements [Fig. 13(b)] with parameter values $a = 0.03$ m, $n = 10\,\mathrm{m}^{-2}$ (which is the maximum compaction allowed by the chosen value of b^*), $b^* = (2a + 0.1\,\mathrm{m})$, $R_s = 700\,\mathrm{kPa\,s\,m}^{-2}$, $\Omega = 0.4$, and $T = 1/\Omega$. The best fit value for R_s for the heuristically extended model is rather higher than obtained with the model for porous roughness elements on a hard surface and is more realistic for a soil. Note that for the roughness sizes used in this section, the small ka approximation is valid up to roughly 1000 Hz.

V. CONCLUSIONS

Twersky's theory for the effective admittance of a surface containing porous roughness elements on a hard plane requires knowledge of the complex density and compressibility of the porous material, the roughness dimensions, the number of roughness elements per unit area n, the radius of the exclusion region around the roughness b^*, and the source–receiver geometry. Together with a tortuous slit-pore model for rigid porous material, the theory has been found to give predictions of relative SPL and effective impedance spectra data in good agreement with data obtained over various artificial surfaces in the laboratory. These surfaces include randomly distributed hemispherical polystyrene, pyra-

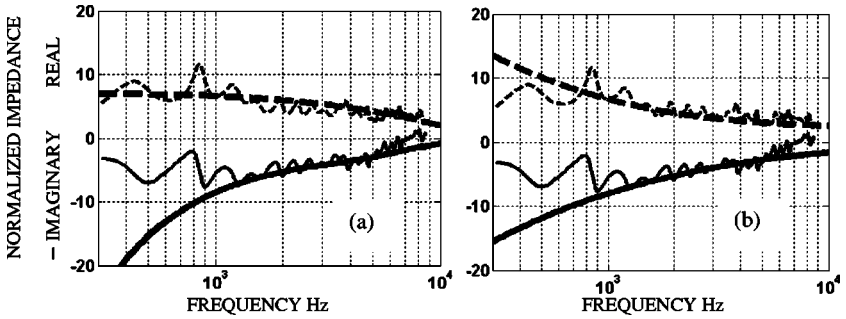


FIG. 13. Real (dashed lines) and imaginary parts (solid lines) of the effective impedance spectrum deduced from the relative SPL spectrum measured (thin lines) over a rough soil ($d = 1$ m, $h = 0.1$ m) compared to predictions (thick lines) over hemispherical porous roughness ($\Omega = 0.4$, $T = 1/\Omega$, $b^* = 2a + 0.1$ m) using (a) the original 3D Twersky theory—Ref. 12 ($a = 0.05$ m, $n = 15\,\mathrm{m}^{-2}$, $R_s = 70\,\mathrm{kPa\,s\,m}^{-2}$) and (b) the heuristic 3D Twersky theory ($a = 0.03$ m, $n = 10\,\mathrm{m}^{-2}$, $R_s = 700\,\mathrm{kPa\,s\,m}^{-2}$).

midal polyurethane bosses, and hemispheroidal sand bosses on a glass plate. For the latter surface, the effective impedance near grazing incidence has been measured and predicted to be much less than the surface impedance of smooth sand.

The theory has been extended heuristically to enable predictions of the effective impedance of rough porous surfaces. Comparison of the new Twersky-based model with the previously published BTHT model shows some agreement for high flow resistivity surfaces but less agreement for large roughness elements. The new heuristic model has been shown to give predictions in relatively good agreement with data for relative SPL and measurement-deduced effective impedance spectra obtained over corrugated sand in the laboratory and an uncultivated soil outdoors.

ACKNOWLEDGMENTS

This investigation has been carried out under a contract awarded by the European Commission, ''Sonic Boom European Research Program: Numerical and Laboratory-Scale Experimental Simulation'' (SOBER), Contract No. G4RD-CT-2000-00398.

[1] K. Attenborough and S. Taherzadeh, ''Propagation from a point source over a rough finite impedance boundary,'' J. Acoust. Soc. Am. **98**, 1717–1722 (1995).
[2] J. P. Chambers, R. Raspet, and J. M. Sabatier, ''Incorporating the effects of roughness in outdoor sound propagation models,'' Noise Con, 1996.
[3] Y. Berthelot and J. P. Chambers, ''On the analogy between sound propagation over a rough surface and sound propagation over a smooth surface with modified surface impedance,'' Seventh International Symposium on Long-Range Sound Propagation, Lyon, France, 1996.
[4] J. P. Chambers and Y. H. Berthelot, ''An experimental investigation of the propagation of sound over a curved rough, rigid surface,'' J. Acoust. Soc. Am. **102**, 707–714 (1997).
[5] J. P. Chambers, J. M. Sabatier, and R. Raspet, ''Grazing incidence propagation over a soft rough surface,'' J. Acoust. Soc. Am. **102**, 55–59 (1997).
[6] I. Tolstoy, ''Coherent sound scatter from a rough interface between arbitrary fluids with particular reference to roughness element shapes and corrugated surfaces,'' J. Acoust. Soc. Am. **72**, 960–972 (1982).
[7] I. Tolstoy, ''Smoothed boundary conditions, coherent low-frequency scatter, and boundary modes,'' J. Acoust. Soc. Am. **75**, 1–22 (1984).
[8] R. J. Lucas and V. Twersky, ''Coherent response to a point source irradiating a rough plane,'' J. Acoust. Soc. Am. **76**, 1847–1863 (1984).
[9] P. Boulanger, K. Attenborough, S. Taherzadeh, T. Waters-Fuller, and K. M. Li, ''Ground effect over hard rough surfaces,'' J. Acoust. Soc. Am. **104**, 1474–1482 (1998).
[10] K. Attenborough and T. Waters-Fuller, ''Effective impedance of rough porous ground surfaces,'' J. Acoust. Soc. Am. **108**, 949–956 (2000).
[11] P. Boulanger and K. Attenborough, ''Effective impedance spectra for predicting rough sea effects on atmospheric impulsive sounds,'' J. Acoust. Soc. Am. (to be published).
[12] V. Twersky, ''Reflection and scattering of sound by correlated rough surfaces,'' J. Acoust. Soc. Am. **73**, 85–94 (1983).
[13] IMSL Math Library User's Manual, Version 3.0, Visual Numerics, Inc., Houston, TX, 1994.
[14] S. Taherzadeh and K. Attenborough, ''Deduction of ground impedance from measurements of relative SPL spectra,'' J. Acoust. Soc. Am. **105**, 2039–2042 (1999).
[15] J. F. Allard, *Propagation of Sound in Porous Media* (Elsevier Applied Science, New York, 1993).
[16] K. Attenborough, ''Models for the acoustical properties of air-saturated granular media,'' Acustica **1**, 213–226 (1993).
[17] H. Lamb, *Hydrodynamics* (Dover, New York, 1945), p. 85.
[18] M. S. Howe, ''On the long range propagation of sound over irregular terrain,'' J. Sound Vib. **98**, 83–94 (1985).
[19] C. F. Chien and W. W. Soroka, ''Sound propagation along an impedance plane,'' J. Sound Vib. **43**, 9–20 (1975).

Using listening difficulty ratings of conditions for speech communication in rooms

Hiroshi Sato[a)]
Institute for Research in Construction, National Research Council, Ottawa, K1A 0R6 Canada

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

Masayuki Morimoto
Environmental Acoustics Laboratory, Faculty of Engineering, Kobe University, Rokko, Nada, Kobe 657-8501 Japan

(Received 23 March 2004; revised 2 November 2004; accepted 24 November 2004)

The use of listening difficulty ratings of speech communication in rooms is explored because, in common situations, word recognition scores do not discriminate well among conditions that are near to acceptable. In particular, the benefits of early reflections of speech sounds on listening difficulty were investigated and compared to the known benefits to word intelligibility scores. Listening tests were used to assess word intelligibility and perceived listening difficulty of speech in simulated sound fields. The experiments were conducted in three types of sound fields with constant levels of ambient noise: only direct sound, direct sound with early reflections, and direct sound with early reflections and reverberation. The results demonstrate that (1) listening difficulty can better discriminate among these conditions than can word recognition scores; (2) added early reflections increase the effective signal-to-noise ratio equivalent to the added energy in the conditions without reverberation; (3) the benefit of early reflections on difficulty scores is greater than expected from the simple increase in early arriving speech energy with reverberation; (4) word intelligibility tests are most appropriate for conditions with signal-to-noise (S/N) ratios less than 0 dBA, and where S/N is between 0 and 15-dBA S/N, listening difficulty is a more appropriate evaluation tool. [DOI: 10.1121/1.1849936]

PACS numbers: 43.55.Hy, 43.71.Gv [NX] Pages: 1157–1167

I. INTRODUCTION

There are many situations in rooms where the results of speech intelligibility tests would suggest that conditions are reasonably acceptable, with intelligibility scores of 90% or greater and signal-to-noise ratios (S/N) above 0 dBA. Such conditions are actually very common. For example, Pearsons[1] found speech-to-noise ratios in public spaces varying from 0 dBA in an aircraft cabin to 15 dBA in a classroom. It is easy to appreciate that conditions of 0-dBA S/N are quite different than those with a 10- or 15-dBA S/N, even though word intelligibility scores for both conditions are similar and quite close to 100%. Such results disguise the fact that in these conditions, speech intelligibility is only possible with a great amount of extra effort by the listener. In fact, it is quite difficult to understand speech in many of these conditions and it seems incorrect to suggest that they represent good conditions for speech communication.

The authors have previously developed the use of subjective ratings of listening difficulty as a better indicator of the quality of acoustical conditions in rooms for speech communication[2] using speech tests in Japanese. Listening difficulty ratings were better able to discriminate among conditions above 0-dB S/N than could speech intelligibility scores. Several other approaches have been used previously to assess speech communication performance instead of speech recognition scores. For example, subjective ratings of the ''easiness'' of speech recognition or ''ease of listening'' were considered.[3–5] Because ''ease'' ratings were measured using paired comparison tests and category scaling methods, they result in a relative scale, and one cannot say how easy is good enough, or when a low rating corresponds to unacceptably bad acoustical conditions. Apoux[6] rated ''ease of listening'' with reaction time for consonants and other studies have measured ''listening effort'' such as those by Downs,[7] and Hicks and Tharpe[8] using a dual-task paradigm. In these studies, the probe reaction time was measured after the word recognition task. The reaction time was treated as listening effort. This is suitable as a clinical technique to assess the effectiveness of hearing-aid devices and/or the application of signal-processing techniques to such devices. However, it is difficult to measure reaction time for groups of typical listeners in real rooms, the focus of this study. This type of measure also has the problem that there is no reaction time value that can be said to be good enough or that corresponds to no listening difficulty.

In rooms, early-arriving reflections have been shown to be particularly important for good speech communication. Lochner and Burger[9] and others have provided a solid basis for the importance of such early reflections. However, con-

[a)]Presently at Institute for Human Science & Biomedical Engineering, National Institute of Advanced Industrial Science and Technology, 1-1-1 Higashi, Tsukuba, Ibaraki 305-8566, Japan. Electronic mail: sato.hiro@aist.go.jp

ventional room acoustics design is still usually based only on consideration of appropriate reverberation times.[10] A recent study by the authors demonstrated the expected improvements to speech intelligibility scores due to early reflections and indicated that in realistic situations early reflection energy in real rooms is equivalent to increasing the level of the direct sound by up to 9 dB.[11] That is, the effective useful speech level could be as much as 9 dB greater than the direct sound due to the beneficial effects of early reflections of speech sounds. Of course, later-arriving speech sounds are usually found to be detrimental to the intelligibility of speech.[9]

The present study examines the use of listening difficulty ratings to rate speech transmission performance as details of the sound-field components were varied. This included examining whether the benefits of early-arriving reflections to listening difficulty ratings are similar to their effect on speech intelligibility scores. This was done using two different types of tests to assess word intelligibility and perceived difficulty of listening to speech in simulated sound fields that were similar to those used in a previous study.[11] Although sound fields can be described by the complete details of impulse responses, the perceptually important aspects can be more simply understood by considering the speech levels associated with three basic components: the direct sound, the early reflections, and the later-arriving reverberant speech sound. In this new work, sound fields were varied by varying each of these components as a group and without changing the detailed make-up of each group such as the details of individual early reflections. It was intended that the effects of changes to each of these three component groups would be indicative of the effects of similar changes in a wide range of rooms.

The first experiment used sound fields that included either, only direct sound (D), or direct sound with early reflections ($D+E$), and combined with two different levels of steady ambient noise. The second experiment used three types of sound fields: direct sound only (D), direct sound with reverberation ($D+\text{Rev.}$), and direct sound with early reflections and reverberation ($D+E+\text{Rev.}$), all with a constant level of ambient noise. Additionally, paired comparison tests were used to confirm the significance of the differences among some of the sound fields in the second experiment.

The main goal of the new work reported in this paper was to confirm and extend previous results[2] that indicated listening difficulty is a better rating of the quality of conditions for speech communication than word recognition test scores in the range of S/N that occurs most frequently in actual rooms. This new work extends the previous work by using native and non-native speakers with a range of language skills and used speech tests in English rather than Japanese. It also explored the effects of systematic variations in the sound fields representative of conditions commonly found in real rooms including varied levels of early-arriving reflections. The new work focuses on conditions with S/N of 0 dBA or greater, whereas our previous work[11] considered the benefits of early reflections on intelligibility scores in more adverse conditions with S/N of 5 dBA and less.

II. LISTENING TESTS WITH AMBIENT NOISE AND VARIED EARLY REFLECTIONS

A. Method

1. Sound-field simulation procedures

All simulated sound fields were produced using a seven-channel electro-acoustic system with loudspeakers arranged around the listener in an anechoic room at the Institute for Research in Construction, National Research Council Canada. The seven loudspeakers were located at a distance of 1.7 m from the listener and at varied angular locations relative to the listener to simulate early reflections from various angles. Each of the seven channels of electronics included programmable digital equalizers with time delays that could all be changed under computer control via a MIDI interface. The loudspeaker responses were corrected to be flat ± 3 dB from 80 Hz to 12 kHz.

The loudspeaker located directly in front of the listener produced the simulated direct sound (first arriving sound). The other six loudspeakers each produced one early reflection. The early reflections arrived at the listener within the first 50 ms after the direct sound. Two conditions of the early reflections were used: one in which the early reflections increased the long-term averaged speech level by 3 dB, and the other by 6 dB. Some sound fields included only a direct sound component (D cases); others included a direct sound and early reflections ($D+E$ cases). The overall amplitudes of each of the two component groups (i.e., direct sound or early reflections) were varied, but the arrival times and relative amplitudes of individual early reflections were not changed and such details are not considered in these experiments.

Each loudspeaker also reproduced simulated ambient noise with a spectrum shape corresponding to that of an NCB40 contour.[12] The combined level of the noise signals from all seven loudspeakers was measured at the listener to be 48.4 dBA. A second noise signal with the same spectrum shape but with an overall level of 45.0 dBA was also used in the experiment.

In the first experiment, subjects were exposed to only one noise level during each of two experimental sessions. The noise signals to each loudspeaker included varied time delays so that they were not exactly coherent. The speech and noise levels used in the first experiment are summarized in Table I. All speech and noise levels were obtained with an omnidirectional measurement microphone located at the center of the listener's head position (but without the listener present). This procedure was used because it can be related to typical measurements in rooms.

2. Subjects, speech material, and procedure for the listening test

Subjects varied from 22 to 58 years of age, with nine males and five females, and they reported no hearing disabilities. Although they all spoke English everyday, for seven of the subjects English was their first language, but for the other seven subjects English was not their first language.

The speech material used in the experiments was from the Fairbanks rhyme test as modified by Latham[13] and as

TABLE I. Summary of measured acoustical quantities for D (direct only) cases and $D+E$ (direct+early reflections) cases.

Direct speech level, dBA	Early speech level, dBA	Noise level, dBA	Total speech level, dBA	Total S/N, dB	Direct speech level, dBA	Early speech level, dBA	Noise level, dBA	Total speech level, dBA	Total S/N, dB
		D cases					D cases		
42.0	⋯	45.0	42.0	−3.0	42.0	⋯	48.4	42.0	−6.4
47.2	⋯	45.0	47.2	2.2	47.2	⋯	48.4	47.2	−1.2
51.2	⋯	45.0	51.2	6.2	51.2	⋯	48.4	51.2	2.8
53.9	⋯	45.0	53.9	8.9	53.9	⋯	48.4	53.9	5.5
57.1	⋯	45.0	57.1	12.1	57.1	⋯	48.4	57.1	8.7
62.1	⋯	45.0	62.1	17.1	62.1	⋯	48.4	62.1	13.7
		$D+E$ cases					$D+E$ cases		
42.0	42.4	45.0	45.2	0.2	42.0	42.4	48.4	45.2	−3.2
42.0	47.3	45.0	48.4	3.4	42.0	47.3	48.4	48.4	0.0
47.2	47.2	45.0	50.2	5.2	47.2	47.2	48.4	50.2	1.8
47.2	52.3	45.0	53.5	8.5	47.2	52.3	48.4	53.5	5.1
51.2	51.2	45.0	54.2	9.2	51.2	51.2	48.4	54.2	5.8
51.2	55.9	45.0	57.2	12.2	51.2	55.9	48.4	57.2	8.8

used in previous tests.[11] The test words were embedded in the sentence ''Word number_ is _, write that down'' and were spoken by a male talker. The original test lists consisted of 5 lists of 50 words, phonetically balanced within each list. In this test, each of the 250 words was used separately. Each subject listened to a total of 288 words for the combinations of 12 reflection conditions, by 2 noise levels and 12 repeats of each of these combinations. For each subject, 38 words were used twice and the repeated words were picked randomly from the 250 words. Each word was presented in a random order for each subject, and the test conditions were also presented in random order. Subjects had a break every 24 words. Each sentence was presented in a 3-s interval to the subject.

Subjects first performed a speech recognition test by writing down the first letter of each test word and then they rated the listening difficulty of each test sentence, including the target word, using the following four categories:

(0) Not difficult: no effort required, completely relaxed listening condition
(1) Slightly difficult: slight attention required
(2) Moderately difficult: moderate attention required
(3) Very difficult: considerable attention required

The authors used these same categories for listening difficulty rating in a previous study for Japanese speech.[2]

Listening difficulty was judged immediately after the speech recognition test component so that listening difficulty included the effect of the cognitive process of word recognition on this subjective rating.

Word recognition scores were obtained for each subject as the average score of the repeated tests for each condition. Listening difficulty was obtained for each subject as the percentage of responses that indicated some level of difficulty (i.e., not a ''0'' response) for each condition. This makes it possible to identify desirable conditions as those in which listeners have no difficulty in listening to the test speech material.[2] Both types of scores are presented as mean scores over all subjects.

B. Results of listening test to speech

1. The relation between listening difficulty and word intelligibility

The first comparisons were based on the results of tests in which subjects performed speech intelligibility tests and listening difficulty ratings for sound fields with varied speech signal-to-noise ratio (S/N) and for two types of reflection conditions. In one series of tests the sound fields consisted of only a direct sound and varied S/N was obtained by varying the amplitude of the direct speech sound in combination with either 45.0 or 48.4 dBA of constant noise. In the other series of tests, three levels of direct speech sound were used and the S/N was varied by adding early reflections to increase the total speech levels by 3 or 6 dBA relative to the direct speech levels. These were all presented in combination with the same two constant noise levels. Figure 1 and Fig. 2 show the

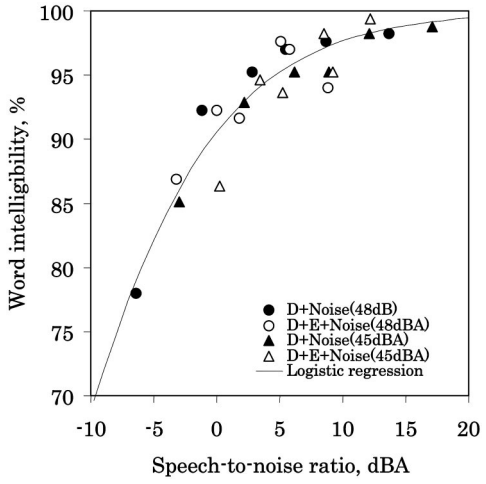


FIG. 1. Mean intelligibility scores for each sound field condition for all listeners for D (direct only) cases (filled symbols) and $D+E$ (direct + early reflections) cases (open symbols) in 48-dBA noise (circles) and 45-dBA noise (triangles). Logistic regression curve for all data is also presented.

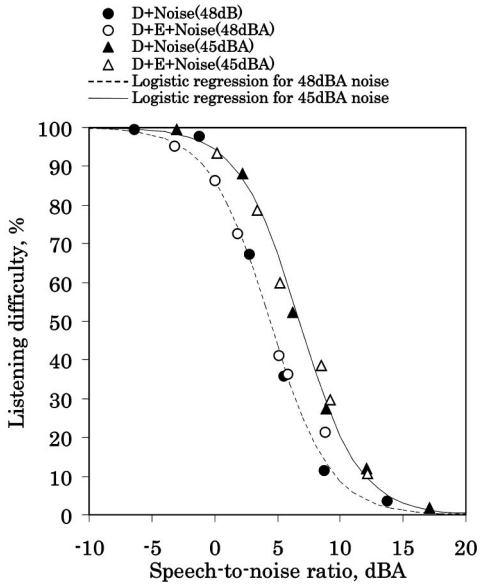


FIG. 2. Mean listening difficulty ratings of each sound-field condition for D (direct only) cases (filled symbols) and $D+E$ (direct+early reflections) cases (open symbols) in 48-dBA noise (circles) and 45-dBA noise (triangles). Logistic regression curves for all data with 45-dBA noise (solid line) and of 48-dBA noise (dashed line) are also presented.

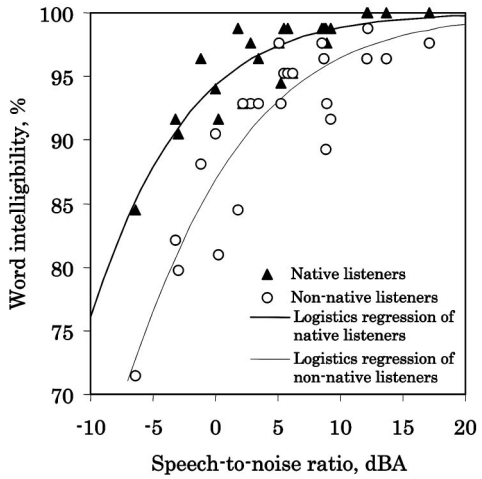


FIG. 3. Mean intelligibility scores of native English speakers (gray triangle) and non-native English speakers (open circle) for all sound-field conditions. Logistic regression curves are included for the native English speakers (solid line) and for the non-native English speakers (dashed line).

resulting plots of word recognition scores and listening difficulty ratings versus S/N, respectively. Both ratings show strong relationships with S/N.

Although word recognition scores exceed 90% for S/N greater than 0 dBA, listening difficulty is about 90% at a S/N of 0 dBA and shows the greatest variation for S/N above 0 dBA. Listening difficulty decreases monotonically as S/N varies from -2.5 to 15 dBA. In other words, listening difficulty varies over a range of 97% (i.e., from 2% to 99%) for the experimental conditions. However, word recognition scores varied over a range of only 21% (i.e., from 78% to 99%) and are seen to be a less sensitive rating of these acoustical conditions.

Analysis of variance (ANOVA) was employed to compare the sensitivity of word recognition and listening difficulty scores. The experimental conditions and individual differences were the two factors tested by the repeated measures ANOVA. The result of ANOVA for word recognition scores indicated that there was a significant effect of conditions ($p<0.0001$) and there was also a significant effect of individual subject differences ($p<0.0001$). Tukey's honestly significant difference (HSD) test for multiple comparisons[14] was employed, and it indicated that differences between conditions of more than 9.12% would be significant. A difference of 9.12% is 42.6% of the complete range of the speech recognition scores.

The ANOVA results for the listening difficulty ratings indicate that there are significant effects of varied condition ($p<0.0001$) and also of the differences among subjects ($p<0.0001$). HSD is 22.8% ($p<0.05$), which is 23.4% of the complete range of listening difficulty ratings, and this is about half of that for word intelligibility. This HSD shows that the listening difficulty ratings have approximately double the sensitivity of the word intelligibility scores.

There was no pair of conditions, with and without early reflections and both having almost the same speech-to-noise ratio, which had difficulty ratings different by 23.4% or more. This indicates that, for listening difficulty ratings, adding early reflections is equivalent to increasing the energy of the direct sound to the same total speech level. In other words, early reflections are equivalent to increased direct speech energy in these conditions without reverberant speech sound.

2. Individual differences among subjects

Because significant differences among individuals were found, ANOVA was employed to test for differences between the subject groups of native and non-native English speakers, all of whom used English every day. Each group had seven subjects. The experimental conditions and subject groups were the two factors included in the ANOVA analysis. The interaction between conditions and subject groups was not significant either for word recognition score or for listening difficulty.

The result of the ANOVA for word recognition scores indicated that there was a significant difference between the two subject groups ($p<0.0001$). Figure 3 shows that there are clearly different trends with respect to S/N for the word intelligibility scores for each subject group. The regression curves in Fig. 3 indicate that the difference in S/N between the two subject groups was 5.7 dB for a 90% word intelligibility score. This means that non-native English speakers had a 5.7-dB disadvantage relative to native English speakers in terms of speech recognition scores. This is comparable to results by Buus *et al.*,[15] who found non-native listeners to require between 3- and 12-dB better conditions depending on their amount of experience with the language of the test. Somewhat similar results were reported by Nabelek and Donahue[16] for varied reverberation time and by van Wijngaarden *et al.*[17] for Dutch subjects listening to both English and German speech material.

On the other hand, for listening difficulty ratings, the ANOVA analysis indicated that there was not a significant

difference between the two subject groups ($p>0.05$).

One can conclude that although word intelligibility and listening difficulty ratings vary among subjects, listening difficulty is not affected by the mother tongue of the subjects. Thus, listeners of varied language skills can be used to provide similar listening difficulty ratings of conditions for speech. However, these ratings would not reflect the expected lower intelligibility scores from listeners less familiar with the language of the test.

3. Effect of noise level on listening difficulty

The difference between the two logistic regression curves (where difficulty is 50%) in Fig. 2 indicates that the effective S/N difference between the two sets of results was only 1.7 dBA. This 1.7-dB difference is only half of the actual difference of noise levels and would correspond to a 20%–30% difference in listening difficulty ratings. This difference is approximately the same as the HSD ($p<0.05$) required to indicate a significant difference between a pair of conditions for listening difficulty. It is likely that this difference is due to context effects in the experiment. Although it was hoped that subjects would judge listening difficulty absolutely, they may have tended to rate conditions relative to the complete range of conditions to which they were exposed in each part of the experiment. In particular, the cases for each noise level were presented as two subgroups in which only one ambient noise level was experienced in each subgroup. This led to a slightly different range of S/N for the two parts of the experiment, and is probably the cause of the 1.7-dB shift between the two sets of listening difficulty rating results.

ANOVA was employed to test the significance of the difference between the results for ambient noise of 48.4 dBA and those of 45.0 dBA. Noise level, speech level, and the interaction between noise level and speech level were included as factors in the ANOVA. The result of the ANOVA was that there was only a marginally significant difference between the two ambient noise conditions ($p<0.1$).

Even though the difference was not statistically significant at $p<0.05$, Fig. 2 suggested that possible context effects should be minimized in subsequent experiments. Accordingly, the second experiment, discussed in the next section, included a wider range of conditions and they were all presented in a single experimental listening session.

III. LISTENING TESTS FOR SPEECH IN AMBIENT NOISE AND REVERBERATION TO CONFIRM THE BENEFIT OF EARLY REFLECTIONS

A. Method

1. Sound-field simulation procedures

All simulated sound fields were produced using a seven-channel electro-acoustic system with loudspeakers arranged around the listener in an anechoic room similar to that described for the first experiments. The seven loudspeakers were located at a distance of 1.6 m from the listener. The delayed early reflections and reverberation were created using programmable digital signal processors (Yamaha DME32) that could be changed under computer control via a MIDI interface. The loudspeaker responses were corrected to be flat ± 1 dB from 80 Hz to 12 kHz.

The loudspeaker located directly in front of the listener produced the simulated direct sound and the other six loudspeakers each produced one early reflection of the speech sounds. The early reflections arrived at the listener with varied delays to distribute them in a realistic manner over the first 50 ms after the direct sound. The details of individual reflections were not varied; only the overall amplitude of the six early reflections was varied as a group. Two levels of the early reflections were used in the experiment, one of which increased the long-term averaged speech level relative to that of the direct sound by 3 dB, and the other by 6 dB. Reverberant speech was produced via all seven loudspeakers with slightly different levels and delay times for each speaker to create a diffuse impression for the reverberant speech component.

Each loudspeaker also reproduced simulated ambient noise with a spectrum shape corresponding to that of an NCB40 contour and with a measured overall level at the position of the listener of 48.6 dBA. The noise signals to each loudspeaker were not exactly coherent to minimize interference effects at the listener and to create the impression of a diffuse sound field for the simulated ambient noise.

There were three series of conditions, which are described in Table II. In one series the sound fields consisted of only a direct sound (D cases) and varied S/N was obtained by varying the amplitude of the direct speech sound relative to the constant level of ambient noise. In the second series, the sound fields consisted of a direct sound and two levels of reverberation ($D+\text{Rev.}A$ or $D+\text{Rev.}B$ cases). The reverberation time was 1.1 s for both reverberant cases, but the ratio of early to late arriving speech sound (C50) varied as described in Table II. The reverberant speech level was 51.6 dBA for the more reverberant case called ''Rev.*A*'' and 45.8 dBA for the less reverberant case called ''Rev.*B*.'' There were four levels of direct sound for each reverberant case, increasing in 3-dBA steps from 49 dBA. In the third series ($D+E+\text{Rev.}A$ or $D+E+\text{Rev.}B$ cases), two levels of early reflections, which increased the effective signal level by 3 and 6 dBA, were added to the 49-dBA direct sound level condition with each of ''Rev.*A*'' and ''Rev.*B*'' late-arriving energy and were compared with cases which had the same effective signal level. The overall amplitudes of each of the three component groups (direct sound, early reflections, and reverberant sound) were varied but the arrival times and relative amplitudes of individual early reflections were not changed.

The measured levels of each component group are summarized in Table II for each of the test conditions. The levels of the direct, early, and reverberant speech sounds are listed where appropriate. This table also lists the effective speech levels consisting of the sum of the direct sound and the early reflection energy arriving within 50 ms after the direct sound, along with the corresponding effective signal-to-noise ratios (E-S/N).

To avoid context effects for a limited series of conditions, as discussed in the previous section, subjects experi-

TABLE II. Summary of measured acoustical quantities for D (direct only) cases, $D+$Rev (direct+reverberant sound) cases and $D+E+$Rev. (direct +early reflections+reverberant sound) cases. The ambient noise level was fixed at 48.6 dBA for all cases listed in this table.

Direct speech level, dBA	Early speech level, dBA	Reverb speech level, dBA	Effective speech level, dBA	Total speech level, dBA	Effective S/N, dB	Total S/N, dB	C50, dBA	U50, dBA	STIr (male)	RT (0.5–1 kHz), s
					D cases					
42.7	⋯	⋯	42.7	42.7	−5.9	−5.9	⋯	−5.9	0.29	⋯
45.7	⋯	⋯	45.7	45.7	−2.9	−2.9	⋯	−2.9	0.39	⋯
48.7	⋯	⋯	48.7	48.7	0.1	0.1	⋯	0.1	0.49	⋯
51.7	⋯	⋯	51.7	51.7	3.1	3.1	⋯	3.1	0.59	⋯
54.7	⋯	⋯	54.7	54.7	6.1	6.1	⋯	6.1	0.68	⋯
57.7	⋯	⋯	57.7	57.7	9.1	9.1	⋯	9.1	0.78	⋯
60.7	⋯	⋯	60.7	60.7	12.1	12.1	⋯	12.1	0.88	⋯
63.7	⋯	⋯	63.7	63.7	15.1	15.1	⋯	15.1	0.94	⋯
					$D+$Rev. cases					
49.4	⋯	51.6	49.4	53.6	0.8	5.0	−2.1	−3.8	0.38	1.1
52.4	⋯	51.6	52.4	55.0	3.8	6.4	0.8	−0.9	0.45	1.1
55.3	⋯	51.6	55.3	56.9	6.7	8.3	3.8	2.1	0.54	1.1
58.3	⋯	51.6	58.3	59.2	9.7	10.6	6.8	5.1	0.64	1.1
49.3	⋯	45.8	49.3	50.9	0.7	2.3	3.7	−0.2	0.43	1.1
52.3	⋯	45.8	52.3	53.2	3.7	4.6	6.7	2.7	0.53	1.1
55.3	⋯	45.8	55.3	55.8	6.7	7.2	9.7	5.7	0.62	1.1
58.3	⋯	45.8	58.3	58.6	9.7	10.0	12.7	8.7	0.72	1.1
					$D+E+$Rev. cases					
49.3	49.4	51.6	52.4	55.0	3.8	6.4	0.8	−0.9	0.44	1.1
49.3	54.0	51.6	55.3	56.8	6.7	8.2	3.7	1.8	0.51	1.1
49.3	49.4	45.8	52.4	53.2	3.8	4.6	6.7	2.7	0.50	1.1
49.3	53.9	45.8	55.2	55.7	6.6	7.1	9.6	5.4	0.58	1.1

enced a full range of experimental conditions in one single test. That is, cases varied from near-zero listening difficulty (S/N of 15 dBA) to nearly complete (100%) listening difficulty (S/N of −6 dBA).

2. Subjects, speech material, and procedure for the listening test

Eleven male subjects and two female subjects were used for the experiment. Subjects varied from 21 to 58 years of age and they didn't report any hearing disabilities. Two non-native English speakers were used in this experiment but their scores on the word intelligibility test were not included in the results.

The speech material used in this experiment was the same as in the previous experiment. Each subject listened to 10 sentences for each of 20 conditions and they heard each test sentence only once. Test words and test conditions were presented in random order to each subject. Subjects could have a small break every 25 words. Each sentence was presented after the subject responded to the former sentence. Hence, the rate of presentation depended on the subject.

The experimental procedure was almost the same as in the previous experiment except a small laptop PC with a 6 in. screen and a full-sized keyboard, positioned on the knees of the subject, was used instead of pen and paper. Subjects were asked to use the keyboard to give the missing first letter of each test word in the speech recognition part of the test. After listening to each sentence, they also rated the listening difficulty by typing the number for the category of difficulty.

To minimize individual differences in the results, subjects were informed about the procedure of the experiment and the definition of listening difficulty in a training session. In this practice session, subjects listened to more than the full range of experimental conditions (S/N was systematically varied in 3-dB steps from +18 to −9 dB and then in reverse order from −9 to +18 dB varying only the direct sound and ambient noise). They responded to the word recognition test and gave a listening difficulty rating to get them used to the idea of listening difficulty and to the complete range of conditions. All of the subjects switched their responses from "not-difficult" to "difficult" and vice versa in the middle of the range of S/N in this training session.

B. Results of the listening tests

1. ANOVA for comparing sensitivity of word intelligibility and listening difficulty

The result of the ANOVA for word recognition scores (the two non-native English speakers were removed) showed that there was a significant effect of varied condition ($p<0.0001$) but there was not a significant effect of individual subject differences ($p>0.05$). Tukey's honestly significant difference (HSD) test was employed for multiple comparisons. HSD was calculated to be 6.88% ($p<0.05$), which is 42.0% of total range of the scores.

The result of the ANOVA for listening difficulty indicated that there was a significant effect of varied condition ($p<0.0001$). There was also a significant effect of differences among individual subjects ($p<0.0001$). HSD for listening difficulty was 24.8%($p<0.05$), which is 25.1% of the complete range of listening difficulty.

Although there is a significant effect of the differences among individual subjects, the HSD is a smaller portion of the complete range of listening difficulty scores than is the HSD for the word intelligibility scores. This is true even

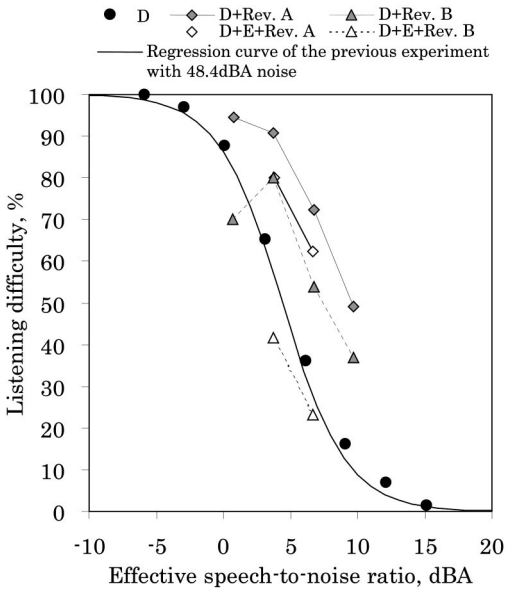


FIG. 4. Relation between effective speech-to-noise ratio and mean difficulty rating of *D* (direct sound) cases (filled circle), *D*+Rev. (direct sound +reverberant sound) cases (gray symbols), and *D*+*E*+Rev. (direct sound +early reflections+reverberant sound) cases (open symbols). The logistic regression curve for the 48.4-dBA-noise case of the previous experiment is also presented.

though there is not a significant effect on word recognition scores of individual differences among subjects when the mother tongue effect was removed.

This again shows that listening difficulty can better discriminate among these experimental conditions than can word recognition scores.

2. Comparison of listening difficulty results with those in the previous experiment

Figure 4 presents the relation between listening difficulty ratings and the effective speech-to-noise ratio (E-S/N), in which the energy of direct sound and the early reflections within 50 ms of direct sound are summed as the effective speech energy. Figure 4 also includes the regression curve obtained from the results of conditions with 48.4-dBA noise in the previous experiment.

The cases with only a direct sound component (filled circles in Fig. 4) are close to the regression curve from the former experiment. Additionally, listing difficulty ranged over almost the full scale from 1.5% to 100% as intended to avoid context effects. This result demonstrates the repeatability of the listening difficulty measure, at least for the equivalent conditions in the previous experiment.

3. Benefit of early reflections in noise and reverberation for listening difficulty

The series of conditions in this experiment were created to confirm that the effect of early reflections also exists in more realistic cases that also included later-arriving speech sounds (reverberation).

The results in Fig. 4 show that the reverberant level influences listening difficulty ratings. The more reverberant Rev.*A* cases created more difficult listening conditions than the less reverberant Rev.*B* cases at the same E-S/N.

The lowest S/N (0.7 dB) case for Rev.*B* without early reflections (filled triangles) deviates from the overall trend of these results and showed lower difficulty for this S/N value than expected. The difference between the lowest case and the second lowest case is smaller than the HSD and is not statistically significant ($p>0.05$). In a separate test, five subjects compared these two conditions ten times and 90% of their responses were the opposite of the result in the main experiment as was expected. All of the subjects who participated in this trial reported that there was a subtle difference between the two conditions and it might relate to a difference in loudness. In all other cases, adding early reflections increased the effective S/N and decreased the resulting listening difficulty rating. This one unusual point was thought to be due to the large scatter in listening difficulty scores; this was verified in a paired comparison test described in next section.

From Fig. 4, it is seen, that for sound fields including Rev.*A* with the direct sound (filled diamonds), the listening difficulty ratings are about the same as those for the direct sound only cases having a 5-dB lower S/N. Adding the reverberant speech level from Table II (51.6 dB for Rev.*A*) to the noise level (48.6 dB) increases the total detrimental sound level by about 5 dB relative to the noise alone. This 5-dB increase in detrimental sound level would relate to a 5-dB decrease in useful-to-detrimental sound ratios between the direct-sound-only cases and the direct-plus-reverberant cases, and supports the use of useful-to-detrimental ratio concept (U50) (Refs. 18, 19) to explain the effects of added reverberant speech sound on listening difficulty for these cases.

Listening difficulty ratings for cases with early reflections are much lower than those for cases without early reflections at the same E-S/N. Early reflections decrease listening difficulty by 10% for Rev.*A* [which is less than the HSD (HSD=24.8%, $p<0.05$)] and by 40% for Rev.*B* (which is greater than HSD) compared to other cases with the same E-S/N. This may be due to early reflections effectively increasing the time window for integrated useful early energy and thus including the early part of late-arriving sound (more than 50 ms after the direct sound) in the integrated useful energy.

Figure 5 presents the relation between listening difficulty ratings and U50(A) values (A-weighted useful-to-detrimental ratio with 50-ms early time interval). As expected, similar trends versus U50(A) are seen for the only-direct-sound cases and the direct-sound-plus-reverberation cases. The cases with early reflections deviate from the main trend and show lower difficulty for a particular U50(A) value than for the cases without early reflections. STIr(Male), which is a new version of the Speech Transmission Index for male voices,[20] shows almost the same relation with listening difficulty as found for U50(A) in Fig. 5. These results may indicate that these measures could be adjusted to better reflect the benefits of early reflections on the listening difficulty scores. Although the previous study[11] discussed the benefit of early reflections on speech recognition scores rang-

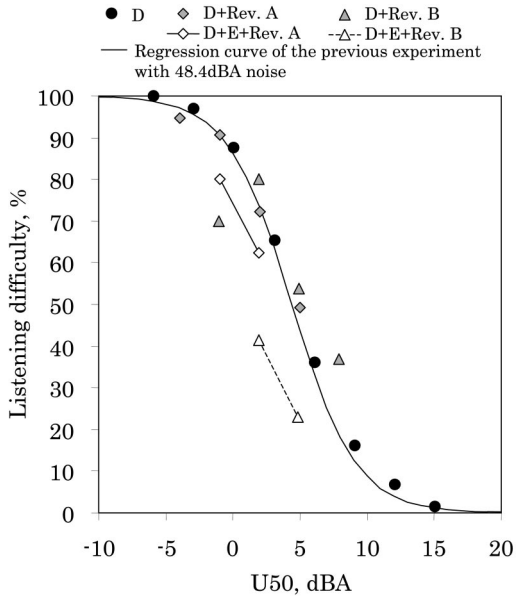


FIG. 5. Relation between A-weighted useful-to-detrimental ratio [U50(A)] and mean difficulty rating of D (direct sound) cases (filled circle), $D+$Rev. (direct sound+reverberant sound) cases (gray symbols), and $D+E+$Rev. (direct sound+early reflections+reverberant sound) cases (open symbols). The logistic regression curve of the 48.4-dBA-noise case of the former experiment is also presented.

ing from 85% to 100% for normal-hearing listeners, listening difficulty better discriminates among conditions in rooms to illustrate the effect of early reflections, as this study shows.

Because the scatter in the listening difficulty ratings, due to individual differences among subjects, is large, it is difficult to precisely relate listening difficulty ratings with physical indices. More precise ratings for some of the cases used in this experiment were obtained using a paired comparison test presented in the next section.

IV. SCHEFFE'S PAIRED COMPARISON TEST TO CONFIRM THE BENEFIT OF EARLY REFLECTIONS ON LISTENING DIFFICULTY TO SPEECH

In order to show the significant benefit of early reflections to listening difficulty, and to discuss the relation of listening difficulty scores with physical numbers, Scheffe's method of paired comparison test[21] was used.

A. Method

1. Sound fields

Seven of the conditions used in the previous experiment were used for the paired comparison tests. The electroacoustic system and the anechoic chamber used in this experiment were the same as in the previous listening test. Two direct-sound-only-plus-noise cases (D), three cases with direct sound plus Rev.B ($D+$Rev.B), and two cases with direct sound, reverberant sound (Rev.B), and early reflections ($D+E+$Rev.B) were used. All conditions are summarized in Table III.

2. Subjects and procedure

Eleven of the subjects who participated in the former experiment were used. Subjects were asked to rate the differences for each pair in one of five categories. They did this by typing a number on the keyboard with the following descriptions presented on the screen of the tiny PC, located in front of subjects, and after listening to each pair of sentences:

(1) Former is much more difficult than latter.
(2) Former is more difficult than latter.
(3) Former is as difficult as latter.
(4) Latter is more difficult than former.
(5) Latter is much more difficult than former.

The categories were assigned scores of -2, -1, 0, 1, and 2 corresponding to the first to the fifth responses in the list above, respectively. A total of 42 different pairs of speech conditions was presented twice to each subject. (Because this psychological scale is a relative scale, it could have used the word ''easy'' instead of ''difficult.'')

B. Results

1. Confirmation of significant benefit of early reflections to listening difficulty

The result of the ANOVA of the results from the psychological scale of listening difficulty showed that there was a significant effect of test condition ($p<0.0001$) and there was not a significant effect of the differences among individual subjects ($p>0.05$). The Yardstick[21] (same concept as HSD) was calculated to be 0.13 ($p<0.05$).

TABLE III. Summary of measured acoustical quantities for Scheffe's paired comparison test selected from some of cases in the former experiment in Table II. The ambient noise level was fixed at 48.6 dBA for all cases listed in this table.

Condition	Direct speech level, dBA	Early speech level, dBA	Reverb speech level, dBA	Effective speech level, dBA	Total speech level, dBA	Effective S/N, dB	Total S/N, dB	C50, dBA	U50, dBA	STIr (male)	RT (0.5–1 kHz), s
					D cases						
(i)	48.7	⋯	⋯	48.7	48.7	0.1	0.1	⋯	0.1	0.49	⋯
(ii)	57.7	⋯	⋯	57.7	57.7	9.1	9.1	⋯	9.1	0.78	⋯
					$D+$Rev$B$ cases						
(iii)	49.3	⋯	45.8	49.3	50.9	0.7	2.3	3.7	−0.2	0.43	1.1
(iv)	52.3	⋯	45.8	52.3	53.2	3.7	4.6	6.7	2.7	0.53	1.1
(v)	55.3	⋯	45.8	55.3	55.8	6.7	7.2	9.7	5.7	0.62	1.1
					$D+E+$Rev$B$ cases						
(vi)	49.3	49.4	45.8	52.4	53.2	3.8	4.6	6.7	2.7	0.50	1.1
(vii)	49.3	53.9	45.8	55.2	55.7	6.6	7.1	9.6	5.4	0.58	1.1

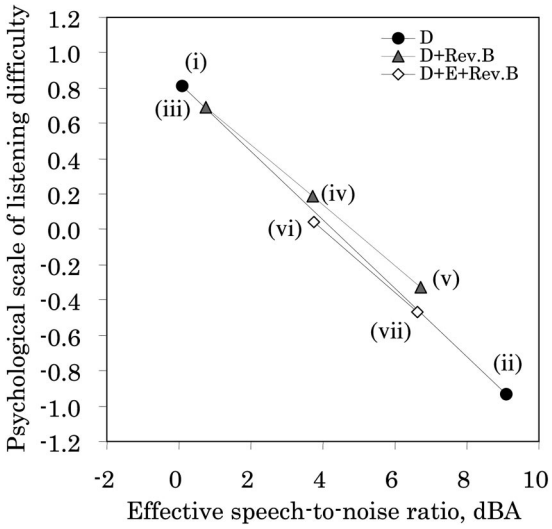


FIG. 6. Relation between effective speech-to-noise ratio and psychological scale value of listening difficulty of *D* (direct sound) cases (filled circle), *D*+Rev.*B* (direct sound+reverberant sound Rev.*B*) cases (gray triangle), and *D*+*E*+Rev.*B* (direct sound+early reflections+reverberant sound Rev.*B*) cases (open diamond). The data points are numbered from (i) to (vii) as defined in Table III.

Comparing the conditions with direct sound and those with direct sound plus reverberation, listening difficulty was increased by reverberation as shown in Fig. 6. When early reflections were added, listening difficulty was reduced relative to cases without early reflections at the same E-S/N. Conditions (i)–(vii) are described in Table III. If the difference between two conditions is greater than the Yardstick, it indicates that these two conditions are significantly different. The difference between condition (iv) (without early reflections) and condition (vi) (with early reflections) is greater than the Yardstick, and the difference between condition (v) (without early reflections) and condition (vii) (with early reflections) is also greater than the Yardstick. These results suggest that the listening difficulty of the conditions with early reflections (vi) is less than the condition without early reflections (iv), which has the same effective speech energy as condition (vi). The same could be said for the relation between (v) and (vii). This result significantly demonstrates that early reflections increase the effectiveness of the speech sounds more than expected due to the summation of the direct and early reflection energy in cases with noise and reverberation.

The key finding from Fig. 5 is the result that the conditions with early reflections and reverberation were rated as less difficult than the direct-sound-only case having the same E-S/N.

Adding reverberant speech (*D*+Rev.*B* cases) to the direct-sound-only cases (*D*) caused increases in listening difficulty that increased with increasing E-S/N. This result suggests that the reason for the lowest E-S/N (0.7 dB) case for Rev.*B* without early reflections in the previous experiment deviating from the overall trend of results was due to the scatter of listening difficulty ratings.

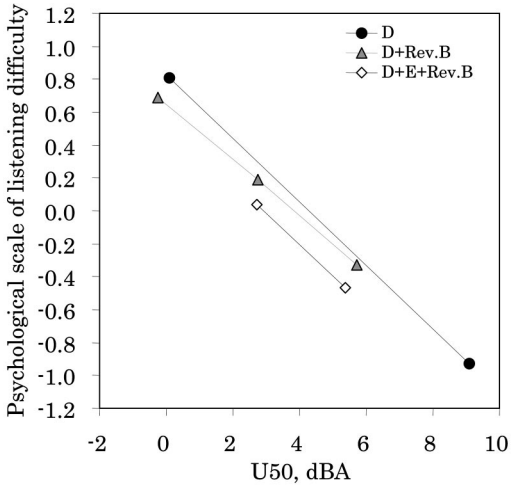


FIG. 7. Relation between A-weighted useful-to-detrimental ratio [U50(A)] and psychological scale values of listening difficulty for *D* (direct sound) cases (filled circle), *D*+Rev.*B* (direct sound+reverberant sound Rev.*B*) cases (filled triangle), and *D*+*E*+Rev.*B* (direct sound+early reflections +reverberant sound Rev.*B*) cases (open diamond).

2. Relation between listening difficulty and physical measures

As Fig. 6 shows, listening difficulty ratings are well related to E-S/N for the *D* cases and the *D*+*E*+Rev.*B* cases. The *D*+Rev.*B* cases are expected to differ from the main trend as E-S/N increases. This is because E-S/N does not take into account the detrimental effect of reverberation as the small differences in listening difficulty show.

Figure 7 shows the relation between the psychological scale of listening difficulty and U50(A). These results suggest that U50(A) overestimates the detrimental effects of reverberant sound in these results. If the time interval for useful energy is increased to 170 ms for *D*+Rev.*B* cases [(iii), (iv), and (v) in Table III], and to 240 ms for *D*+*E*+Rev.*B* cases [(vi) and (vii)], the variation of ratings with useful-to-detrimental ratios would agree with the direct-sound-only cases [(i) and (ii)] as illustrated in Fig. 8. This suggests that the effective early time interval may vary due to differences in the addition of early reflection components, and that a simple energy addition over a fixed 50-ms early time interval may not always be appropriate.

STIr values show the same trend as do the U50(A) results and similarly overestimate the detrimental effects of reverberant sound seen in Fig. 9.

V. DISCUSSION

Figure 10 shows the variation of word intelligibility scores of native English speakers and listening difficulty ratings with U50(A), for all of the conditions in this study. The listening difficulty ratings do exhibit a reasonable amount of scatter. This may be partially due to systematic errors in U50 values that were based on a fixed 50-ms early time interval as pointed out in the results of this study. Further efforts are required to find a more appropriate procedure for determining the boundary for the early time interval and to better understand the process by which early reflection energy is integrated into the useful speech energy in complete room

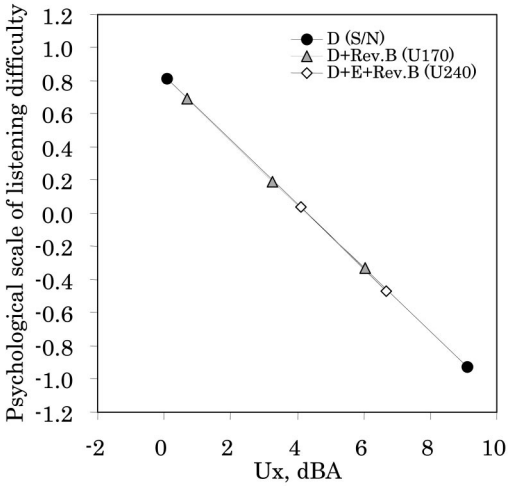


FIG. 8. Relation between A-weighted useful-to-detrimental ratio [Ux(A)] and psychological scale value of listening difficulty for *D* (direct sound) cases (filled circle), *D*+Rev.*B* (direct sound+reverberant sound Rev.*B*) cases (filled triangle, $x=170$ ms), and *D*+*E*+Rev.*B* (direct sound +early reflections+reverberant sound Rev.*B*) cases (open diamond, $x=240$ ms).

impulse responses. Several previous studies[9,22–24] give some clues as to how one might improve the method of evaluating the effect of room acoustics on speech communication.

This study clearly shows, that early reflections at least have an effect equivalent to amplifying the direct sound by as much as the energy increase they provide, and they improve the E-S/N for listening difficulty under noisy conditions. The results of experiments in noisy and reverberant sound fields suggest that early reflections tend to expand the time window of the useful early speech energy. This finding suggests that some late-arriving sound (i.e., greater than 50 ms after the direct sound) is helpful, and that this would influence the determination of optimum reverberation time criteria for speech communication in noise.

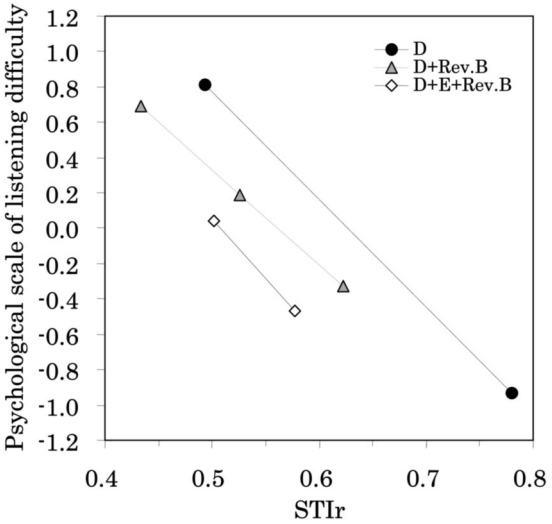


FIG. 9. Relation between revised STI for male (STIr) and psychological scale values of listening difficulty for *D* (direct sound) cases (filled circle), *D*+Rev.*B* (direct sound+reverberant sound Rev.*B*) cases (filled triangle), and *D*+*E*+Rev.*B* (direct sound+early reflections+reverberant sound Rev.*B*) cases (open diamond).

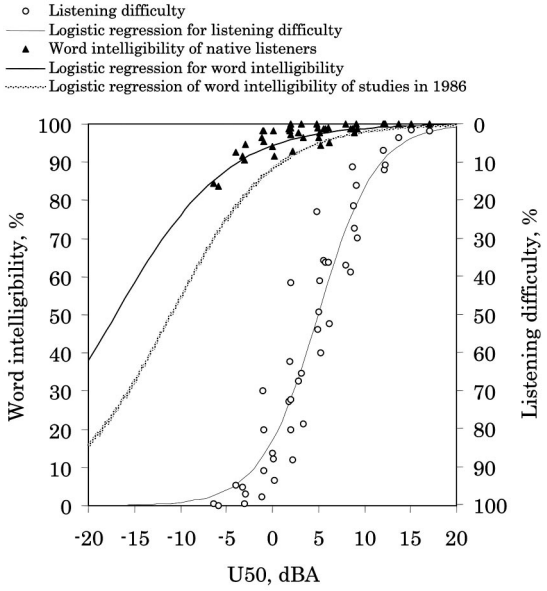


FIG. 10. Relation of word intelligibility (filled triangle) and listening difficulty (open circle) for all conditions in this study with the A-weighted useful-to-detrimental ratio [U50(A)]. Also shown is the regression curve to earlier results for word recognition scores in rooms.

The experimental conditions in this study ranged from -6 to $+15$ dBA in terms of S/N, and this range covered almost all conditions we might find in normal public spaces as identified by Pearsons *et al.*[1] The new results showed that listening difficulty ranged from 0% to 100% in this range of conditions. Values of 5% and 95% of listening difficulty correspond to S/N values of -4.5 and 14.5 dBA, respectively. On the other hand, for this same range of S/N values, word intelligibility scores only varied from 90% to 100%. This small range of word recognition scores makes it more difficult to consider the influence of early reflections on speech communication using word intelligibility scores.

In a previous study,[2] the authors discussed the range of sound-field conditions that would be most suitably evaluated using intelligibility scores, difficulty ratings, or sound quality and easiness ratings. Each rating is most appropriate for a particular range of conditions. Figure 10 also presents the regression line of intelligibility scores of old studies using the same rhyme test and measured in actual sound fields (classroom, gym, auditorium, etc.).[25,26] The curve is slightly lower than the new data presented in this study, but still indicates 90% intelligibility at a U50 of 0 dBA. (In the older studies U50 values were probably a few dB too high, because speech levels were estimated from the known output of the source using simple diffuse field theory.) The new results in the current paper clearly illustrate the range of S/N or U50 for which word intelligibility scores are more appropriate (up to 0 dBA) and the range for which listening difficulty ratings are more appropriate (-4.5 to 14.5 dBA).

One problem with listening difficulty ratings is the scatter among the results of the different subjects, even though they had received a training session. One solution to this problem is to use paired comparison tests, but for large sets of conditions, paired comparison tests can be very time con-

suming. More effort is required to minimize scatter of the difficulty ratings.

VI. CONCLUSIONS

The results demonstrate that

(1) Listening difficulty ratings better discriminate among conditions commonly found in spaces intended for speech communication than do word recognition scores.
(2) For S/N values ranging from -4.5 to $+14.5$ dBA, listening difficulty scores vary from 5% to 95%, a range of 90%. On the other hand, word intelligibility scores vary only a small amount over the upper part of this range, corresponding to near to acceptable conditions for speech communication. Word intelligibility scores are about 95% at an S/N of $+1$ dBA and start to decrease increasingly rapidly below this S/N value.
(3) Early reflection energy has at least the same effect on speech intelligibility and listening difficulty ratings as an equivalent increase in direct sound level. When reverberation exists, the benefit of early reflections is more than expected due to the added early reflection energy.
(4) Listening difficulty isn't affected by the mother tongue of subjects in the English rhyme test. It is therefore a more widely applicable test for evaluating the quality of conditions for speech communication. However, listening difficulty ratings would not reflect the decreased intelligibility scores of non-native listeners.

ACKNOWLEDGMENTS

This study was done while the first author was on sabbatical leave from Faculty of Engineering, Tohoku University, Japan, granted by The Kajima Foundation. The research project was partially supported by the Twenty-first Century Center of Excellence (COE) Program, ''Design Strategy towards Safety and Symbiosis of Urban Space'' awarded to Graduate School of Science and Technology, Kobe University. The Ministry of Education, Culture, Sports, Science and Technology of Japan sponsored the Program.

[1] K. S. Pearsons, R. L. Bennett, and S. Fidell, ''Speech levels in various noise environments,'' Report No. EPA-600/1-77-025, Washington, D.C. (1977).
[2] M. Morimoto, H. Sato, and M. Kobayashi, ''Listening difficulty as a subjective measure for evaluation of speech transmission performance in public spaces,'' J. Acoust. Soc. Am. **116**, 1607–1613 (2004).
[3] H. Sato, H. Yoshino, and M. Nagatomo, ''Relationship between speech transmission index and easiness of speech perception in reverberatory fields,'' J. Acoust. Soc. Am. **103**, Pt. 2, 2999 (1998).
[4] J. E. Greenberg, J. G. Desloge, and P. M. Zurek, ''Evaluation of array-processing algorithms for a headband hearing aid,'' J. Acoust. Soc. Am. **113**, 1646–1657 (2003).
[5] J. F. Feuerstein, ''Monaural versus binaural hearing: Ease of listening, word recognition, and attentional effort,'' Ear Hear. **13**, 80–86 (1992).
[6] F. Apoux, O. Crouzet, and C. Lorenzi, ''Temporal envelope expansion of speech in noise for normal-hearing and hearing impaired listeners: Effects on identification performance and response times,'' Hear. Res. **153**, 123–131 (2001).
[7] D. W. Downs, ''Effects of hearing aid use on speech discrimination and listening effort,'' J. Speech Hear Disord. **47**, 189–193 (1982).
[8] C. B. Hicks and A. M. Tharpe, ''Listening effort and fatigue in school-age children with and without hearing loss,'' J. Speech Lang. Hear. Res. **45**, 573–584 (2002).
[9] J. P. A. Lochner and J. F. Burger, ''The influence of reflections on auditorium acoustics,'' J. Sound Vib. **1**, 426–454 (1964).
[10] ANSI S12.60-2002, ''Acoustical performance criteria, design requirements and guidelines for schools,'' (American National Standards Institute, New York, 2002).
[11] J. S. Bradley, H. Sato, and M. Picard, ''On the importance of early reflections for speech in rooms,'' J. Acoust. Soc. Am. **113**, 3233–3244 (2003).
[12] L. L. Beranek, ''Balanced noise-criterion (NCB) curves,'' J. Acoust. Soc. Am. **86**, 650–664 (1989).
[13] H. G. Latham, ''The signal-to-noise ratio for speech intelligibility—An auditorium acoustics design index,'' Appl. Acoust. **12**, 253–320 (1979).
[14] J. W. Tukey, ''The problem of multiple comparisons,'' Mimeographed Monograph (1953), appears in full in *Collected work of J. W. Tukey*, edited by H. Braun (Chapman & Hall Inc., New York, 1994), Vol. VII.
[15] S. Buus, M. Florentine, B. Scharf, and G. Canevet, ''Native, french listeners' perception of american english in noise,'' Proceedings of Inter Noise 86, Cambridge, MA, July 1986.
[16] A. K. Nabelek and A. M. Donahue, ''Perception of consonants in reverberation by native and non-native listeners,'' J. Acoust. Soc. Am. **75**, 632–634 (1984).
[17] S. J van Wijngaarden, H. J. M. Steeneken, and T. Houtgast, ''Quantifying the intelligibility of speech in noise for non-native listeners,'' J. Acoust. Soc. Am. **111**, 1906–1916 (1998).
[18] J. S. Bradley, ''Relationships among measures of speech intelligibility in rooms,'' J. Audio Eng. Soc. **46**, 396–405 (1998).
[19] J. S. Bradley, R. D. Reich, and S. G. Norcross, ''On the combined effects of signal-to-noise ratio and room acoustics on speech intelligibility,'' J. Acoust. Soc. Am. **106**, 1820–1828 (1999).
[20] H. J. M. Steeneken and T. Houtgast, ''Mutual dependency of the octave-band weights in predicting speech intelligibility,'' Speech Commun. **28**, 109–123 (1999).
[21] H. Scheffe, ''An analysis of variance for paired comparisons,'' J. Am. Stat. Assoc. **47**, 381–400 (1952).
[22] R. Plomp, ''Rate of decay of auditory sensation,'' J. Acoust. Soc. Am. **36**, 277–282 (1964).
[23] V. R. Thiele, ''Richtungsverteilung und zeitfolge der schallruckwurfe in raumen,'' Acustica **3**, 291–302 (1953). (Title in English: ''Directional distribution and time response to impulse sounds in rooms.'')
[24] Y. Tahara and H. Sato, ''Auditory integration characteristics based on critical time delay of echo perception for improvement of indices for room acoustics,'' Proceedings of International Symposium on Room Acoustics, Awaji, Japan, 8 pages on CD-ROM (2004).
[25] J. S. Bradley, ''Predictors of speech intelligibility in rooms,'' J. Acoust. Soc. Am. **80**, 837–845 (1986).
[26] J. S. Bradley, ''Speech intelligibility studies in classrooms,'' J. Acoust. Soc. Am. **80**, 846–854 (1986).

Absorption measurement of acoustic materials using a scanning laser Doppler vibrometer

Steve Vanlanduit,[a] Joris Vanherzeele, Patrick Guillaume, and Gert De Sitter
Department of Mechanical Engineering, Vrije Universiteit Brussel Pleinlaan 2, B-1050 Brussels, Belgium

(Received 5 November 2004; revised 20 December 2004; accepted 20 December 2004)

In this article a method is proposed to estimate the normal incidence reflection ratio and absorption coefficient of acoustical materials using measurements in a transparent tube excited with a loudspeaker and terminated with the material under investigation. The waveforms are measured at different locations in the tube using a scanning laser Doppler vibrometer. Because the measurement probe (i.e., the laser beam) does not interfere with the wave in the tube, narrow tubes can be used. This means that—in contrast to the standardized wide tube tests using microphones—the proposed experiment could be used for high frequencies (in the paper an 8 mm tube was used, resulting in a 25 kHz upper frequency limit). It is shown based on theoretically known scenarios (i.e., an open tube and a rigid termination) that the absorption coefficient can be obtained with an error of about three percent. In addition, the absorption coefficient of two commonly used absorption materials—glass fiber wool and carpet—were determined and found to be in good agreement with material databases. [DOI: 10.1121/1.1859233]

PACS numbers: 43.55.Ev, 43.58.Bh, 43.35.Sx Pages: 1168–1172

I. INTRODUCTION

Different well-accepted standardized test methods exist to estimate normal incidence absorption coefficients of acoustical materials from experiments in the so-called Kundt's tube. On the one hand, one has the standing wave ratio methods, where a microphone is moved inside the tube.[1,2] An other class of Kundt-tube methods is the ''Transfer-function method'' that uses two microphones that are sidewall mounted.[3,4] The former method clearly has the disadvantage that the microphone disturbs the sound field in the tube, and therefore this implies a limitation on the upper frequency in the test. In the standard[1] it is stated that the microphone blockage cannot exceed 5% of the cross section of the tube. For a 1/4 in. microphone this means that the tube should be at least 12.7 cm (=5 in.) and that the upper frequency f_u is consequently limited by 1575 Hz ($f_u = 200/d$, with d the tube diameter). In the transfer function method, on the other hand, the microphone does not interfere with the sound field due to its mounting. However, in this case problems can arise due to the limited number of locations and the limitation of the separating distance between the microphones (according to Ref. 3 the spacing should be five times the diameter of the microphone). The optimal separating distance between the microphones, that produces good quality measurements, is frequency dependent (i.e., decreases with an increase in frequency).

In this article we will present a method to estimate absorption coefficients from experiments in a transparent (glass) tube. The incident and reflected waves are determined using a scanning laser Doppler vibrometer (which is a full field optical vibration measurement instrument). Since the laser beam does not interfere with the acoustic wave in the tube, measurements on narrow tubes (in the order of 1 cm) are possible, and therefore high-frequency absorption coefficient estimates can be obtained. Moreover, since many measurement locations with a small separation distance are available, the result is accurate for both low wave number and high spatial wave number wave shapes.

The use of optical measurement techniques for sound field measurements has been applied extensively in literature. In the early 1990s, different articles appeared on the use of holography for acoustic field visualization.[5,6] In addition to pure visualization, adapted signal processing techniques were also proposed to extract material properties based on the holographic recordings of the sound field or on spatially distributed pressure measurements. One of the most widespread methods is Tamura's technique[7] that uses a spatial Fourier transform of the measurements of the complex pressure distributions on two parallel planes. One common important disadvantage of all holography based methods is that with the current laser technology no broadband measurements can be performed (because the laser cannot shoot pulses fast and often enough). Indeed, most publications on the subject used sinusoidal excitation signals. By using a scanning laser Doppler vibrometer (SLDV)[8]—the sound field can be measured in a wide frequency band, as was illustrated in Ref. 9 for measurements in air, and in Ref. 10 for underwater ultrasonic measurements. Other more general high-frequency absorption measurements were also proposed (in Ref. 11, for instance, a method of up to 20 kHz is described). In this paper, SLDV measurements of the sound field in a glass tube will be used to estimate the normal incidence reflection ratio (and absorption coefficient) of acoustical materials.

The paper is outlined as follows: in Sec. II the absorption measurement procedure is described. Experimental results on different scenarios (i.e., an open tube, a rigid termination, and two commonly used acoustical materials) are

[a]Electronic mail: steve.vanlanduit@vub.ac.be

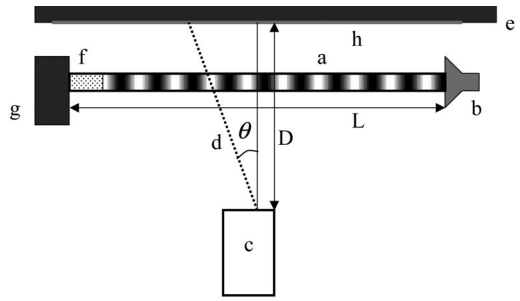


FIG. 1. Experimental set up of the tube measurements using a SLDV: (a) glass tube; (b) loudspeaker; (c) SLDV; (d) laser beam; (e) rigid block; (f) material under test; (g) rigid acoustical termination; and (h) retroreflective tape.

given in Sec. III. Finally, as a conclusion, the merits and limitations of the methods are summarized in Sec. IV.

II. THE PROPOSED ABSORPTION MEASUREMENT METHOD

A. Measurement of pressure waves in a transparent tube

Consider the arrangement as shown in Fig. 1, where the laser beam of a SLDV is directed through a glass tube onto a solid reflector (covered with retro-reflective tape to increase the amount of reflected laser light). The optical pathlength of the laser beam incident at an angle θ is given by

$$l(\theta,t)=\int_0^{D/\cos\theta} n(\theta,t,y)dy, \tag{1}$$

with D the standoff distance between the SLDV and the rigid reflector and n the refraction index (the y variable denotes the coordinate along the laser beam, starting from $y=0$ at the vibrometer position). The beat frequency of the Doppler vibrometer (at 40 MHz) will be modulated by a Doppler shift frequency Δf, which is given by

$$\Delta f=\frac{\dot{l}}{\lambda}, \tag{2}$$

where $\dot{l}$ is the time derivative of the optical pathlength and λ is the wavelength of the laser light (632 nm for a HeNe laser). Using the Gladstone–Dale equation, $n-1=K\rho$ (with K the Gladstone–Dale constant) and the ideal gas law for adiabatic conditions, $P/P_0=(\rho/\rho_0)^\gamma$ (with γ the adiabatic constant), the following relation between the time derivative of the pressure and the time derivative of the refractive index can be derived:

$$\dot{n}(\theta,t,y)=\frac{(n_0-1)}{\gamma}\frac{\dot{p}(\theta,t,y)}{p_0}, \tag{3}$$

where n_0 and p_0 are the undisturbed refractive index and pressure, respectively. By combining Eqs. (1)–(4), the following relation between the Doppler shift and the acoustic pressure is obtained:

$$\Delta f(\theta,t)=\frac{n_0-1}{\gamma p_0\lambda}\int_0^{D/\cos\theta}\dot{p}(\theta,t,y)dy. \tag{4}$$

Under the following assumptions:

(i) there is no acoustic field outside the tube;
(ii) a purely plane wave propagates inside the tube;
(iii) and the laser beam angle θ is smaller than 20° (this implies that for a tube with a diameter of 8 mm and an upper frequency of 20 kHz, the integration is performed over a maximum area along the tube of 20% of the acoustic wavelength).

Under these conditions, Eq. (4) can be rewritten as

$$\Delta f(\theta,t)=\frac{d(n_0-1)}{\cos\theta\gamma p_0\lambda}\dot{p}(\theta,t), \tag{5}$$

with d the diameter of the tube and where $\dot{p}(\theta,t)$ denotes the time derivative of the acoustic pressure at the location in the tube intersecting with the laser beam at angle θ. From Eq. (5) it is clear that one can obtain quantitative pressure values from measured Doppler shifts. However, it will be shown in the next section that this is not necessary in order to obtain reflection ratios and absorption coefficients.

B. Estimation of absorption using transfer functions

Assume that N measurements of Doppler frequency shifts $\Delta f(\theta_i,t)$ (for $i=1,...,N$) at different laser angles are performed with a SLDV. When using a periodic excitation of the tube with period T, both the Doppler shifts and pressure signals can be represented by a discrete set of Fourier coefficients $\Delta F(\theta_i,\omega_k)$ and $P(\theta_i,\omega_k)$ at frequencies $\omega_k=2\pi/T$, which are obtained from the time domain signals by applying a FFT. A relation between the two sets of Fourier coefficients can be derived from Eq. (5):

$$P(\theta_i,\omega_k)=\frac{\cos\theta\gamma P_0\lambda}{j\omega_k d(n_0-1)}\Delta F(\theta_i,\omega_k). \tag{6}$$

When using the transfer function approach[3] to estimate the reflection ratio r starting from two pressure measurements at two arbitrary locations x_1 and x_2 on the tube (corresponding to laser beam angles θ_1 and θ_2) in the tube, the following expression has to be calculated:

$$r(\omega_k)=\left|\frac{H_{12}(\omega_k)-H_I(\omega_k)}{H_R(\omega_k)-H_{12}(\omega_k)}\right|, \tag{7}$$

with $H_I=\exp(-jk_0(\omega_k)(x_1-x_2))$, $H_R=\exp(jk_0(\omega_k)(x_1-x_2))$, and $H_{12}(\omega_k)=P(\theta_2,\omega_k)/P(\theta_1,\omega_k)$, where $k_0(\omega_k)=\omega_k/c$. Of course, it is easy to see—using Eq. (6)—that

$$H_{12}(\omega_k)=\frac{P(\theta_2,\omega_k)}{P(\theta_1,\omega_k)}=\frac{\Delta F(\theta_2,\omega_k)}{\Delta F(\theta_1,\omega_k)}, \tag{8}$$

and thus the reflection ratio can be directly determined from the Doppler shift measurements.

In contrast to the classical transfer function method that uses two microphones, several hundreds of locations can easily be measured using a SLDV (in the experiments in Sec. III, $N=441$). This means that for every combination of two measurement locations x_l and x_m for $m=1,...,N$ and $l=1,...,N$ with $l\neq m$, a reflection ratio $r_{lm}(\omega_k)$ can be computed using Eq. (7). In order to limit the computation time and to guarantee that the pressure waves are still sufficiently correlated only the combinations $r_{lm}(\omega_k)$ for which $m-l$

$\leqslant M$ are calculated ($M<N$ and in the paper the value of $M=5$ is taken without loss of generality). This results in $MN-M(M-1)/2$ combinations [instead of $N(N+1)/2$]. The median of the reflection ratios $r_{lm}(\omega_k)$ is taken as the estimate of the reflection ratio $r(\omega_k)$ (in this way outliers corresponding to low signal-to-noise locations are eliminated).

An important implication of the use of a narrow tube is that the sound wave attenuation due to thermal and viscous losses in the tube becomes important. This effect can be incorporated in the transfer function method by introducing a complex wave number $k=k_0-jk_1$, where k_1 is the attenuation in nepers per meter. The following expression can be used to estimate this attenuation k_1:[1]

$$k_1=\frac{1.94\times 10^{-2}\sqrt{f}}{cd}, \tag{9}$$

with c the sound velocity. In general, however, it is advisable to experimentally determine the attenuation value k_1. This can easily be done from the determined reflection ratios at varying locations x. Indeed, consider the reflection ratio $r(\omega_k)$ without wave attenuation (which is consequently independent of the location x):

$$r(\omega_k)=\frac{P_r(x,\omega_k)}{P_i(x,\omega_k)}, \tag{10}$$

with P_i and P_r the amplitudes of the incident and reflected waves at frequency ω_k and

$$r_{k_1}(x,\omega_k)=\frac{P_r^{k_1}(x,\omega_k)}{P_i^{k_1}(x,\omega_k)} \tag{11}$$

the reflection ratio at location x with wave attenuation k_1 (as it occurs in the tube). The attenuation of the waves is given by $P_r^{k_1}(x,\omega_k)=P_r\exp(-k_1(L+x))$ and $P_i^{k_1}(x,\omega_k)=p_i\exp(-k_1(L-x))$, and therefore

$$r_{k_1}(x,\omega_k)=r(\omega_k)\exp(-2k_1x). \tag{12}$$

Because the value $r_{k_1}(x,\omega_k)$ is measured for $N-1$ different locations x both $r(\omega_k)$ and k_1 can be estimated using a least-squares approximation:

$$(\hat{r}(\omega_k),\hat{k}_1)=\min_{(r(\omega_k),k_1)}\sum\,(r_{k_1}(x,\omega_k)-r(\omega_k)\exp(-2k_1x))^2. \tag{13}$$

After the estimation of the reflection ratios, the absorption coefficients $\alpha(\omega_k)$ are obtained from the reflection ratios $r(\omega_k)$ by using the relation $\alpha(\omega_k)=1-r(\omega_k)^2$.

III. EXPERIMENTAL RESULTS

In the experiments in this paper a glass tube with a length of 830 mm, an inner diameter of 8 mm, and a 1.5 mm tube wall thickness was used. The SLDV (a Polytec PSV300) was positioned at a standoff distance of 840 mm from the rigid block. 441 locations on the tube were measured with a spacing of 1 mm. An overview picture of the measurement setup is shown in Fig. 2.

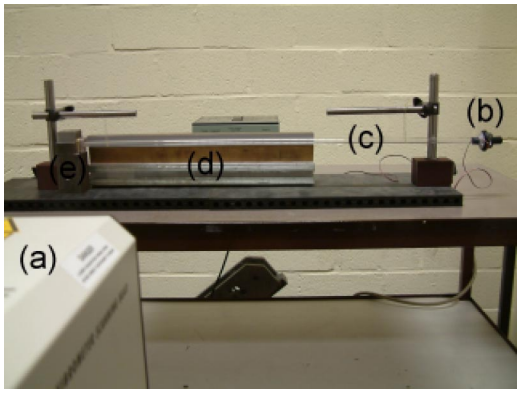


FIG. 2. Overview of the measurement setup: (a) scanning laser Doppler vibrometer; (b) loudspeaker; (c) glass tube; (d) rigid support; and (e) rigid terminations with absorption material (carpet).

At one end of the tube a loudspeaker was attached using a rubber connecting tube [see a detailed photo in Fig. 3(a)]. At the other end, different terminations were created.

(1) A rigid termination.
(2) An open termination.
(3) Glass fiber wool in front of a rigid termination [see Fig. 3(b)].
(4) Carpet attached onto a rigid termination [see Fig. 3(c)].

The beam was excited with a periodic chirp signal (with period $T=100$ ms) and 10 averages were performed. Although the method in this paper could theoretically be used for frequencies starting from $f_l=3c/4L=290$ Hz up to f_u

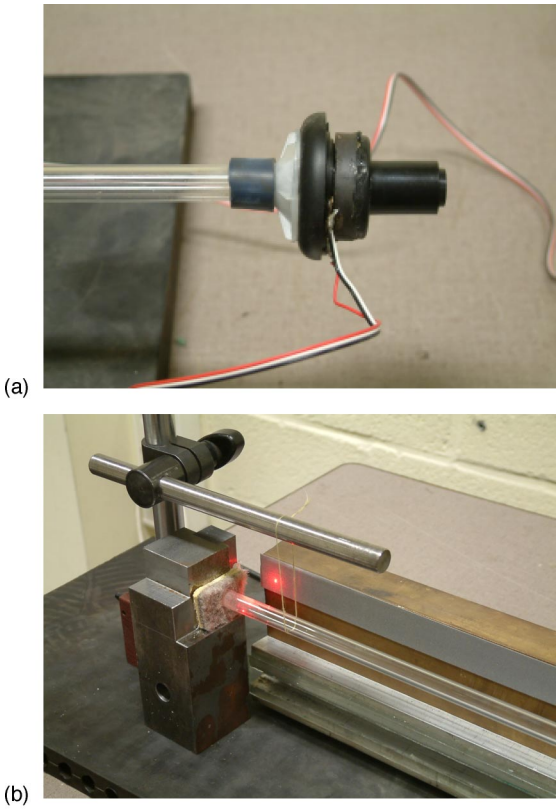


FIG. 3. Details of the measurement setup: (a) Loudspeaker and (b) termination with carpet.

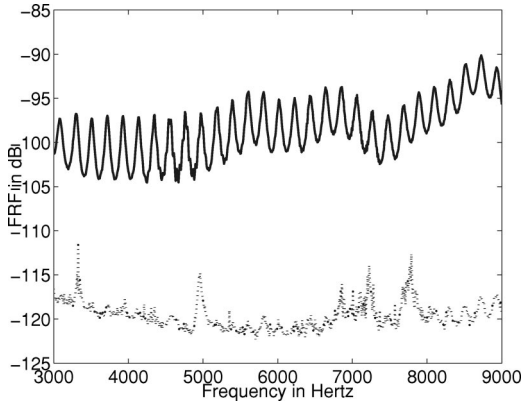


FIG. 4. FRF measured in the tube and (solid line) and FRF measured 5 mm above the tube (dotted line).

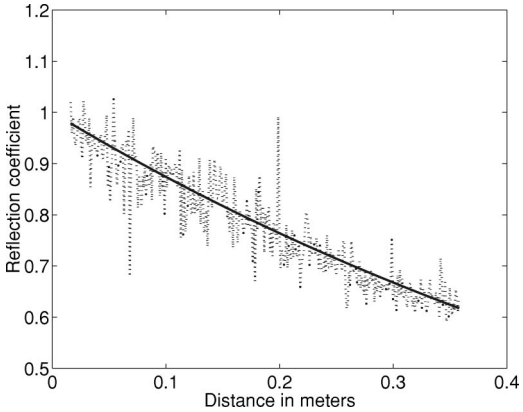


FIG. 6. Decay of the identified reflection ratio in function of the distance from the termination. The solid line gives the approximaton values [see Eq. (13)] while the dotted line shows the experimentally determined values.

$=200/d=25$ kHz (following the limits in the standard),[3] the loudspeaker response was only flat enough in the region from 3 up to 9 kHz. In Fig. 4(a) it can be seen that—if one uses a rigid termination—the resonance frequencies (which are multiples of $c/2L=207$ Hz) are clearly visible in the Frequency Response Functions FRFs plot (the FRF is the ratio between the spectrum of the Doppler shifts and the electrical signals sent to the loudspeaker). Moreover, the FRFs at a location 5 mm above the tube is about 20 dB lower, which indicates that the acoustic isolation of the tube walls is sufficient.

The quality of the measurements is given by the coherence function,[12] which is shown in Fig. 5. Although the average coherence (i.e., averaged over all measurement locations) in Fig. 5 drops below 0.7 between the resonance frequencies (which is normal at nodal locations), the overall measurement quality is quite good.

In Fig. 6 the decay of the reflection ratio as a function of the measurement location is given. The solid line in Fig. 6 shows the estimated values determined using the least squares approximation from Eq. (13). The estimated value $k_1=0.67$ (at 8523 Hz) corresponded to the value one would obtain from Eq. (9) when using a tube diameter of 7.5 mm (instead of the actual value of 8 mm).

The results of the absorption coefficients for a rigid reflector are shown in Fig. 7. The average deviation from the 100% theoretical absorption values is 0.028. It can be remarked that because of the use of a least-squares approximation [see Eq. (13)], the absorption coefficients in Fig. 7 are much smoother than the measurement results in Fig. 6.

For an open tube the theoretical reflection ratios r are given by[13]

$$r=1-0.14k_0^2d^2. \tag{14}$$

When comparing these values to the experimentally determined values using the proposed approach (see Fig. 8) again about 2% error is present.

Both validation cases on the rigid termination and the open tube show that the results are quite reliable. In Fig. 9, absorption coefficient results of the presented method when applied to commonly used acoustical materials are given (3 cm glass fiber wool and carpet).

Although in this case no theoretical values are available, the experimental values agree with data available from material databases (however, it has to be stressed that material manufacturers do not provide high-frequency absorption values nor do they supply narrow band measurements). The results in the experiment in this paper (see Fig. 9) agree quite well with the values given in the graphs in Ref. 14 for a 2.54 cm (=1 in.) board: approximately 98% at 3 kHz and close to 100% from 4 kHz on (the deviation in the region between 3

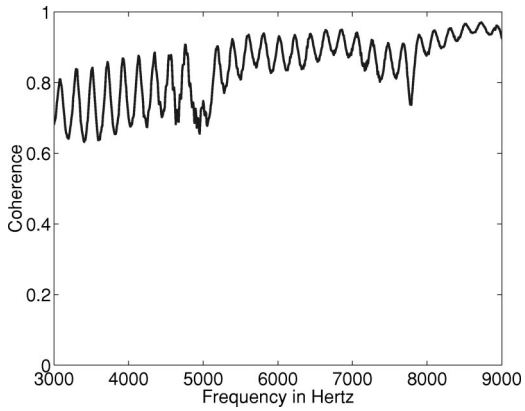


FIG. 5. Coherence of the measurements (average over all 441 measurement locations).

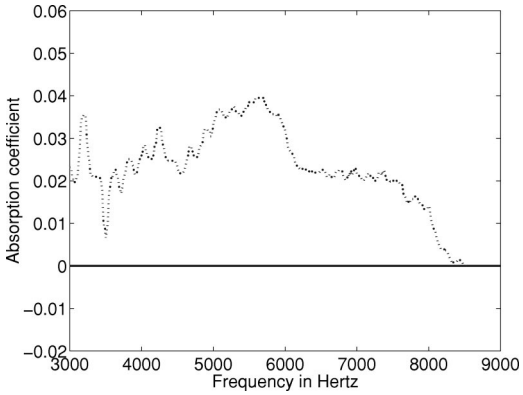


FIG. 7. Absorption coefficient for a rigid termination. The solid line gives the theoretical values while the dotted line shows the experimentally determined values.

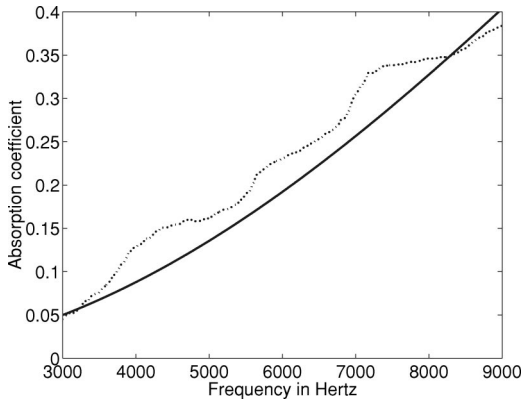


FIG. 8. Absorption coefficient for the open tube termination. The solid line gives the theoretical values [see Eq. (14)] while the dotted line shows the experimentally determined values.

and 6 kHz can be due to the fact that the density of the glass fiber wool is slightly increased when inserting the specimen in the narrow tube (this could be improved by using a slightly wider tube). The results of the carpet measurements are also in agreement with data from the literature: Everest[14] tabulates absorptions values of 0.65 at 4 kHz that agrees well with what we obtain (however, it has to be remarked that there is a large variation of the absorption coefficients of carpet depending on the type of carpet).

IV. MERITS AND LIMITATIONS OF THE METHOD

The method presented in this paper has the potential to fill the gap of measuring normal incidence absorption coefficients at high frequencies (up to 20 kHz). Since broadband excitation is used measurements with a high-frequency resolution are available. From the validation experiments included in the paper it can be concluded that the absorption coefficient can be obtained with an accuracy of about 3%. The number of measurement locations that has to be measured to obtain this accuracy is much higher than in the classical two-microphone technique. This, however, does not pose a problem since the scan can be completely automated

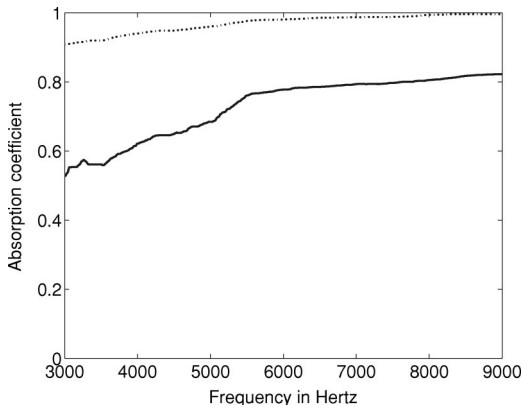


FIG. 9. Absorption coefficient for glass wool (dotted) and carpet (solid line) termination.

(the total measurement time was only 7 min). Since the method is noncontacting, there is no interference with the sound wave in the tube. This also means that the method is also able to measure the absorption behavior of materials in hostile environments (e.g., high temperatures, humidity, etc.).

The major limitation of the method is the disability to extract small samples from a material. In the present paper an adapted tool was developed for this goal (a steel tube with one of the edges sharpened). For materials that are difficult to introduce in the glass tube, the steel tube with material inside can be directly coupled onto the glass tube. Of course, the method is limited to materials with a texture, pattern or roughness that is small compared to the diameter of the tube.

ACKNOWLEDGMENTS

This research has been sponsored by the Flemish Institute for the Improvement of the Scientific and Technological Research in Industry (IWT), the Fund for Scientific Research—Flanders (FWO) Belgium. The authors also acknowledge the Flemish government (GOA-Optimech) and the research council of the Vrije Universiteit Brussel (OZR) for their funding. The first author holds a grant as a Postdoctoral Researcher from the FWO Vlaanderen.

[1] ISO Standard, ''Acoustics—Determination of sound absorption coefficient and impedance in impedance tubes. Part 1: Method using standing wave ratio,'' ISO 10534-1, International Standardization Organization, 1996, pp. 12–15.
[2] ASTM Standard, ''Standard test method for impedance and absorption of acoustical materials by the impedance tube method,'' ASTM384, American Society for Testing of Materials, March 1999.
[3] ISO Standard, ''Acoustics—Determination of sound absorption coefficient and impedance in impedance tubes. Part 1: Transfer-function method,'' ISO 10534-2, International Standardization Organization, 1998, pp. 11–15.
[4] ASTM Standard, ''Standard test method for impedance and absorption of acoustical materials using a tube, two microphones and a digital frequency analysis system,'' ASTM1050-90, American Society for Testing of Materials, September 1990.
[5] O. J. Lokberg, ''Sound in flight: measurement of sound fields by use of tv holography,'' Appl. Opt. **33**, 2574–2584 (1994).
[6] O. J. Lokberg, ''Recording of sound emmission and propagation in air using tv-holography,'' J. Acoust. Soc. Am. **96**, 2244–2250 (1994).
[7] M. Tamura, ''Spatial fourier transform method of measuring reflection coefficient at oblique incidence,'' J. Acoust. Soc. Am. **88**, 2259–2264 (1990).
[8] N. A. Halliwell, ''Laser Doppler measurement of vibrating surfaces: A portable instrument,'' J. Sound Vib. **62**, 312–315 (1979).
[9] L. Zipser, H. Franke, E. Olsson, N. Molin, and M. Sjodahl, ''Reconstructing two-dimensional acoustic object fields by use of digital phase conjugation of scanning laser vibrometry recordings,'' Appl. Opt. **42**, 5831–5838 (2003).
[10] A. R. Harland, J. N. Petzing, and J. R. Tyrer, ''Nonperturbing measurements of spatially distributed underwater acoustic fields using a scanning laser Doppler vibrometer,'' J. Acoust. Soc. Am. **115**, 187–196 (2004).
[11] D. Lafarge, J. F. Allard, B. Brouard, C. Verhaegen, and W. Lauriks, ''Characteristic dimensions and prediction at high frequencies of the surface impedance of porous layers,'' J. Acoust. Soc. Am. **93**, 2474–2478 (1993).
[12] N. M. Maia and J. M. Silva, *Theoretical and Experimental Modal Analysis* (Research Studies Press Ltd., Taunton, UK, 1997).
[13] M. L. Munjal, *Acoustics of Ducts and Muffler* (Wiley, New York, 1987).
[14] F. A. Everest, *Master Handbook of Acoustics* (McGraw-Hill, New York, 2001).

Retrofocusing techniques for high rate acoustic communications[a)]

Milica Stojanovic[b)]
Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

(Received 8 June 2004; revised 14 December 2004; accepted 16 December 2004)

High rate underwater communications have traditionally relied on equalization methods to overcome the intersymbol interference (ISI) caused by multipath propagation. An alternative technique has emerged in the form of time-reversal, which comes at virtually no cost in computational complexity, but sacrifices the data rate and relies on the use of large arrays to reduce ISI. In this paper, spatiotemporal processing for *optimal* multipath suppression is addressed analytically. A communication link between a single element and an array is considered in several scenarios: uplink and downlink transmission, with and without channel state information and varying implementation complexity. Transmit/receive techniques are designed which simultaneously maximize the data detection signal-to-noise ratio and minimize the residual ISI, while maintaining maximal data rate in a given bandwidth and satisfying a constraint on transmitted energy. The performance of so-obtained focusing techniques is compared to the standard ones on a shallow water channel operating in a 5 kHz bandwidth around a 15 kHz center frequency. Results demonstrate benefits of focusing techniques whose performance is not conditioned on the array size. Optimal configurations are intended as a basis for adaptive system implementation in which channel estimates will replace the actual values.
[DOI: 10.1121/1.1856411]

PACS numbers: 43.60.Ac, 43.60.Dh, 43.60.Fg, 43.60.Gk [EJS] Pages: 1173–1185

I. INTRODUCTION

High rate, bandwidth-efficient underwater communications have traditionally relied on adaptive equalization methods to overcome the intersymbol interference (ISI) caused by multipath propagation. Excellent performance of these receivers comes at a price of high computational complexity. While processing complexity can be somewhat reduced by use of sophisticated spatiotemporal multichannel equalizers,[1] retrofocusing techniques appear to offer a different approach. In traditional equalization, all of the signal processing is performed at the receiver side, while the transmitter uses standard signaling waveforms, which are designed *a priori*, and, hence, are not matched to the channel. A different, and possibly better approach results if signal processing can be split between the transmitter and receiver. Such an approach forms the basis of spatiotemporal retrofocusing.

In its simplest form, retrofocusing is achieved by transmitting a time-reversed (or equivalently, phase-conjugated in the frequency domain) replica of a probe signal received earlier from the source location. This technique has been used for medical imaging, therapy, and material testing,[2] while more recently, time-reversal has been investigated as a communication technique that offers lower computational complexity as compared to traditional equalization.[3–23] However, in high rate communications time-reversal alone does not eliminate ISI, the fact that motivates present analysis and the search for optimal retrofocusing techniques.

A. Previous work

Several research groups have been involved in application of time-reversal arrays to undersea acoustic communications, addressing active phase-conjugation for two-way communication, as well as passive phase-conjugation for one-way communication from a single-element source to an array.

The Scripps group has been engaged in experimental work, using large arrays to demonstrate spatial and temporal focusing of phase-coherent communication signals.[3–8] Communication begins with a single-element source transmitting the initial probe to an array. The array then uses the time-reversed replicas of the received, channel-distorted probe to generate *transmit* filters that are subsequently used for pulse-shaping the information sequence that is sent back to the source. This method of two-way communication is called active phase-conjugation. In one of the experiments, a 30 element array, operating at a center frequency of 3.5 kHz, was used to transmit PSK signals pulse-shaped at the transmitter array by a time-reversed replica of the probe signal received earlier. No signal processing was employed at the receiver, and transmission at 1 kbit/s over 10 km was reported.

The University of Washington group has addressed experimentally the technique of passive phase-conjugation,[9–13] in which the same principle of time-reversal is used for one-way communication. In this technique, the single-element source sends a probe, waits for the channel reverberation to subside, and then transmits the information-bearing signal to the array. The received channel-distorted probe is time-reversed and used at each array element as a *receive* filter for subsequent detection of the information-bearing signal. Peri-

[a)]Portion of this work was presented at the High Frequency Ocean Acoustics Conference, March 2004.
[b)]Electronic mail: millitsa@mit.edu

odic insertion of the probe signal is necessary to account for the time-variability of the channel. The technique was demonstrated experimentally using a 14 element array, operating in the 5–20 kHz band, to transmit data over distances of about 1 km. Transmission at 2 kbits/s was reported; however, such transmission could only be sustained for a short period of time, after which the probe signal had to be retransmitted to account for the channel time-variability (probe retransmission effectively reduced the data rate by a factor of 2). To recover the loss in data rate, decision-directed phase-conjugation was used, in which the detected data stream was used to regenerate the channel estimates. Passive phase conjugation was compared to standard equalization, showing that significant performance degradation, which increases with signal-to-noise ratio (SNR) and eventually leads to saturation, is the price to be paid for low computational complexity.[14]

In parallel with these experimental efforts that emphasize low-complexity processing using time-reversal, a third group of researchers at IST, Portugal, focused on analytical work.[15–20] Realizing that probe retransmission considerably reduces the effective data rate in both active and passive phase-conjugation, this group proposed the use of adaptive channel estimation to generate the up-to-date time-reversed filters directly from the received information-bearing signal, thus eliminating the need for probe retransmission (inevitably at the expense of increased computational complexity). This group also proposed the use of low-complexity adaptive equalization *in conjunction* with time-reversal.

As the early experiments devoted to implementing time-reversal in the ocean and testing the basic concepts have shown, suppression of multipath effects through time-reversal can be achieved at the expense of reduced data throughput and/or the need for a large array. For a better utilization of channel resources, additional signal processing is necessary to eliminate ISI and enable high-rate communications.

Ensuring ISI-free transmission in a system that has multiple transmit/receive elements is a major asset in a channel whose bandwidth is severely limited. In particular, it lays ground for capacity improvement through the use of space-time coding and multi-input multi-output (MIMO) signaling. This technique, developed originally for radio channels, was shown to increase the fundamental channel capacity in proportion to the number of transmit/receive elements used.[24] However, capacity-approaching codes are known only for ISI-free channels. MIMO signal processing for controlling the intersymbol and the inter-channel interference in underwater acoustic systems was addressed in the framework of multiple-user communications,[25] and, more recently, for single-user communications using multiple transmit elements.[26] Recent experimental results demonstrate large bandwidth efficiency improvement over acoustic channels, provided that accurate channel estimation is available and that residual ISI is kept at a minimum.

B. Problem definition

While recent research demonstrates the potential of time-reversal in spatial localization of acoustical energy, it often fails to recognize that time-reversal alone does not provide temporal focusing necessary to eliminate ISI caused by multipath propagation. Time-reversal recombines multipath energy in a manner of matched filtering, whose function is to maximize the SNR at a given time instant, and not to eliminate ISI. In fact, matched filtering *increases* temporal dispersion of the signal, i.e., the duration of the overall impulse response of the system. While SNR maximization is an appropriate optimization criterion for single pulse focusing, its application to communication problems, where a *sequence* of data-modulated pulses is transmitted at a high rate, must be approached judiciously. For signals that contain temporal dispersion, matched filter represents only the front end of the optimal receiver, and must be followed by a sequence estimator or an equalizer.[27]

Multipath components that remain after matched filtering contribute to residual ISI, whose severity depends on the channel. If not equalized, residual ISI may completely prevent detection. Increasing the number of array elements in a time-reversal array only helps to *reduce* residual ISI, but it does not eliminate it. Hence, if time-reversal is to be used toward eliminating the multipath distortion without sacrificing the data rate, it must be combined with equalization to remove residual ISI. However, the advantage of this approach to standard equalization that uses fixed transmitter wave forms is not apparent. The use of retrofocusing for *complete* suppression of multipath thus remains an open question.

In this paper, a solution is proposed to the following problem: If the channel responses between a single element and an array are known, determine the optimal transmit/receive technique that the two can use to simultaneously (i) eliminate ISI *and* (ii) maximize SNR, while maintaining *maximal* data rate in a given bandwidth and satisfying a constraint on transmitted energy. Note that because it allows for transmitter as well as receiver optimization, the solution differs from standard equalization. Also, because it explicitly requires minimization of ISI, it differs from time-reversal. The resulting system does *not* depend on the number of array elements to minimize the multipath distortion, i.e., it does not trade the computational complexity for the array size, but instead provides an answer for a variety of applications that cannot afford large arrays.

For those applications that also cannot afford processing power at both ends of the link, a constrained optimization problem is considered. A complexity restriction, likely to be imposed on the remote, single-element end, forces its transmit/receive filters to use no knowledge of the channel. The resulting one-sided focusing solution sacrifices some of the performance of the two-sided focusing in exchange for minimal implementation complexity. Analytical results are provided to quantify this trade-off.

A related question that emerges during the study of optimal focusing is the following: If the requirement for no ISI is relaxed, and the use of both channel-dependent transmit filtering *and* equalization at the receiver is allowed, what is the optimal system configuration, and how does its performance compare to that of optimal focusing? Analytical solution to this problem provides an upper bound on the perfor-

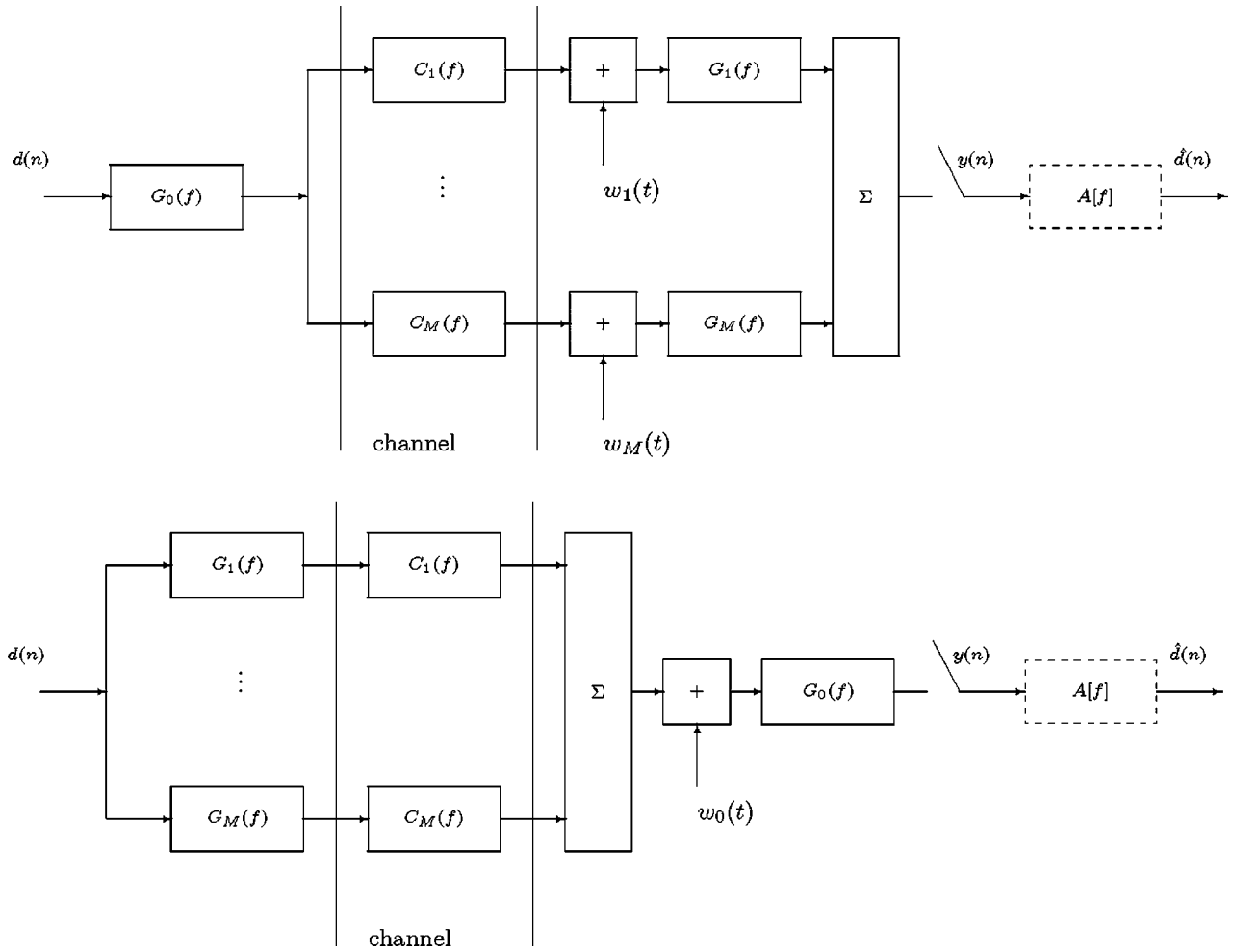


FIG. 1. Uplink (above) and downlink (below) transmission. An equalizer (dashed box) may or may not be used.

mance of all spatiotemporal processing methods.

System optimization is addressed in Sec. II. under various optimization criteria. In Sec. III., the performance of resulting techniques is compared through numerical computation of the analytical expressions for predicted performance of a system operating at 10 kbits/s over a 3 km shallow water channel. Results demonstrate the benefits of optimal focusing which outperforms time-reversal, and whose performance is not contingent on the array size. The conclusions are summarized in Sec. IV.

Optimal configurations discussed in this paper are intended as a basis for adaptive system implementation in which channel estimates will replace the unknown, time-varying responses. An adaptive channel estimation procedure, which uses a low-complexity decision-directed approach,[28] is suitable for this task.

II. SYSTEM OPTIMIZATION

In this section, system optimization is addressed for uplink and downlink communication (to/from array), as shown in Fig. 1. Performance is assessed using SNR as the figure of merit, and compared to that of time-reversal, standard linear equalization, and transmit time-reversal in conjunction with equalization (receive time-reversal is identical to matched filtering, which, when followed by an equalizer, reduces to standard equalization). System optimization is first addressed under no ISI requirement, while optimization of transmit filtering for use in conjunction with equalization is addressed subsequently.

A. Transmitter/receiver optimization for no ISI (focusing)

The sequence of data symbols $d(n)$ is transmitted at a symbol rate $1/T$. Referring to Fig. 1, the problem is to find the transmit/receive filters $G_0(f)$ and $G_1(f),\ldots G_M(f)$ such that the SNR at the receiver is maximized, subject to the constraint that there is no ISI in the decision variables $\hat{d}(n) = y(nT)$, and that finite transmitted energy per symbol E is used. The channel responses $C_m(f)$, $m=1,...,M$, and the power spectral density $S_w(f)$ of the uncorrelated noise processes $w_m(t)$, $m=0,...,M$, are assumed to be known.

Let the composite equivalent baseband channel transfer function be denoted by

$$F(f) = G_0(f) \sum_{m=1}^{M} G_m(f) C_m(f). \tag{1}$$

The received signal after filtering is then given by

$$y(t) = \sum_n d(n) f(t-nT) + z(t) \tag{2}$$

where the noise $z(t)$ has power spectral density

$$S_z(f)=S_w(f)\times\begin{cases}\sum_{m=1}^{M}|G_m^2(f)| & \text{for uplink transmission}\\ |G_0^2(f)| & \text{for downlink transmission.}\end{cases} \tag{3}$$

The requirement for no ISI is expressed as

$$F(f)=X(f), \tag{4}$$

where $X(f)$ is a Nyquist transfer function, i.e., it is band-limited to $|f|<1/T$, and its waveform in time, $x(t)$, satisfies the condition

$$x(nT)=\begin{cases}x_0 & \text{for } n=0\\ 0 & \text{otherwise.}\end{cases} \tag{5}$$

Without the loss of generality, we take that $X(f)=|X(f)|$. For example, $X(f)$ can be chosen as a raised cosine spectrum,[27] whose bandwidth B is controlled by the roll-off factor $\alpha\in[0,1]$, $B=(1/T)(1+\alpha)$. When $\alpha\to 0$, $X(f)\to T$ for $|f|\leqslant 1/2T$, and no excess bandwidth is used, i.e., maximal symbol rate is achieved for which ISI-free transmission is possible within the available bandwidth.

When there is no ISI, the received signal, sampled at times nT, is given by

$$y(nT)=d(n)x_0+z(nT) \tag{6}$$

and the SNR is

$$\mathrm{SNR}=\frac{\sigma_d^2x_0^2}{\sigma_z^2}. \tag{7}$$

where $\sigma_d^2=E\{|d^2(n)|\}$ and $\sigma_z^2=\int_{-\infty}^{+\infty}S_z(f)df$.

The total transmitted energy is the energy of the signal $u_0(t)=\Sigma_n d(n)g_0(t-nT)$ for the uplink scenario, or the sum of energies of the signals $u_m(t)=\Sigma_n d(n)g_m(t-nT)$, $m=1,...,M$ for the downlink scenario. The power spectral density of these signals is $S_{u_m}(f)=(1/T)S_d(f)|G_m(f)|^2$, where $S_d(f)$ is the power spectrum of the data sequence.[27] If the transmitted energy per symbol is set to E, assuming uncorrelated data symbols ($S_d(f)=\sigma_d^2$), the energy constraint is expressed as

$$E=\sigma_d^2\times\begin{cases}\int_{-\infty}^{+\infty}|G_0^2(f)|df, & \text{for uplink transmission}\\ \Sigma_{m=1}^{M}\int_{-\infty}^{+\infty}|G_m^2(f)|df, & \text{for downlink transmission.}\end{cases} \tag{8}$$

1. Unrestricted optimization (two-sided filter adjustment)

Let us first consider uplink transmission. Taking into account the energy constraint, and the no-ISI requirement, $G_0(f)=X(f)/\Sigma_{m=1}^{M}G_m(f)C_m(f)$, the SNR is expressed as

$$\mathrm{SNR}=\frac{Ex_0^2}{\sigma_z^2\int_{-\infty}^{+\infty}|G_0^2(f)|df}=\frac{Ex_0^2}{\int_{-\infty}^{+\infty}\frac{X^2(f)}{|\Sigma_{m=1}^{M}G_m(f)C_m(f)|^2}df\int_{-\infty}^{+\infty}S_w(f)\Sigma_{m=1}^{M}|G_m^2(f)|df}. \tag{9}$$

This function is to be maximized with respect to the receive filters $G_m(f)$, $m=1,...,M$. To do so, we use a two-step procedure, each step involving one Schwarz inequality. The first inequality states that

$$\left|\sum_{m=1}^{M}G_m(f)C_m(f)\right|^2\leqslant\sum_{m=1}^{M}|G_m^2(f)|\sum_{m=1}^{M}|C_m^2(f)|, \tag{10}$$

where the equality holds for

$$G_m(f)=\alpha(f)C_m^*(f). \tag{11}$$

We note similarly with time-reversal in that receive filters should be proportional to the phase-conjugate of the channel transfer functions. However, there is room for additional improvement through optimization of the function $\alpha(f)$.

Denoting the composite channel power spectral density by

$$\gamma(f)=\sum_{m=1}^{M}|C_m^2(f)| \tag{12}$$

and using the inequality (10) we have that

$$\mathrm{SNR}\leqslant\frac{Ex_0^2}{\int_{-\infty}^{+\infty}\frac{X^2(f)}{\gamma(f)\Sigma_{m=1}^{M}|G_m^2(f)|}df\int_{-\infty}^{+\infty}S_w(f)\Sigma_{m=1}^{M}|G_m^2(f)|df}. \tag{13}$$

Applying a second Schwarz inequality to the denominator of the SNR bound yields

$$\int_{-\infty}^{+\infty}\frac{X^2(f)}{\gamma(f)\Sigma_m|G_m^2(f)|}df\int_{-\infty}^{+\infty}S_w(f)\sum_{m=1}^{M}|G_m^2(f)|df\geqslant\left[\int_{-\infty}^{+\infty}\frac{X(f)}{\sqrt{\gamma(f)}}\sqrt{S_w(f)}df\right]^2, \tag{14}$$

where the equality holds for

$$\frac{X(f)}{\sqrt{\gamma(f)S_w(f)}}=\beta\sum_{m=1}^{M}|G_m^2(f)| \tag{15}$$

and β is a constant. Combining the conditions (11) and (15) we obtain the optimal value

$$\alpha(f)=\frac{1}{\sqrt{\beta}}\frac{1}{S_w^{1/4}(f)}\frac{\sqrt{X(f)}}{\gamma^{3/4}(f)}. \tag{16}$$

The transmit filter is now obtained from the no-ISI condition (4) as

$$G_0(f) = \frac{X(f)}{\alpha(f)\gamma(f)} \tag{17}$$

and the constant β then follows from the energy constraint (8, uplink):

$$E = \sigma_d^2 \beta \int_{-\infty}^{+\infty} \sqrt{S_w(f)} \frac{X(f)}{\sqrt{\gamma(f)}} df. \tag{18}$$

The desired filters are given in the following:

$$G_0(f) = K(f)\sqrt{X(f)}\gamma^{-1/4}(f),$$

$$G_m(f) = K^{-1}(f)\sqrt{X(f)}\gamma^{-3/4}(f)C_m^*(f), \quad m = 1,...,M$$

where

$$K(f) = \sqrt{\frac{E/\sigma_d^2}{\int_{-\infty}^{+\infty} \sqrt{S_w(f)} \frac{X(f)}{\sqrt{\gamma(f)}} df}} S_w^{1/4}(f). \tag{19}$$

We can verify that the result reduces to the known case by setting $M = 1$.[27]

This selection of filters achieves maximal SNR,

$$\mathrm{SNR}_2 = E x_0^2 \left[\int_{-\infty}^{+\infty} \sqrt{S_w(f)} \frac{X(f)}{\sqrt{\gamma(f)}} df \right]^{-2}, \tag{20}$$

where the index "2" signifies the fact that both sides of the link adjust their filters in accordance with the channel.

Filter optimization in the downlink transmission case is accomplished similarly, by a double application of the Schwarz inequality. The resulting filters are given in the same form as for the uplink case, except that the factor $K(f)$ is reciprocal of that given in Eq. (19). The same maximal SNR, SNR_2 (20), is achieved.

2. Restricted optimization (one-sided filter adjustment)

We now turn to the situation in which one side of the communication link is restricted to have minimal complexity, such as when limited processing power is available at one end of the link. Namely, we constrain the single-element side to use a filter $G_0(f)$ that is fixed, i.e., it may not be computed as a function of the channel responses.

To illustrate the optimization procedure, we look at the downlink transmission case. The no-ISI condition still must hold, $F(f) = X(f)$, and to maximize the SNR, this transfer function should be divided between the transmitter and receiver so that

$$G_0(f) = \beta \frac{\sqrt{X(f)}}{\sqrt{S_w(f)}}, \tag{21}$$

where β is a constant. For this selection, the SNR achieves the Schwarz inequality (matched filter) bound,

$$\mathrm{SNR} = \frac{\sigma_d^2 \left| \int_{-\infty}^{+\infty} G_0(f) \Sigma_{m=1}^{M} G_m(f) C_m(f) \right|^2}{\int_{-\infty}^{+\infty} S_w(f) |G_0^2(f)| df}$$

$$\leqslant \sigma_d^2 \int_{-\infty}^{+\infty} \frac{1}{S_w(f)} \left| \sum_{m=1}^{M} G_m(f) C_m(f) \right|^2 df. \tag{22}$$

Applying the Schwarz inequality (10) to the above integrand yields

$$\mathrm{SNR} \leqslant \sigma_d^2 \int_{-\infty}^{+\infty} \frac{1}{S_w(f)} \gamma(f) \sum_{m=1}^{M} |G_m^2| df \tag{23}$$

with equality holding for $G_m(f) = \alpha(f) C_m^*(f)$, $m = 1,...,M$. Combining this condition with the matched-filter requirement (21) and the no-ISI constraint, we obtain the optimal value

$$\alpha(f) = \frac{1}{\beta} \frac{\sqrt{X(f)}}{\gamma(f)} \sqrt{S_w(f)}. \tag{24}$$

The constant β can now be determined from the energy constraint (8, downlink):

$$E = \sigma_d^2 \frac{1}{\beta^2} \int_{-\infty}^{+\infty} S_w(f) \frac{X(f)}{\gamma(f)} df. \tag{25}$$

The desired filters are given in the following:

$$G_0(f) = K^{-1}(f)\sqrt{X(f)},$$

$$G_m(f) = K(f)\sqrt{X(f)}\gamma^{-1}(f)C_m^*(f), \quad m = 1,...,M$$

where

$$K(f) = \sqrt{\frac{E/\sigma_d^2}{\int_{-\infty}^{+\infty} S_w(f) \frac{X(f)}{\gamma(f)} df}} S_w^{1/2}(f). \tag{26}$$

This selection of filters achieves the maximal SNR available with one-sided adjustment,

$$\mathrm{SNR}_1 = E x_0 \left[\int_{-\infty}^{+\infty} S_w(f) \frac{X(f)}{\gamma(f)} df \right]^{-1}. \tag{27}$$

In the uplink transmission case with restricted computational complexity, the transmit filter is simply chosen as the standard (e.g., square-root raised cosine) function, and an analogous optimization procedure results in the following solution:

$$G_0(f) = K\sqrt{X(f)},$$

$$G_m(f) = K^{-1}\sqrt{X(f)}\gamma^{-1}(f)C_m^*(f), \quad m = 1,...,M$$

where

$$K = \sqrt{\frac{E/\sigma_d^2}{x_0}}. \tag{28}$$

The same SNR, SNR_1 (27), is achieved.

Comparing the SNR available with and without complexity restriction, we find that $\mathrm{SNR}_1 \leqslant \mathrm{SNR}_2$. The two SNRs are equal only when $\gamma(f)$ is proportional to $S_w(f)$.

In what follows, we shall focus on the usual case of white noise, $S_w(f) = N_0$. Note that the factors $K(f)$ then become constants independent of frequency in both two-sided and one-sided focusing, and the same set of filters may be used for uplink and downlink transmission. The SNR expressions reduce to

$$\mathrm{SNR}_2 = \frac{E}{N_0} x_0^2 \left[\int_{-\infty}^{+\infty} \frac{X(f)}{\sqrt{\gamma(f)}} df \right]^{-2} \tag{29}$$

and

$$\mathrm{SNR}_1 = \frac{E}{N_0} x_0 \left[\int_{-\infty}^{+\infty} \frac{X(f)}{\gamma(f)} df \right]^{-1}. \tag{30}$$

We now want to compare these values of SNR, achieved through optimal focusing, to the SNR achieved by time-reversal and methods based on equalization.

B. Time-reversal performance with residual ISI

When no care is taken to ensure focusing, the samples of the received signal contain residual ISI:

$$y(nT) = f(0)d(n) + \sum_{k \neq n} f(kT)d(n-k) + z(nT). \tag{31}$$

Assuming uncorrelated data symbols, the SNR is given by

$$\mathrm{SNR}_0 = \frac{\sigma_d^2 |f^2(0)|}{\sigma_d^2 \Sigma_{k \neq 0} |f^2(kT)| + \sigma_z^2}. \tag{32}$$

We look at the following scenarios. On the uplink, the transmitter uses a standard filter, $G_0(f) = K_u \sqrt{X(f)}$, and the receiver uses $G_m(f) = G_0^*(f) C_m^*(f)$, $m = 1,...,M$. This scenario is analogous to ideal (noiseless) passive phase-conjugation. On the downlink, the transmitter uses $G_m(f) = K_d \sqrt{X(f)} C_m^*(f)$, $m = 1,...,M$, and the receive filter is simply $G_0(f) = \sqrt{X(f)}$. This scenario is analogous to active phase-conjugation. The constants K_u, K_d are determined from the energy constraint (8). The resulting SNR is the same in the uplink and the downlink scenarios, and it is given by

$$\mathrm{SNR}_0 = \frac{\dfrac{E}{N_0}}{\dfrac{E}{N_0}\rho + \dfrac{x_0}{\int_{-\infty}^{+\infty} X(f)\gamma(f) df}}, \tag{33}$$

where

$$\rho = \frac{\Sigma_{k \neq 0} |f^2(kT)|}{|f^2(0)|} = \frac{T \int_{-1/2T}^{+1/2T} |X\gamma[f]|^2 df}{\left| \int_{-\infty}^{+\infty} X(f)\gamma(f) df \right|^2} - 1 \tag{34}$$

and $X\gamma[f]$ is used to denote the folded spectrum of $X(f)\gamma(f)$:

$$X\gamma[f] = \frac{1}{T} \sum_{k=-\infty}^{+\infty} X\left(f + \frac{k}{T}\right) \gamma\left(f + \frac{k}{T}\right). \tag{35}$$

It is interesting to observe that as the noise vanishes, i.e., $E/N_0 \to +\infty$, unlike with optimal focusing when $\mathrm{SNR}_{1,2} \to +\infty$, the performance of time-reversal saturates: $\mathrm{SNR}_0 \to 1/\rho$. The value of ρ depends on the channel characteristics, expressed through the function $\gamma(f)$, and the system bandwidth, expressed through the function $X(f)$.

C. Time-reversal performance with equalization

The performance of time-reversal saturates because of residual ISI. To overcome this limitation, an equalizer may be used. We look at the downlink scenario, where an optimal, minimum mean squared error (MMSE) linear processor is employed. It consists of a receiving matched filter followed by a symbol rate sampler and a linear MMSE equalizer. For the received signal samples given in the form (31) with uncorrelated data sequence, the MMSE equalizer has a transfer function

$$A[f] = \frac{\sigma_d^2 F^*[f]}{\sigma_d^2 |F[f]|^2 + S_z[f]}, \tag{36}$$

where $F[f]$ is the folded spectrum of the overall response $F(f)$, and $S_z[f]$ is the power spectral density of the discrete-time noise process $z(nT)$. The receiving filter is matched to the overall response,

$$G_0(f) = \sum_{m=1}^{M} G_m^*(f) C_m^*(f) \tag{37}$$

so that $F(f) = |G_0^2(f)|$, and, hence, $S_z[f] = N_0 F[f]$. The equalizer transfer function thus reduces to

$$A[f] = \frac{\dfrac{\sigma_d^2}{N_0}}{1 + \dfrac{\sigma_d^2}{N_0} F[f]}. \tag{38}$$

The SNR at the equalizer output is[27]

$$\mathrm{SNR} = \frac{1}{\mathrm{MSE}} - 1 = \left[T \int_{-1/2T}^{+1/2T} \frac{1}{1 + \dfrac{\sigma_d^2}{N_0} F[f]} df \right]^{-1} - 1. \tag{39}$$

For the transmit filter selection as in active phase-conjugation, $G_m(f) = K_d \sqrt{X(f)} C_m^*(f)$, $m = 1,...,M$, we have that

$$\frac{\sigma_d^2}{N_0} F(f) = \frac{E/N_0}{\int_{-\infty}^{+\infty} X(f)\gamma(f) df} X(f)\gamma^2(f). \tag{40}$$

This transfer function is used to compute the resulting SNR (39):

$$\mathrm{SNR}_{3,\mathrm{tr}} = \left[T \int_{-1/2T}^{+1/2T} \frac{1}{1 + \dfrac{E/N_0}{\int_{-\infty}^{+\infty} X(f)\gamma(f) df} X\gamma^2[f]} df \right]^{-1} - 1 \tag{41}$$

where $X\gamma^2[f]$ is the folded spectrum of $X(f)\gamma^2(f)$.

In the uplink transmission case, time-reversal filtering at the receiver is followed by equalization. Because passive phase-conjugation is equivalent to matched filtering, and the transmitter uses a fixed filter, this case is identical to standard equalization, which is treated next.

D. Equalizer performance

A standard equalizer does not rely on time-reversal at the transmitter, but instead uses pre-determined, channel-independent filters. In the downlink case, the transmit filters are $G_m(f)=K_d\sqrt{X(f)}$, $m=1,...,M$, and we have that

$$\frac{\sigma_d^2}{N_0}F(f)=\frac{E/N_0}{Mx_0}X(f)\Sigma^2(f), \tag{42}$$

where

$$\Sigma(f)=\left|\sum_{m=1}^{M}C_m(f)\right|. \tag{43}$$

Note that this case may represent a poor system design as transmission over multiple channels with different delays creates additional time spreading, and the channel transfer functions add directly in the above expression, possibly in a destructive manner. The resulting SNR is computed from Eq. (39):

$$\mathrm{SNR}_{3,\mathrm{down}}=\left[T\int_{-1/2T}^{+1/2T}\frac{1}{1+\dfrac{E/N_0}{M\int_{-\infty}^{+\infty}X(f)df}X\Sigma^2[f]}df\right]^{-1}-1, \tag{44}$$

where $X\Sigma^2[f]$ is the folded spectrum of $X(f)\Sigma^2(f)$.

In the uplink scenario, the MMSE linear processor consists of a bank of matched filters, $G_m(f)=G_0^*(f)C_m^*(f)$, $m=1,...,M$, as in passive phase-conjugation, whose outputs are summed, sampled at the symbol rate, and processed by a linear equalizer. This process is also called multichannel equalization. The MMSE equalizer is again defined by Eq. (38) where the overall transfer function is now $F(f)=|G_0^2(f)|\gamma(f)$. For the standard transmit filter selection, $G_0(f)=K_u\sqrt{X(f)}$, we have that

$$\frac{\sigma_d^2}{N_0}F(f)=\frac{E/N_0}{x_0}X(f)\gamma(f). \tag{45}$$

The resulting SNR is computed from Eq. (39):

$$\mathrm{SNR}_{3,\mathrm{up}}=\left[T\int_{-1/2T}^{+1/2T}\frac{1}{1+\dfrac{E/N_0}{\int_{-\infty}^{+\infty}X(f)df}X\gamma[f]}df\right]^{-1}-1, \tag{46}$$

where, as before, $X\gamma[f]$ is the folded spectrum of $X(f)\gamma(f)$.

Comparing uplink and downlink equalization, we have that $\mathrm{SNR}_{3,\mathrm{up}}\geqslant\mathrm{SNR}_{3,\mathrm{down}}$. The two are equal if the channel transfer functions $C_m(f)$, $m=1,...,M$ are identical and constant within the signal bandwidth. It is not clear, however, how $\mathrm{SNR}_{3,\mathrm{down}}$ compares with $\mathrm{SNR}_{3,\mathrm{tr}}$, i.e., what is the advantage, if any, of using transmit time-reversal in conjunction with equalization. This question gives rise to a broader one of optimal transmit filtering for use with equalization.

E. Transmitter/receiver optimization for a system with equalization

So far, we have looked at optimal focusing (filter optimization under no-ISI constraint) and at MMSE equalization using *a priori* selected transmit filters. However, it is possible to look at a system in which both channel-dependent transmit filtering *and* equalization are used. In other words, if the requirement for no ISI is relaxed in the optimal system design, and the equalizer is allowed at the receiver, the question is what transmit/receive filtering should be used to maximize the SNR. Note that because this optimization criterion is less restrictive than that of focusing (the no-ISI constraint has been removed) improved performance may be expected. Also, performance must be improved with respect to standard equalization, which represents only a special case of transmit filter selection.

For any given transmit filtering, an optimal linear receiver consists of a matched filter (or a bank of matched filters for the uplink scenario) followed by a symbol-spaced MMSE equalizer (36). Both the matched filter and the equalizer transfer functions depend on the transmit filter selection, and consequently, so does the achieved SNR. We want to find the transmit filter(s) for which the SNR at the equalizer output (39) is maximized. Transmit filtering, in turn, will determine receive filtering and the equalizer.

Let us first consider the uplink case. Maximizing the SNR is equivalent to minimizing the MSE, which is defined by the overall transfer function $F(f)=|G_0^2(f)|\gamma(f)$. We assume that the system operates in minimal bandwidth B required to support ISI-free transmission at symbol rate $1/T$, i.e., $G_0(f)$ is zero for $|f|>1/2T$. Then, the optimization problem is to find the function $G_0(f)$ for which

$$\mathrm{MSE}_{\mathrm{up}}=T\int_{-1/2T}^{+1/2T}\frac{1}{1+\dfrac{\sigma_d^2}{N_0T}|G_0^2(f)|\gamma(f)}df \tag{47}$$

is minimized, subject to the constraint on transmitted energy,

$$\sigma_d^2\int_{-1/2T}^{1/2T}|G_0^2(f)|df=E. \tag{48}$$

In the downlink case, the MSE is defined by the overall transfer function $F(f)=|\Sigma_{m=1}^{M}G_m(f)C_m(f)|^2$, and we want to find a set of functions $G_m(f)$, $m=1,...,M$ for which

$$\mathrm{MSE}_{\mathrm{down}}=T\int_{-1/2T}^{+1/2T}\frac{1}{1+\dfrac{\sigma_d^2}{N_0T}|\Sigma_{m=1}^{M}G_m(f)C_m(f)|^2}df \tag{49}$$

is minimized, subject to the constraint on transmitted energy,

$$\sigma_d^2\int_{-1/2T}^{1/2T}\sum_{m=1}^{M}|G_m^2(f)|df=E. \tag{50}$$

Realizing that the downlink MSE is bounded by

$$\mathrm{MSE}_{\mathrm{down}}\geqslant T\int_{-1/2T}^{+1/2T}\frac{1}{1+\dfrac{\sigma_d^2}{N_0T}|\alpha^2(f)|\gamma(f)}df, \tag{51}$$

which is achieved for $G_m(f)=\alpha(f)C_m^*(f)$, $m=1,...,M$, the downlink optimization problem can be reduced to the same form as that of the uplink problem. Namely, if we define

$$\Phi(f)=\frac{\sigma_d^2}{N_0T}\times\begin{cases}|G_0(f)|^2 & \text{for uplink transmission}\\ |\alpha(f)|^2\gamma(f) & \text{for downlink transmission,}\end{cases}$$

$$|f|\leqslant\frac{1}{2T} \tag{52}$$

then

$$\mathrm{MSE}=T\int_{-1/2T}^{1/2T}\frac{1}{1+\Phi(f)\gamma(f)}df \tag{53}$$

is to be minimized with respect to a real, non-negative function $\Phi(f)$, subject to the constraint that

$$T\int_{-1/2T}^{1/2T}\Phi(f)df=\frac{E}{N_0}. \tag{54}$$

Using the Lagrange method, we form

$$\Lambda(\Phi)=T\int_{-1/2T}^{1/2T}\frac{1}{1+\Phi(f)\gamma(f)}df+\lambda\left[T\int_{-1/2T}^{1/2T}\Phi(f)df-\frac{E}{N_0}\right], \tag{55}$$

where λ is a constant. Differentiating the above function with respect to Φ, and setting the derivative equal to zero, provides the following solution:

$$\Phi(f)=\frac{1}{\gamma(f)}\left[\frac{1}{\sqrt{\lambda}}\sqrt{\gamma(f)}-1\right],\quad f\in B_0=\left[-\frac{1}{2T},\frac{1}{2T}\right]. \tag{56}$$

Substituting this solution into the constraint (54) we obtain

$$\frac{1}{\sqrt{\lambda}}=K_0=\frac{\frac{E}{N_0}+T\int_{-1/2T}^{1/2T}\frac{1}{\gamma(f)}df}{T\int_{-1/2T}^{1/2T}\frac{1}{\sqrt{\gamma(f)}}df}. \tag{57}$$

To ensure a valid solution for $\Phi(f)$, we must verify that $K_0\geqslant 1/\sqrt{\gamma(f)}$, $\forall f\in[-1/2T,1/2T]$. This condition will hold if $K_0\geqslant 1/\sqrt{\gamma_{\min}}$, where $\gamma_{\min}$ is the smallest value of $\gamma(f)$ within the available bandwidth,

$$\gamma_{\min}=\min_{f\in B_0}\{\gamma(f)\}. \tag{58}$$

If this is not the case, the expression (56) does not represent a valid solution. We then modify the solution as follows:

$$\Phi(f)=\begin{cases}\frac{1}{\gamma(f)}[K_L\sqrt{\gamma(f)}-1] & f\in B_L\subset B_0\\ 0 \quad \text{otherwise}\end{cases}, \tag{59}$$

where K_L is determined from the energy constraint (54),

$$K_L=\frac{\frac{E}{N_0}+T\int_{B_L}\frac{1}{\gamma(f)}df}{T\int_{B_L}\frac{1}{\sqrt{\gamma(f)}}df} \tag{60}$$

and B_L is the maximal bandwidth for which $K_L\geqslant 1/\sqrt{\gamma(f)}$, $\forall f\in B_L$.

To gain insight into this definition, we note that whenever it is decided a priori that $\Phi(f)$ is zero for some frequency region $\bar{B}_L$ (the complement of B_L within B_0), the MSE (53) can be expressed as

$$\mathrm{MSE}=T\int_{\bar{B}_L}df+T\int_{B_L}\frac{1}{1+\Phi(f)\gamma(f)}df \tag{61}$$

and the energy constraint becomes

$$T\int_{B_L}\Phi(f)df=\frac{E}{N_0}. \tag{62}$$

The solution (59) then represents $\Phi(f)$ for which the second MSE term is minimized subject to the energy constraint. In order to minimize the first MSE term as well, the smallest possible frequency region $\bar{B}_L$ (i.e., the largest B_L) should be chosen, hence the definition of B_L.

We further define the set of frequencies B_L to be

$$B_L=\{f\in B_0:\gamma(f)\geqslant\gamma_L\}, \tag{63}$$

where γ_L is the smallest value of $\gamma(f)$ within B_0, for which $K_L\geqslant 1/\sqrt{\gamma_L}$,

$$\gamma_L=\min_{f\in B_0:K_L\geqslant 1/\sqrt{\gamma_L}}\{\gamma(f)\}. \tag{64}$$

When B_L is defined via the threshold γ_L, then this threshold should be minimized. Separation of the frequency regions B_L and $\bar{B}_L$ based on thresholding of the channel function $\gamma(f)$ is intuitively satisfying, because it states that if transmit energy is limited, it should not be wasted on those regions where $\gamma(f)$ is low.

The solution for K_L can be obtained numerically, starting with $B_L=B_0=[-1/2T,1/2T]$. If the resulting $K_L=K_0\geqslant 1/\sqrt{\gamma_{\min}}$, then $\Phi(f)$ has a full nonzero solution (56). If this is not the case, K_L is computed from Eq. (60) iteratively, increasing γ_L from the initial value $\gamma_{\min}$ by a small amount $\Delta\gamma_L$ in each step, until the condition $K_L\geqslant 1/\sqrt{\gamma_L}$ is met. The search for γ_L then stops, and the solution for $\Phi(f)$ follows from Eq. (59).

The desired transmit/receive filters are given in the following:

$$\begin{aligned}\text{uplink:}\quad & G_0(f)=K\sqrt{\Phi(f)},\\ & G_m(f)=K\sqrt{\Phi(f)}C_m^*(f),\quad m=1,...,M,\\ \text{downlink:}\quad & G_m(f)=K\sqrt{\Phi(f)}\gamma^{-1/2}(f)C_m^*(f),\\ & \qquad m=1,...,M,\\ & G_0(f)=K\sqrt{\Phi(f)}\gamma^{1/2}(f),\end{aligned}$$

where

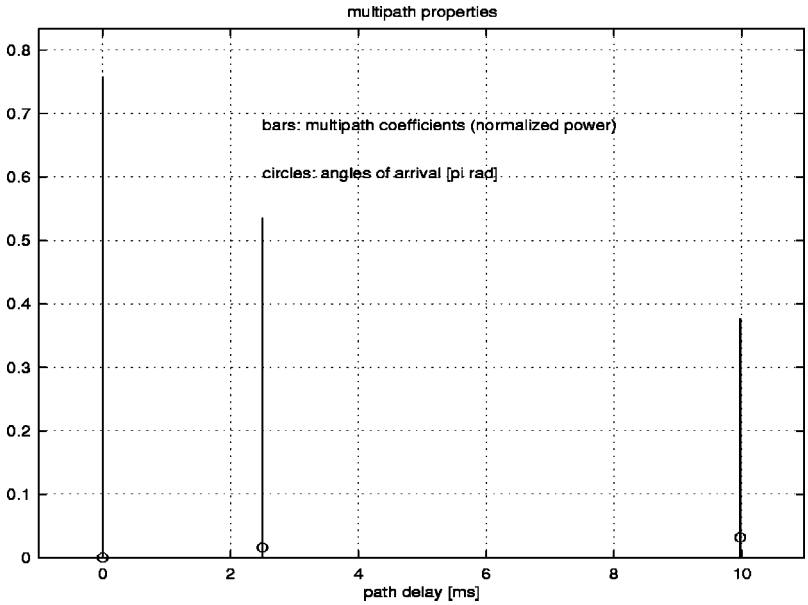


FIG. 2. Multipath characteristics of the example channel: path gain magnitudes $|c_p|$ and angles of arrival θ_p are shown at corresponding delays τ_p (reference delay is $\tau_0=0$).

$$K=\sqrt{N_0T/\sigma_d^2}. \tag{65}$$

The equalizer transfer function (up to the normalizing constant σ_d^2/N_0) is the same for uplink and for downlink transmission:

$$A[f]=\frac{1}{1+\Phi(f)\gamma(f)}, \quad |f|\leqslant\frac{1}{2T}. \tag{66}$$

This system achieves the SNR,

$$\mathrm{SNR}_4=\left\{1-T\int_{B_L}df+\left[T\int_{B_L}\frac{1}{\sqrt{\gamma(f)}}df\right]^2 \times\left[\frac{E}{N_0}+T\int_{B_L}\frac{1}{\gamma(f)}df\right]^{-1}\right\}^{-1}-1. \tag{67}$$

In the case when $B_L=B_0$, it is easily shown that

$$\mathrm{SNR}_4=\mathrm{SNR}_2+\frac{\mathrm{SNR}_2}{\mathrm{SNR}_1}-1\geqslant\mathrm{SNR}_2. \tag{68}$$

Thus, this signaling scheme outperforms optimal focusing. The question, of course, is how great is the difference in performance.

III. PERFORMANCE COMPARISON

In this section, we use an illustrative example to compare the performance of various techniques discussed: optimal focusing with two-sided filter adjustment, optimal focusing with one-sided filter adjustment, time-reversal, time-reversal in conjunction with equalization, standard equalization (with fixed transmit filters), and equalization using optimized transmit filters. Performance is evaluated through numerical computation of the analytical SNR expressions for a particular channel model.

A. Channel model

The channel model is based on geometry of shallow water multipath. We look at repeated surface-bottom reflections and take into account a certain number of multipath arrivals, P. Each multipath component is characterized by a gain c_p, delay τ_p, and angle of arrival θ_p, which are computed from the propagation path length l_p. The path gain magnitude is computed as $|c_p|=\Gamma_p/\sqrt{A(l_p)}$, where $\Gamma_p\leqslant 1$ may be used to model loss due to reflection (we choose each reflection to introduce a $\sqrt{2}$ loss in amplitude) and $A(l_p)$ is the nominal acoustic propagation loss, $A(l_p)=l_p^k[a(f_c)]^{l_p}$, calculated assuming practical spreading, $k=1.5$, a carrier frequency $f_c=15$ kHz, and absorption according to Thorp. [For f_c in kHz, $a(f_c)$ is given in dB/km as $10\log a(f_c)=0.11f_c^2/(1+f_c^2)+44f_c^2/(4100+f_c^2)+2.75\times10^{-4}f_c^2+0.003$.] The path gain phase is computed as $\angle c_p=-2\pi f_c\tau_p$. Observed across the array, there is a phase delay $\varphi_p=2\pi(d/\lambda_c)\sin\theta_p$ between the elements spaced by d, where $\lambda_c=c/f_c$, and $c=1500$ m/s is the nominal sound speed. In reference to the first element, the channel transfer functions are given by

$$C_m(f)=\sum_{p=0}^{P-1}c_{m,p}e^{-j2\pi f\tau_p},$$

where

$$c_{m,p}=c_pe^{-j(m-1)\varphi_p}, \quad m=1,...,M. \tag{69}$$

As an example, we use a channel of depth 75 m, range of 3 km, and the system mounted near the bottom. Three propagation paths are taken into account (direct, surface reflected, and surface-bottom-surface reflected). Figure 2 shows the resulting multipath profile of the channel. We note that the total multipath spread is 10 ms, which is on the order of that observed experimentally.

The channel function $\gamma(f)$ is shown in Fig. 3 for $M=4$ and $M=32$. Shown on the same plot is the desired system response $X(f)$ chosen as a raised cosine with roll-off factor close to 0, which provides maximal bit rate for ISI-free transmission in a bandwidth $B=1/T$. The symbol duration is chosen to be $T=0.2$ ms, corresponding to the bandwidth of 5 kHz and transmission at 10 kbits/s if 4-PSK is used, or 15 kbits/s if 8-PSK is used. The impulse response of

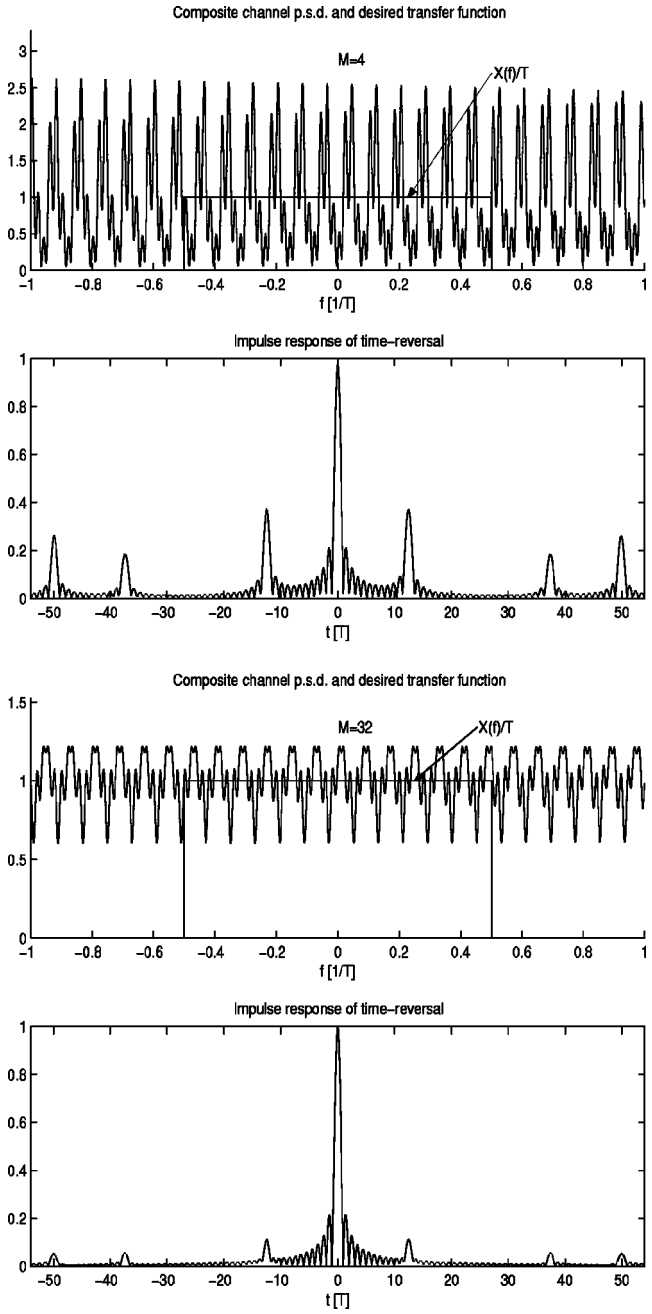


FIG. 3. Composite channel power spectral density $\gamma(f)$, and the impulse response of time-reversal which corresponds to $X(f)\gamma(f)$. Multipath coefficients are normalized such that $M\Sigma_p|c_p^2|=1$, and half-wavelength spacing between array elements is used.

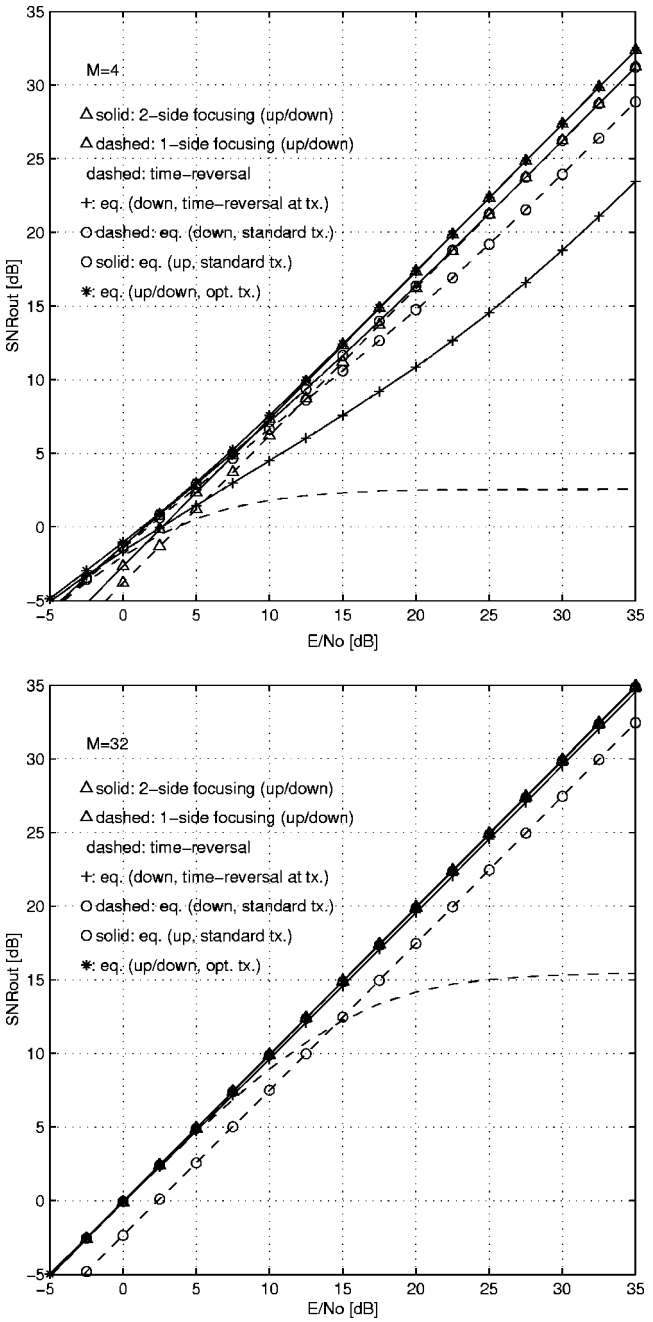


FIG. 4. Performance of various techniques on the example channel: output SNR vs E/N_0 for fixed number of array elements M.

the overall system obtained with time-reversal is also shown, and it is evidently far from ideal. As the number of array elements is increased, $\gamma(f)$ tends to flatten out, resulting in better, but not complete suppression of multipath through time-reversal.

B. Performance analysis

Figure 4 summarizes performance results for the two examples. Let us focus on the $M=4$ case. We first confirm that two-sided focusing [Eq. (29), solid curve labeled "Δ"] outperforms one-sided focusing [Eq. (30), dashed curve labeled "Δ"], but more interestingly, we observe that the difference in performance is small. This is an encouraging observation from the viewpoint of designing a practical system with restricted processing complexity. The performance of time-reversal [Eq. (33), dashed curve] is inferior to optimal focusing and to all other schemes at practical values of SNR. By practical, those values are meant that yield at least several dB of output SNR, as this is required for an adaptive system to perform in a decision-directed manner. The loss of time-reversal becomes quite large even at a moderate E/N_0 of 10–15 dB, and the performance saturates thereafter at a value $1/\rho$ determined by the channel (34). Some of the loss is recovered by the use of an equalizer in conjunction with transmit time-reversal [Eq. (41), curve labeled "+"]. However, it is interesting to observe that this system compares poorly with the standard equalizer that uses square-root raised cosine transmit filters and equal energy allocation

across the array [Eq. (44), dashed curve labeled "○"]. Standard downlink equalization is inferior to its uplink counterpart [Eq. (46), solid curve labeled "○"], which offers a consistently good performance. Note that this performance is closely matched by one-sided focusing at moderate to high SNR. Finally, we confirm that equalization using optimized transmit filters [Eq. (67), curve labeled "*"] provides an upper bound on the performance of all other schemes. More important, we observe that this scheme offers negligible improvement over focusing, which allows a much easier implementation.

With $M=32$, the performance of time-reversal improves; however, the saturation effect is still notable. Nonetheless, it has to be noted that at low and moderate E/N_0 up to about 10 or 15 dB, with the increased number of elements, time-reversal becomes the technique of choice as it offers near-optimal performance at minimal computational complexity. Equalization in conjunction with transmit time-reversal now outperforms standard downlink equalization, while the performance of both focusing methods, as well as that of standard uplink equalization, tends to the same optimal curve. Comparing the performance achieved with 32 and with 4 elements demonstrates that optimal focusing is much less sensitive to the array size than either of the techniques based on time-reversal (recall also that the power in the channel is kept constant with changing M).

Performance sensitivity to the array size is summarized in Fig. 5, which shows the output SNR as a function of M for a given symbol SNR E/N_0. Two values of E/N_0 are taken as an example, 10 and 20 dB. In each instance, we note superiority of focusing methods over time-reversal. Performance of focusing methods shows fast improvement with initial increase in M, and a small increment thereafter. Hence, good performance can be achieved without unduly increasing the number of array elements. For example, if one-sided focusing is used, and $E/N_0 \geqslant 10$ dB, increasing the number of elements beyond 6 offers less than 3 dB total improvement in performance. In contrast to this situation, time-reversal steadily gains in performance within the range of M shown; however, it fails to achieve the performance of focusing methods. Most important, we observe that at moderate to high symbol SNR, one-sided focusing needs a relatively small number of array elements to approach the optimal performance.

Results of Fig. 5 also offer an interesting comparison between standard equalization and equalization in conjunction with transmit time-reversal. Using time-reversal at the transmitter offers an improvement provided that M is greater than a certain number. In the example considered, this number is 11 at E/N_0 of 10 dB, and 8 at E/N_0 of 20 dB.

So far, we have considered up to $M=32$ elements, but it is interesting to observe the performance of time-reversal with a further increase in M, to see what array size is needed to bring its performance to that of other techniques. Figure 6 shows the performance of time-reversal for an extended range of M. Shown on the same plot is the performance of optimal focusing, which, on this scale, is indistinguishable from the system bound or multichannel equalization. An interesting effect is immediately apparent: the performance of

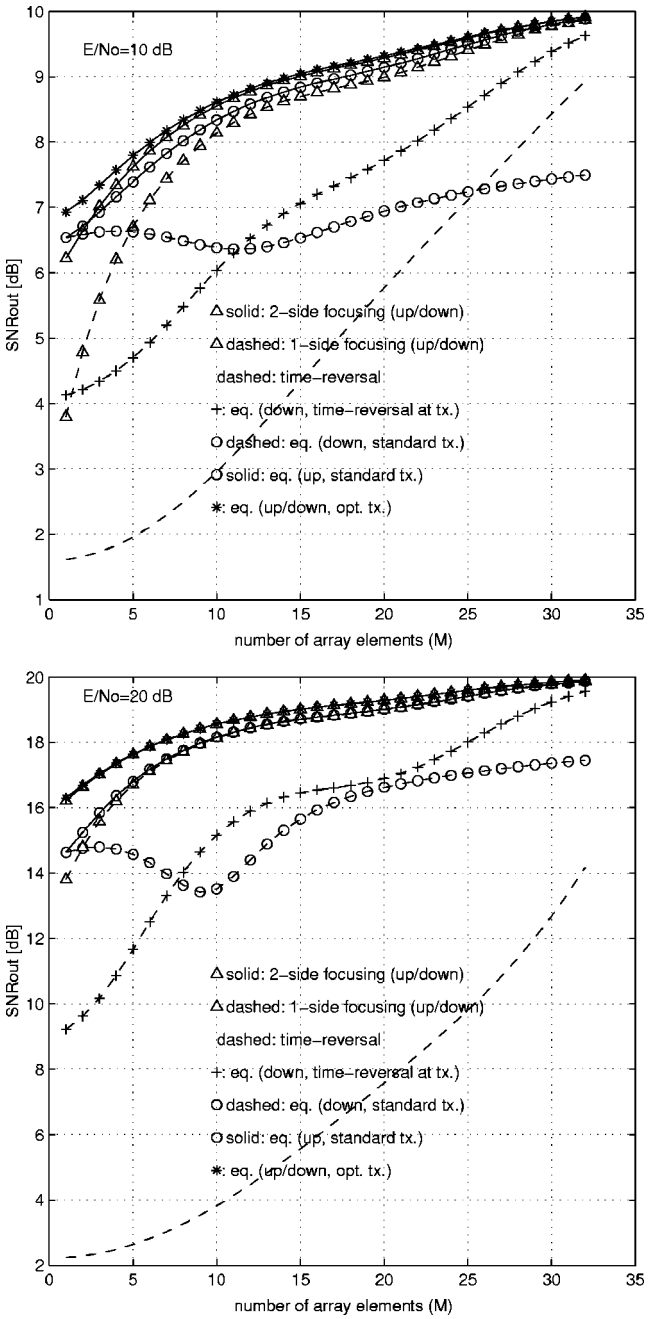


FIG. 5. Performance of various techniques on the example channel: output SNR vs M for fixed symbol SNR E/N_0.

time-reversal does not consistently improve with increasing M (as does the performance of focusing techniques and uplink equalization) but instead exhibits an oscillatory behavior, tending to the optimum only as $M \to \infty$. The values of M for which the performance is best are those values for which it happens so that the composite channel function $\gamma(f)$ flattens out almost completely, i.e., $\gamma(f) \approx 1$. At these values of M, time-reversal approaches the performance bound. However, due to the nature of the function $\gamma(f)$, the performance does not remain at optimum, but deviates from it with an increase in M. This fact underlines the suboptimality of system design based on time-reversal only. In practice, it could be difficult to rely on finding the optimal number of elements every time the array is deployed and system configuration changes.

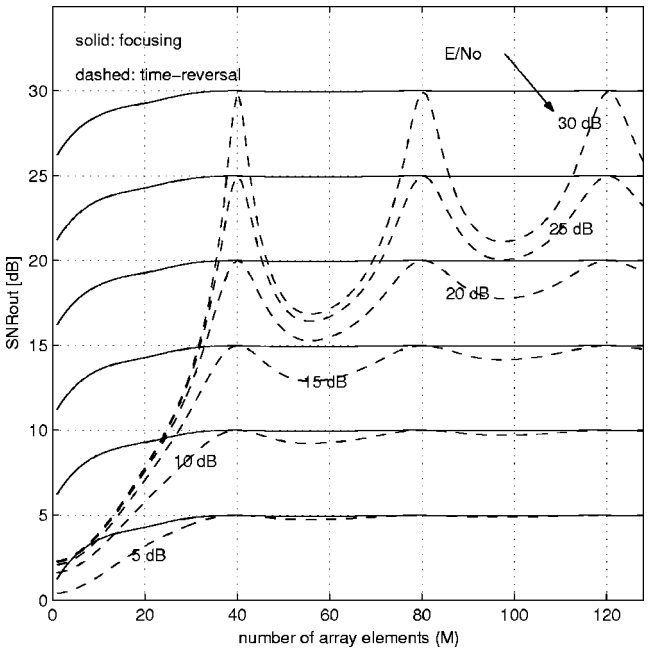


FIG. 6. Performance of time-reversal and optimal focusing for extended range of M.

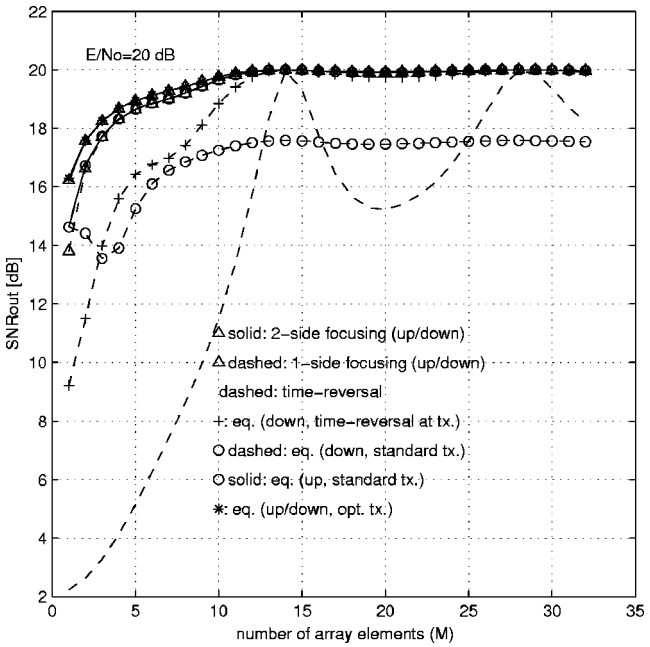


FIG. 8. Output SNR vs M for the three-path channel model, $d=1.41\lambda_c$.

Finally, we investigate performance sensitivity to the changes in multipath composition and the array element spacing, both of which influence the channel function $\gamma(f)$. Figure 7 shows performance results obtained for the same channel model, but with six, instead of three multipath components taken into account. The total multipath spread is now somewhat greater than 60 ms, with the additional arrivals' strength approximately 9, 12, and 15 dB below the principal arrival. The performance differs little as compared to the three-path channel. If anywhere, the difference can be seen when only a few elements are used—the performance of time-reversal is then worse on the six-path channel, while that of other techniques is better. More important, we observe that the same conclusions regarding performance comparison between different techniques hold in the presence of extended multipath. To assess performance sensitivity to changes in relative strength of the multipath arrivals, a hypothetical case of a lossless six-path channel was investigated. Suffice it to say that the same general conclusions were made in this case.

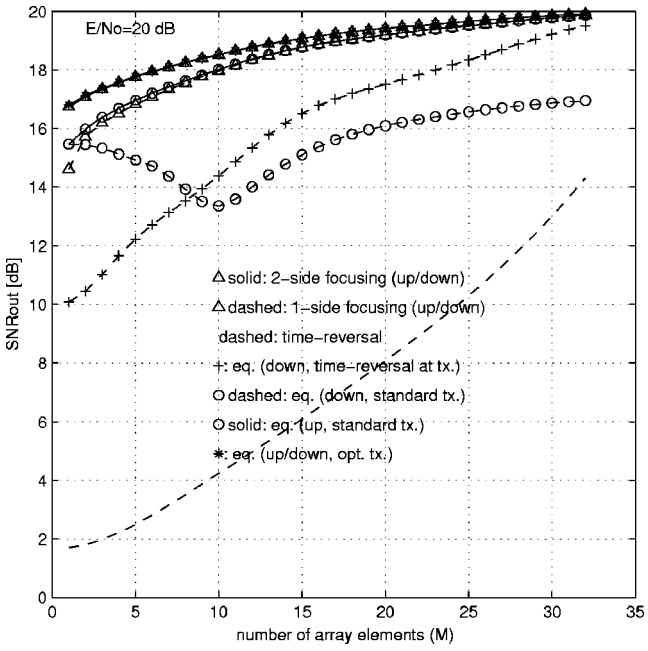


FIG. 7. Output SNR vs M for the six-path channel model, $d=\lambda_c/2$.

The effect of changing the array element spacing is illustrated in Fig. 8. This figure shows performance results for the original three-path channel, but with $d=1.41\lambda_c$ instead of $d=\lambda_c/2$. Evidently, the performance of time-reversal shows more sensitivity to the changes in the array element spacing than the other techniques (the same would be true for changes in the carrier frequency). The optimum is now approached with a smaller number of elements (this number corresponds to the same total array length, i.e., 40 elements spaced by $\lambda_c/2$ or 14 elements spaced by $1.41\lambda_c$ in the example considered). However, the improvement in performance is not consistent with an increase in the element spacing. For a given number of elements $M>1$, performance improves with an initial increase in element spacing, but exhibits an oscillatory behavior afterwards. For example, with $M=4$, performance starts to degrade after element spacing increases beyond $5\lambda_c$. With elements spaced by $10\lambda_c$, at $E/N_0=20\,\mathrm{dB}$, the output SNR does not exceed a value of about 8 dB if more than two elements are used in a time-reversal array. Compared to this situation, performance of retrofocusing techniques shows negligible sensitivity to the changes in element spacing.

IV. CONCLUSION

A number of techniques have been investigated for communication over an underwater acoustic link where one end is equipped with a single transmit/receive element and another with an array. To achieve maximal bit rate within a fixed bandwidth, an optimization criterion of maximizing the

data detection SNR, while eliminating or minimizing the residual ISI was chosen. Transmit/receive filters were obtained analytically for uplink and downlink transmission, with varying degrees of system complexity. The so-obtained optimal techniques were compared to standard time-reversal and equalization.

Because it ignores residual ISI, time-reversal exhibits performance saturation, and strongly depends on the use of a large array. When this can be afforded (e.g., in a network whose base station uses an array to isolate multiple users) time-reversal offers a solution for minimal-complexity processing. With a smaller array, however, standard equalization outperforms time-reversal, and the use of equalization in conjunction with transmit time-reversal does not guarantee performance improvement over standard equalization, unless the number of elements exceeds a certain value. The proposed method of spatiotemporal retrofocusing guarantees maximal SNR and elimination of ISI for an arbitrary array size. It outperforms time-reversal at the expense of additional filtering. The filters needed for optimal focusing include phase-conjugation augmented by a channel-dependent scaling function. If filter adjustment is constrained to the array side only, one-sided focusing offers an excellent trade-off between complexity and performance. Its performance has only a small loss with respect to the two-sided focusing, and it stays close to that of uplink multichannel equalization. It thus represents a solution for systems that cannot deploy large arrays and have limited processing power. In addition, when two-way communication is to be established over a white-noise channel, the same set of filters can be used for transmission and reception.

The system analysis was completed by optimizing both ends of an equalization based system. While it is unlikely that a practical system would be based on this approach due to the difficulty of finding optimal filters, it provides an upper bound on the performance of all other techniques. Results demonstrate that optimal focusing performs very close to this bound.

Future work will concentrate on an experimental validation of spatiotemporal focusing aided by adaptive channel estimation. Two types of errors will guide the system performance: the error due to noise and the error due to time-variability of the channel. In particular, the latter may prove as the limiting factor for the performance of an acoustic system with a long round-trip delay and high rate of channel variation. Future analytical work will address system optimization with imperfect channel knowledge.

[1] M. Stojanovic, J. Catipovic, and J. Proakis, ''Reduced-complexity multichannel processing of underwater acoustic communication signals,'' J. Acoust. Soc. Am. **98**, 961–972 (1995).

[2] M. Fink, F. Wu, J.-L. Tomas, and D. Cassereau, ''Time-reversal of ultrasonic fields—Parts I, II and III,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**, 555–592 (1992).

[3] W. A. Kuperman, W. S. Hodgkiss, H. C. Song, T. Akal, C. Ferla, and D. R. Jackson, ''Phase conjugation in the ocean: Experimental demonstration of an acoustic time-reversal mirror,'' J. Acoust. Soc. Am. **103**, 25–40 (1998).

[4] W. S. Hodgkiss, H. C. Song, W. A. Kuperman, T. Akal, C. Ferla, and D. R. Jackson, ''A long-range and variable focus phase-conjugation experiment in shallow water,'' J. Acoust. Soc. Am. **105**, 1597–1604 (1999).

[5] W. S. Hodgkiss, J. D. Skinner, G. E. Edmonds, R. A. Harriss, and D. E. Ensberg, ''A high frequency phase-conjugation array,'' in Proceedings of the IEEE Oceans'01 Conference, pp. 1581–1585, 2001.

[6] J. S. Kim, H. C. Song, and W. A. Kuperman, ''Adaptive time-reversal mirror,'' J. Acoust. Soc. Am. **109**, 1817–1825 (2001).

[7] G. Edelmann, W. S. Hodgkiss, W. A. Kuperman, and H. C. Song, ''Underwater acoustic communication using time-reversal,'' in Proceedings of the IEEE Oceans'01 Conference, 2001.

[8] H. C. Song, G. Edelmann, S. Kim, W. S. Hodgkiss, W. A. Kuperman, and T. Akal, ''Low and high frequency ocean acoustic phase conjugation experiments,'' Proc. SPIE **4123**, 104–108 (2000).

[9] D. Jackson and D. Dowling, ''Phase-conjugation in underwater acoustics,'' J. Acoust. Soc. Am. **89**, 171–181 (1991).

[10] D. Rouseff, D. R. Jackson, W. L. J. Fox, C. D. Jones, J. A. Ritcey, and D. R. Dowling, ''Underwater acoustic communication by passive-phase-conjugation: Theory and experimental results,'' IEEE J. Ocean. Eng. **26**, 821–831 (2001).

[11] D. Jackson, J. A. Ritcey, W. L. J. Fox, C. D. Jones, D. Rouseff, and D. R. Dowling, ''Experimental testing of passive phase-conjugation for underwater acoustic communication,'' in Proceedings of the 34th Asilomar Conference, 2000, pp. 680–683.

[12] J. Flynn, W. L. J. Fox, J. A. Ritcey, D. R. Jackson, and D. Rouseff, ''Decision-directed passive phase-conjugation for underwater acoustic communications with results from a shallow-water trial,'' in Proceedings of the 35th Asilomar Conference, 2001, pp. 1420–1427.

[13] D. Rouseff, J. A. Flynn, W. L. J. Fox, and J. A. Ritcey, ''Decision-directed passive phase-conjugation for underwater acoustic communication: experimental results,'' in Proceedings of the IEEE Oceans'02 Conference (2002).

[14] J. A. Flynn, W. L. J. Fox, J. A. Ritcey, and D. Rouseff, ''Performance of reduced-complexity multi-channel equalizers for underwater acoustic communications,'' in Proceedings of the 36th Asilomar Conference, 2002, Vol. 1, pp. 453–460.

[15] J. Gomes and V. Barroso, ''The performance of sparse time-reversal mirrors in the context of underwater communications,'' in Proceedings of the 10th IEEE Workshop on Statistical Signal and Array Processing, 2000, pp. 727–731.

[16] J. Gomes and V. Barroso, ''A matched-field processing approach to underwater acoustic communication,'' in Proceedings of the IEEE Oceans'99 Conference, 1999, pp. 991–995.

[17] J. Gomes and V. Barroso, ''Asymmetric underwater acoustic communication using time-reversal mirror,'' in Proceedings of the IEEE Oceans'01 Conference, 2000.

[18] J. Gomes and V. Barroso, ''Wavefront segmentation in phase conjugate arrays for spatially modulated acoustic communication,'' in Proceedings of the IEEE Oceans'01 Conference, 2001, pp. 2236–2243.

[19] J. Gomes and V. Barroso, ''Ray-based analysis of a time-reversal mirror for underwater acoustic communication,'' in Proceedings of the ICASSP'00 Conference, 2000, pp. 2981–2984.

[20] J. Gomes and V. Barroso, ''Time-reversed communication over Doppler-spread underwater channels,'' in Proceedings of the ICASSP'02 Conference, 2002, pp. 2849–2852(III).

[21] P. Hursky, M. Porter, J. Rice, and V. McDonald, ''Passive phase-conjugate signaling using pulse-position modulation,'' in Proceedings of the IEEE Oceans'01 Conference, 2001.

[22] T. C. Yang, ''Temporal resolutions of time-reversal and passive phase-conjugation for underwater acoustic communications,'' IEEE J. Ocean. Eng. **28**, 229–245 (2003).

[23] J. Skim and K. C. Shin, ''Multiple focusing with adaptive time-reversal mirror,'' J. Acoust. Soc. Am. **115**, 600–606 (2004).

[24] D. Gesbert, M. Shafi, D. Shiu, P. J. Smith, and A. Naguib, ''From theory to practice: An overview of MIMO space-time coded wireless systems,'' IEEE J. Sel. Areas Commun. **21**, 281–302 (2003).

[25] M. Stojanovic and Z. Zvonar, ''Multichannel processing of broadband multiuser communication signals in shallow water acoustic channels,'' IEEE J. Ocean. Eng. **21**, 156–166 (1996).

[26] D. Kilfoyle, J. Preisig, and A. Baggeroer, ''Spatial modulation over partially coherent multi-input/multi-output channels,'' IEEE Trans. Signal Process. **51**, 794–804 (2003).

[27] J. G. Proakis, *Digital Communications* (Mc-Graw Hill, New York, 1995).

[28] M. Stojanovic, ''Efficient acoustic signal processing based on channel estimation for high rate underwater information,'' J. Acoust. Soc. Am. (submitted) (available at http://www.mit.edu/millitsa/publications.html)

The synthesis of spatially correlated random pressure fields

Stephen J. Elliott
Institute of Sound and Vibration Research, University of Southampton, Highfield, Southampton, SO17 1BJ, United Kingdom

Cédric Maury
Institute of Sound and Vibration Research, University of Southampton, Highfield, Southampton, SO17 1BJ, United Kingdom, and Université de Technologie de Compiègne, Laboratoire Roberval CNRS UMR 6066, Secteur Acoustique, BP20529, 60205, Compiègne Cedex, France

Paolo Gardonio
Institute of Sound and Vibration Research, University of Southampton, Highfield, Southampton, SO17 1BJ, United Kingdom

(Received 10 May 2004; revised 11 November 2004; accepted 27 November 2004)

The feasibility is considered of synthesizing a spatially correlated random pressure field having specified statistical properties. Of particular interest is the use of a near-field array of acoustic sources to synthesize a pressure field whose statistical properties are similar to either a diffuse acoustic sound field or to that generated by a turbulent boundary layer (TBL). A formulation based on least-squares filter design is presented. Initially, the more fundamental question is addressed of how many uncorrelated signal components are required to approximate the pressure field. A one-dimensional analysis suggests that two uncorrelated components per acoustic wavelength are required to approximate a diffuse pressure field. Similarly, for a TBL pressure field, about one uncorrelated component per correlation length is required in the spanwise direction and about two uncorrelated components per correlation length are required in the streamwise direction. These estimates are in good agreement with theoretical predictions for an infinite array, based on the Fourier transform of the spatial correlation function. When a full simulation is performed, including the acoustic effect of an appropriately positioned array of monopole sources, it is found that the number of acoustic sources required to reasonably approximate the diffuse or TBL pressure field is only slightly greater than the lower bound on this number, set by the number of uncorrelated components required. [DOI: 10.1121/1.1850231]

PACS numbers: 43.60.Fg, 43.28.Lv, 43.55.Br, 43.50.Nm [DKW] Pages: 1186–1201

I. INTRODUCTION

In this article a formulation is presented for using an array of acoustic sources to simulate the spatial correlation properties of a random pressure field, as specified by the spectral density matrix between the elements of an array of acoustic sensors. It is used in the subsequent sections to assess the feasibility of using such an array of sources close to a surface to generate a random pressure field with similar spatial statistical properties to those of either a diffuse acoustic sound field, or the pressure field generated by a turbulent boundary layer.

A diffuse acoustic field is normally defined as being due to an infinite number of uncorrelated plane waves, generated by remote acoustic sources, which, on average, provide equal incident energy from all directions.[1,2] It is characterized below in terms of its two-dimensional spatial correlation structure, which is the same in free space or when uncorrelated waves in a half-space fall on an infinite rigid surface.[3] Laboratory tests of the sound transmission properties of building structures are often conducted using random noise in a sound transmission suite. The sound field in the reverberant source room of a sound transmission suite is assumed to be diffuse for these measurements, but this is only a reasonable approximation for excitation frequencies above the room's Schröder frequency.[4,5] Below this frequency, individual modes of the source room can dominate the pressure field, which is then not diffuse. The use of a near-field array of acoustic sources, next to the structure being tested, to generate a diffuse pressure field in this low-frequency region has been considered by Bravo and Elliott[6] and in the present article we present a general theoretical background that can be used to better understand this problem.

The pressure field generated on the surface of a structure by a turbulent boundary layer (TBL) is also random in both time and space. In a fully developed turbulent boundary layer, the pressure fluctuations over a small region are stationary in time and homogeneous in space and can be characterized by their power spectral density at a point and their cross-spectral density between two separate points.[7,8] It is important to be able to measure the response of structures to TBL excitation, particularly in aeronautical applications, and this is currently achieved either with flight tests or in low noise wind tunnels. Both of these test methods are expensive and time consuming. These difficulties have led to the measurement of the acoustic properties of aircraft structures in sound transmission suites, although this is known to underestimate their sound reduction properties when exposed to TBL pressure fluctuations, as will be illustrated below.

Previously, Fahy[9] has considered the problem of simu-

lating TBL pressure fluctuations using jet noise, a siren tunnel, a loudspeaker or a shaker, but concluded that none of these single sources could reproduce the decay characteristics or spatial correlation properties in a TBL. Although there is a brief discussion in this paper about the use of an array of shakers, the electrical and mechanical difficulties of implementing such a system were considered, in 1966, to be such that it would be a difficult practical proposition. Robert and Sabot[10] have taken another approach to the laboratory simulation of TBL-excited structures, by noting that the modes of a structure are excited almost independently by such an excitation, and that there is generally a relatively small number of significantly excited structural modes. They considered the use of an array of appropriately-driven shakers acting on the panel, driven so that each structural mode was excited to the required extent. This approach assumes that the mode shapes of the structure are known and so, in practice, some preliminary modal analysis would have to be performed before the drive signals to the shakers could be determined. Dodds[11] has also considered combining together four independent random signals to generate the actuator drive signals in the problem of simulating the partially correlated displacements exciting the four wheels of a vehicle when driven over a rough road. These studies are related to the simulation of multiple channels of partially correlated random data, as described, for example, by Shinozuka,[12] Mignolet and Spanos,[13] and Soize and Poirion.[14] This data is typically used for numerical modelling of the dynamic response of structures, particularly nonlinear structures, to excitations due to the wind, earthquakes, and road excitation. Although the basis of the simulation method varies between these authors, they all begin the analysis with a factorization of the spectral density matrix of the target data, as used later in Sec. IV, and implicitly assume that the multiple channels of the simulated random data are generated by combining together an equal number of uncorrelated random signals.

The direct simulation of a spatially correlated random pressure field with a dense array of acoustic sources is considered here, by arranging for the sources to be mutually correlated so that they reproduce the spatial correlation properties of the target pressure field. The price that must be paid for such a general approach, however, is the large number of sources required for an accurate simulation, particular at high frequencies, as we shall see below. It is shown that the overall problem can be broken down into two parts. The first is concerned with the number of statistically uncorrelated signals required to accurately model the target spectral density matrix, which is related to the proper orthogonal decomposition (POD) of the pressure field. The second part is concerned with the invertibility of a matrix of physical responses between the acoustic sources and the sensors, over which the correlation properties of the target pressure field are reproduced.

II. SPATIAL CORRELATION FUNCTIONS

The spatial correlation structures of a diffuse acoustic field are compared with that of a simple model of a TBL pressure field at high subsonic Mach number, and the different effects that these two pressure fields have in exciting the response of a simple panel is discussed. The spatial correlation structure of the acoustic diffuse field, which is homogeneous and isotropic,[3] leads to an expression for the cross-spectral density between the pressures at two points, A and B, as

$$S_{\mathrm{AB}}(\omega) = S_{pp}(\omega)\frac{\sin k_\omega r}{k_\omega r}, \tag{1}$$

where $S_{pp}(\omega)$ is the power spectral density of the pressure at any point, r is the distance between the two measurement points, A and B, in any direction and k_ω is the acoustic wave number given by $k_\omega = \omega/c_0$ where c_0 is the speed of sound, which is taken later to be 343 ms^{-1}.

A widely used model for the correlation structure of a fully developed TBL, which will be used for illustration here, is that due to Corcos.[7] This predicts that the cross-spectral density between the pressures at two points, A and B, on a rigid surface is given by

$$S_{\mathrm{AB}}(\omega) = S_{pp}(\omega)e^{-|r_x|/L_x}e^{-|r_y|/L_y}e^{-j\omega r_y/U_c}, \tag{2}$$

where differences between the x and y coordinates of points A and B are denoted r_x and r_y. $S_{pp}(\omega)$ is again the power spectral density at any point, L_x is the correlation length in the spanwise direction, L_y is the correlation length in the streamwise direction, U_c is the convection velocity, and an $e^{j\omega t}$ time dependence has been assumed. It should be emphasized that this is an empirical model and, in contrast to the diffuse field, has not been derived from a theoretical understanding of the underlying physical phenomena. It should also be noted that the simulation of the random pressure field would not reproduce all the physics of a thin plate interacting with a turbulent flow, in particular the aeroelastic–mechanical coupling, as described, for example, by Clark and Frampton,[15] is not included.

The random pressure field is thus modeled as being homogeneous, but nonisotropic, and the Corcos model assumes that the variation of the correlation structure in the spanwise and streamwise directions is independent. In the simulations below the correlation lengths were assumed to be inversely proportional to frequency, and to have the form[7]

$$L_x = \frac{\alpha_x U_c}{\omega}, \tag{3}$$

$$L_y = \frac{\alpha_y U_c}{\omega}, \tag{4}$$

where α_x and α_y are constants, which were taken to be 1.2 and 8 in the simulations described below. The convection velocity was also assumed to be 135 ms^{-1}, which corresponds to a free-stream velocity of about 225 ms^{-1} and a Mach number of $M \approx 0.66$. Other authors have suggested different formulations for the correlation lengths in Eqs. (3) and (4), particularly at low frequencies (see, for example, Refs. 16 and 17) but still use the exponential decay of the spatial correlation in Eq. (2). In the limiting case of small r_x and r_y, it is known that viscous effects round off the cusp in the exponential peak of the correlation function in Eq. (2),[18] but in the frequency range of interest here, this occurs over a much smaller length scale than the correlation length. The

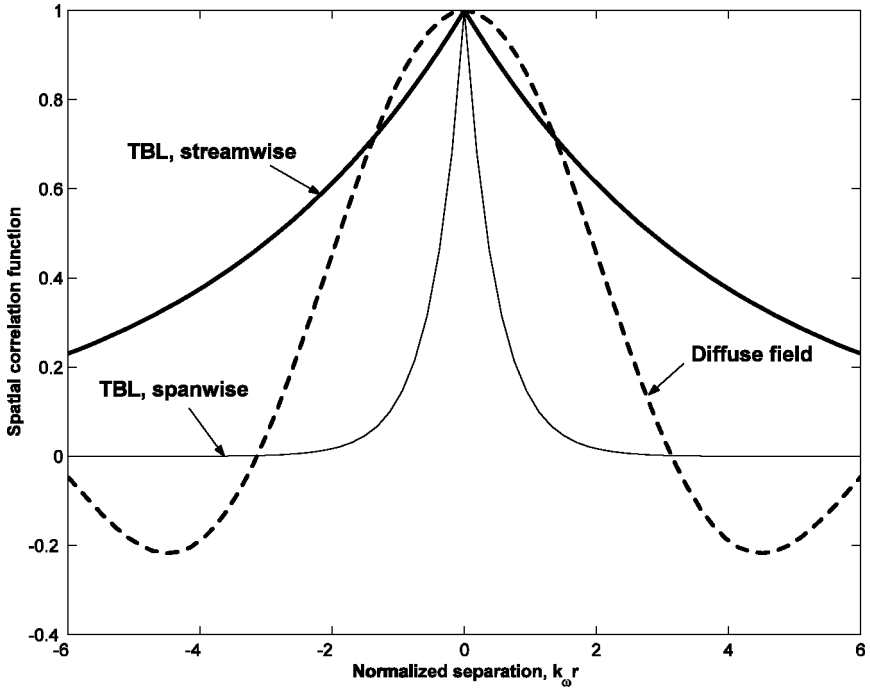


FIG. 1. Normalized magnitude of the spatial correlation function of an acoustic diffuse field (dashed line) and of the Corcos model of a TBL pressure field in the spanwise (thin solid line), and streamwise (thick solid line) directions, plotted as a function of separation distance normalized by the acoustic wavelength, $\lambda = 2\pi/k_\omega$.

normalized spatial correlation structure of the Corcos TBL model in the streamwise and spanwise direction for the conditions above is compared with that for the acoustic diffuse field in Fig. 1.

Figure 2 shows the calculated value of the total kinetic energy of a 0.7 m×0.5 m×0.003 m simply supported aluminum panel, calculated from the sum of the squared mode amplitudes, when subject to either the diffuse or TBL models of the random excitation fields.[19,20] The streamwise direction of the TBL was assumed to be parallel to the longer dimension of the panel. The power spectral density at any one point in the pressure fields, $S_{pp}(\omega)$ in Eqs. (1) and (2), was assumed to be independent of frequency, and equal in the two cases. It can be seen that the TBL pressure field is considerably less efficient at exciting the vibration of the panel, with the average level of the kinetic energy being about 15 dB

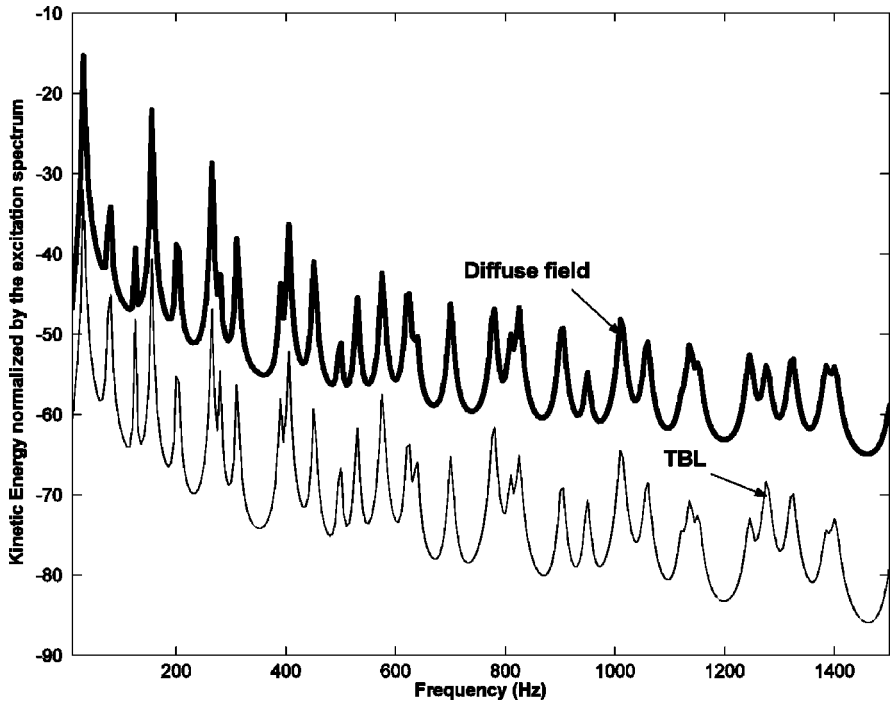


FIG. 2. Spectral density of a panel's kinetic energy when subject to an acoustic diffuse field (thick line), and a Corcos model of the TBL pressure field at $M \approx 0.66$ (thin line) having the same mean-square pressure.

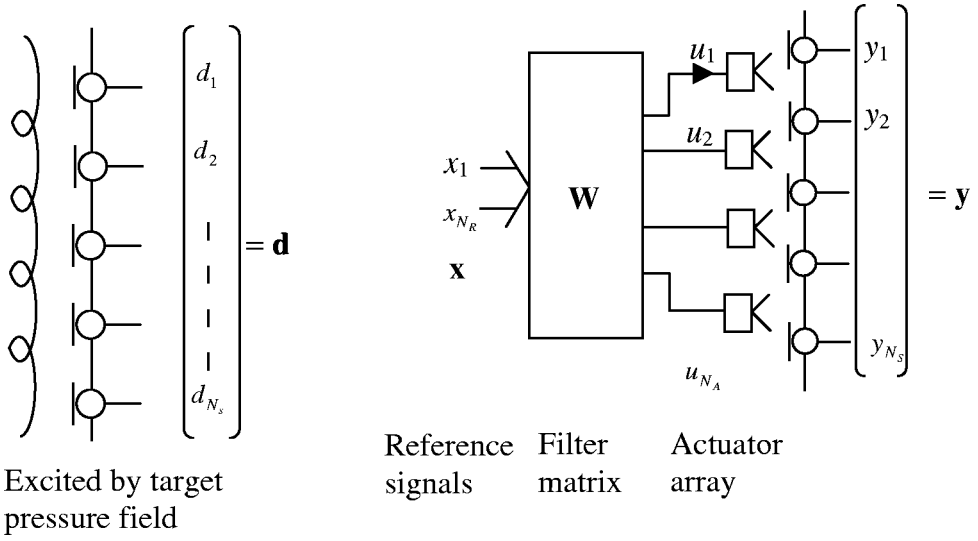


FIG. 3. The outputs from an array of microphones on a panel excited by either a target pressure field (left-hand picture), which define the desired signals, **d**, for a simulation of the pressure field using an array of acoustic sources (right-hand picture), for which the outputs of the microphone array is **y**.

below that of the diffuse field at low frequencies, and about 20 dB below the diffuse field level at 1500 Hz. These differences in level can partly be explained in terms of the relative widths of the spatial correlation functions shown in Fig. 1, particularly the fact that the width of the spatial correlation function for the TBL in the spanwise direction is about one-tenth that of the diffuse field.[19,20]

III. DENSITY OF THE UNCORRELATED COMPONENTS

Figure 3 illustrates a one-dimensional array of N_S acoustic sensors, which in this case are microphones, mounted on a rigid surface and subject either to the pressure field generated by the target random pressure field, left-hand picture, or an array of N_A acoustic actuators, which are represented as loudspeakers in the right-hand picture. The actuators are assumed to be driven by the output of a matrix of filters, **W**, which are themselves excited by an array of N_R reference signals, **x**. Figure 4 shows the equivalent block diagram for this system, in which **G** is the matrix of responses from each actuator to each sensor in Fig. 3, and the original set of desired sensor signals **d** is assumed to be generated from the set of reference signals in a way that will be explained in Sec. III.

The statistical properties of the target pressure field at the sensor positions can be characterized by the elements of the spectral density matrix,

$$\mathbf{S}_{dd} = E[\mathbf{d}\mathbf{d}^{\mathrm{H}}], \tag{5}$$

where **d** is the vector of Fourier transforms, at a single frequency, of one record of the outputs from the microphone array, subject to excitation by the desired random pressure field, E denotes the expectation operator over the ensemble of records,[21] and $^{\mathrm{H}}$ denotes the Hermitian, complex conjugate, transpose. The diagonal elements of $\mathbf{S}_{dd}$ correspond to the power spectral densities of the outputs from each individual sensor and the off-diagonal elements to the cross-spectral densities between the outputs of pairs of sensors. The frequency dependence of $\mathbf{S}_{dd}$ and the other variables used below, has been suppressed for notational convenience.

The use of a finite number of discrete measurements allows the problem of approximating the continuous pressure field to be treated using linear algebraic methods, but does assume that the measurements are made on a sufficiently dense grid to accurately sample the correlation properties of the random pressure field under consideration. The number of sensors required to accurately represent diffuse and TBL pressure fields has been discussed by Maury *et al.*[22] It should

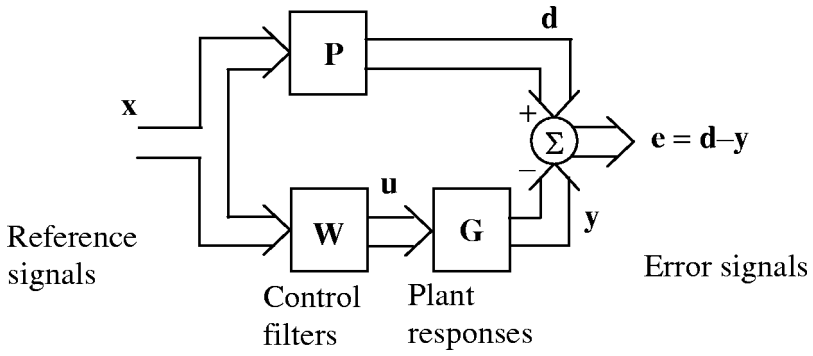


FIG. 4. Block diagram for the calculation of the least-squares control filters designed to reproduce the desired pressures, **d**.

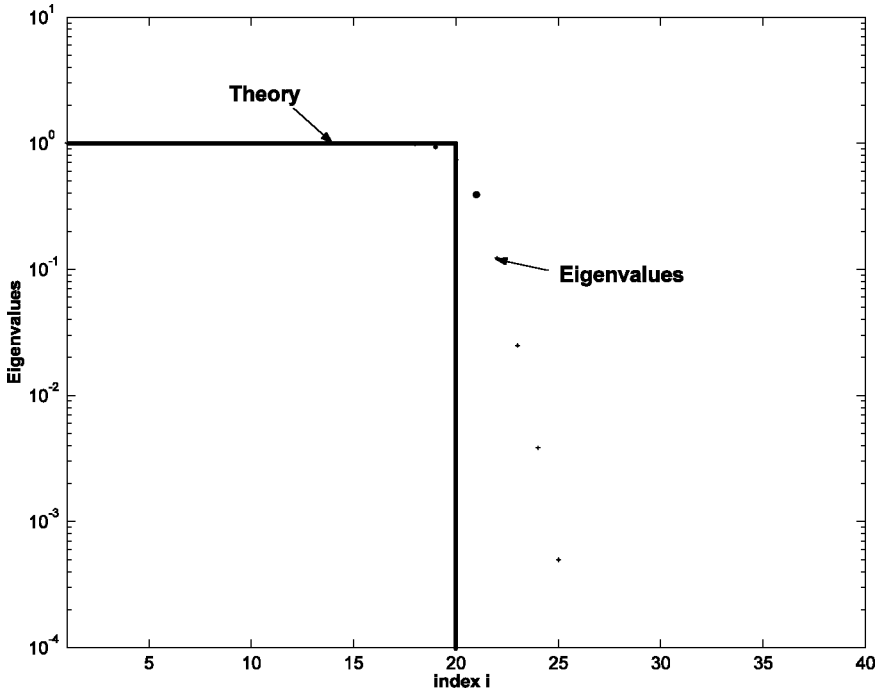


FIG. 5. The eigenvalues of the spectral density matrix calculated for a linear array of 100 microphones exposed to a diffuse acoustic soundfield (dots) together with the theoretical form of this function for an infinite array (solid).

be noted, however, that since, in the formulation below, the random pressure field to be reproduced is defined entirely by its spectral density matrix at the sensor positions, the field is not restricted to being homogeneous in the most general case.

In this section a reduced-rank model for the spectral density matrix for the target pressure field, $\mathbf{S}_{dd}$, will be considered. If $\mathbf{S}_{dd}$, which is an $N_S \times N_S$ matrix, can be approximated to acceptable accuracy with a model of rank N_U, then this must be the number of uncorrelated components underlying the pressure field. Moreover, if the number of reference signals is made equal to this number of uncorrelated components, then the rank of the spectral density matrix for the signals driving the actuator array, $\mathbf{S}_{uu}$, can be no larger than N_U, and so N_U also defines the lower bound on the number of actuators required to reproduce the pressure field to the specified accuracy. In practice it may not be possible to realize such a set of actuators, and when practical actuators are used, a larger number may be required. This is further explored in Sec. IV.

Consider the eigenvalue/eigenvector decomposition, at each frequency, of the Hermitian spectral density matrix for the N_S outputs of the microphone array, excited by the pressure field to be reproduced,

$$\mathbf{S}_{dd} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{\mathrm{H}}, \tag{6}$$

which can also be written as

$$\mathbf{S}_{dd} = \sum_{i=1}^{N_S} \lambda_i \mathbf{q}_i \mathbf{q}_i^{\mathrm{H}}, \tag{7}$$

where $\mathbf{q}_i$ are the columns of the eigenvector matrix $\mathbf{Q}$, and λ_i are the real, positive diagonal elements of the eigenvector matrix $\mathbf{\Lambda}$, ordered such that $\lambda_i > \lambda_{i+1}$. This expansion of the spectral density matrix for the discrete set of pressure measurements is analogous to the proper orthogonal decomposition described, for example, by Berkooz *et al.*[23]

The best least-squares approximation to $\mathbf{S}_{dd}$ with a rank of N_U, $\mathbf{S}_{\hat{d}\hat{d}}$, is now obtained by taking the N_U terms with the largest eigenvalues in Eq. (7) and so has the spectral density matrix

$$\mathbf{S}_{\hat{d}\hat{d}} = \sum_{i=1}^{N_U} \lambda_i \mathbf{q}_i \mathbf{q}_i^{\mathrm{H}}. \tag{8}$$

Defining a vector of error signals to be

$$\mathbf{e} = \mathbf{d} - \hat{\mathbf{d}}, \tag{9}$$

where $\hat{\mathbf{d}}$ is the vector of approximated pressure signals, the sum of the mean square errors is equal to

$$J_e = \mathrm{trace}(\mathbf{S}_{ee}). \tag{10}$$

The ratio of this sum of mean square errors to the sum of the mean square values of $\mathbf{d}$, $J_d = \mathrm{trace}(\mathbf{S}_{dd})$, is then equal to

$$\frac{J_e(N_u)}{J_d} = \frac{\sum_{i=N_U+1}^{N_S} \lambda_i}{\sum_{i=1}^{N_S} \lambda_i}. \tag{11}$$

This expression allows the accuracy of the reduced rank approximations to be quantified. Figure 5 shows the magnitude of the eigenvalues of the $\mathbf{S}_{dd}$ matrix calculated from a simulation of a diffuse field measured with 100 microphones in a linear array that extends over ten wavelengths.

If the microphones used to characterize the pressure field in Fig. 3 are uniformly separated, and the field is homogeneous, then the matrix $\mathbf{S}_{dd}$ has a Toeplitz form, and as the number of microphones becomes very large, such a ma-

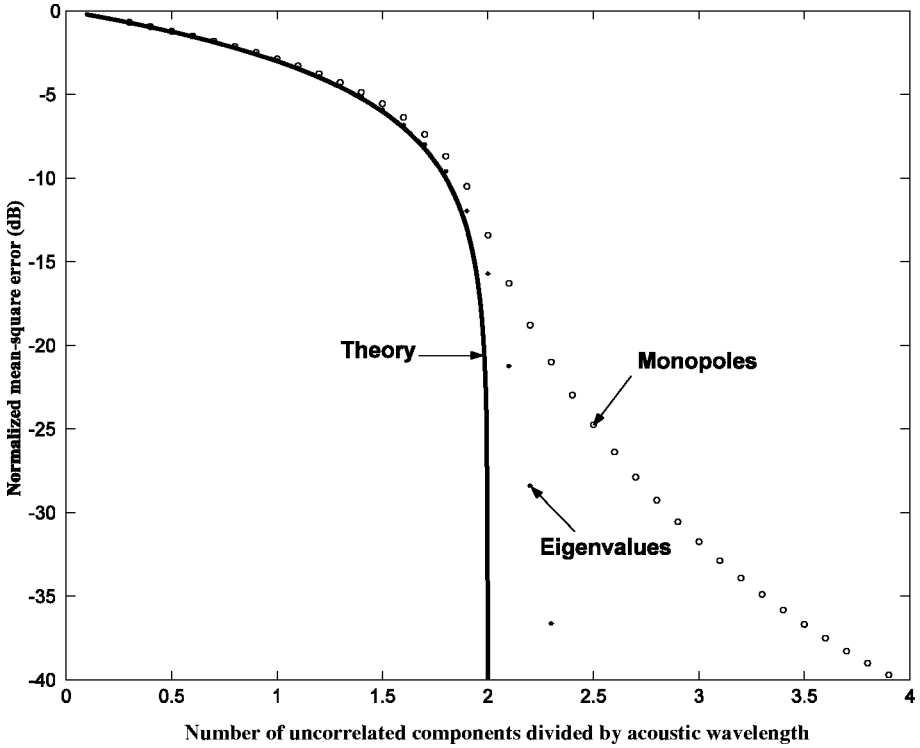


FIG. 6. The normalized mean square error associated with a reduced-rank approximation to the diffuse field spectral density matrix, calculated for the 100 microphone array, as a function of the number of uncorrelated components per wavelength (dots), and the theoretical form of this error for an infinite array (solid line). Also shown is the error when synthesizing the pressure field with an equal number of acoustic monopoles (circles), as described in Sec. IV.

trix becomes asymptotically equivalent to a circulant matrix.[24] Since the eigenvectors of a circulant matrix are Fourier components, the eigendecomposition of $\mathbf{S}_{dd}$ for a homogeneous pressure field will tend to a spatial Fourier transform. The Fourier transform of the spatial correlation function for an acoustic diffuse field, Eq. (1), is proportional to

$$\begin{aligned} F(k) &= 1, \quad |k| \leqslant k_\omega \\ &= 0, \quad |k| > k_\omega, \end{aligned} \tag{12}$$

where k is the spatial wave number and k_ω is ω/c_0, as above. This wave number spectrum is also plotted in Fig. 5, where $k/k_\omega = i/2N_\lambda$, i is the index for the eigenvalues, and N_λ is the number of wavelengths in the analysis window, which is 10 in this case. It can be seen that the theoretical prediction of the eigenvalue distribution, calculated from the wave number spectrum of an infinite array, is reasonably good, and predicts the fall-off in value after the 20th eigenvalue.

Figure 6 shows, in dB, the ratio of the mean square error involved in taking a reduced-rank approximation to the $\mathbf{S}_{dd}$ matrix for the simulation of the diffuse acoustic field to the mean square value of the original field, Eq. (11), as a function of the number of uncorrelated components being considered, divided by the number of wavelengths over the array. It is clear that a 10 dB reduction in the mean square error can be obtained with about two uncorrelated components per wavelength. The solid line in Fig. 6 is the theoretical prediction of the error for an infinite array, obtained by integrating the spatial Fourier transform above over a finite range of k, to give J_e and over all k, to give J_d, so that

$$\begin{aligned} \frac{J_e(N_U)}{J_d} &= 1 - \frac{N_U}{2N_\lambda}, \quad N_U \leqslant 2N_\lambda, \\ &= 0, \quad N_U > 2N_\lambda, \end{aligned} \tag{13}$$

where N_U is the number of uncorrelated components and N_λ is the number of wavelengths in the analysis window. This slightly underpredicts the actual error, because of the finite size of the array, but is clearly a good guide to the number of uncorrelated components required. The modeling error when the diffuse pressure field is simulated with an array of loudspeakers is also shown in Fig. 6, as will be discussed in Sec. IV.

Similar simulations for a one-dimensional microphone array of 100 microphones over 10 correlation lengths of a TBL pressure field in the spanwise direction give the eigenvalue distribution shown in Fig. 7. Also plotted is the theoretical prediction for these eigenvalues, obtained from the spatial Fourier transform of the infinite spatial correlation function in the x direction, from Eq. (2), which is proportional to

$$F(k_x) = \frac{1}{1 + (k_x L_x)^2}, \tag{14}$$

where L_x is the spanwise correlation length and in this case $k_x L_x = 2\pi i/10$. Maury *et al.*[22] have also derived the wave number spectrum for a finite-size array and show that the array must cover at least six correlation lengths for the result to approach that of an infinite array. Figure 8 shows the normalized mean square error in taking a reduced-rank approximation to $\mathbf{S}_{dd}$ for the spanwise TBL pressure field, plot-

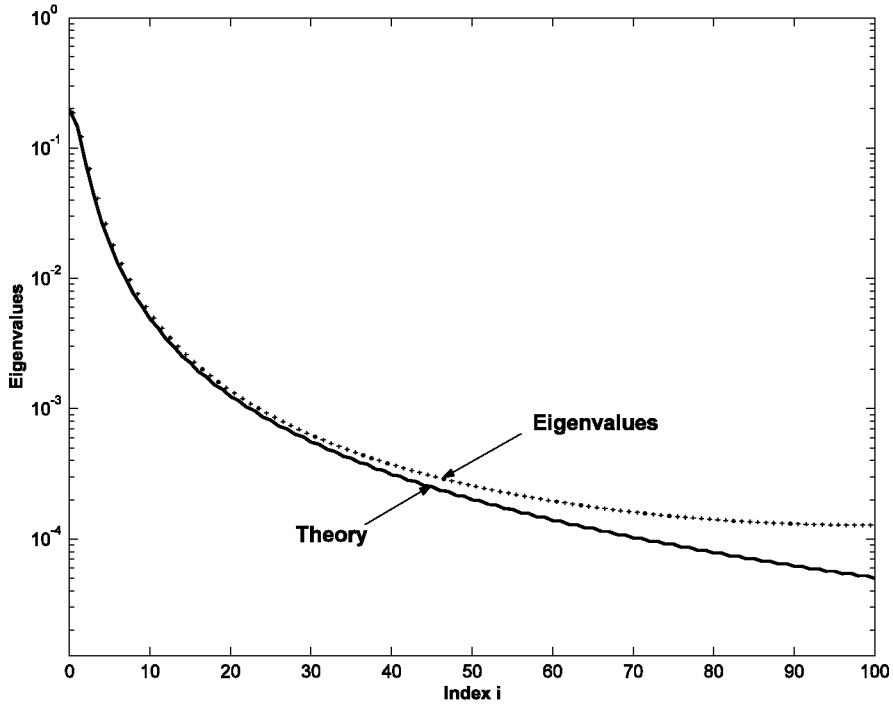


FIG. 7. The eigenvalues of the spectral density matrix calculated for a linear array of 100 microphones exposed to a TBL pressure field in the spanwise direction (dots) together with the theoretical form of this function for an infinite array (solid).

ted against the number of uncorrelated components per correlation length. Also plotted is the theoretical prediction of the normalized error for an infinite array, obtained by integrating $F(k_x)$ over a finite range of k_x, to give

$$\frac{J_e(N_U)}{J_d} = 1 - \frac{2}{\pi}\tan^{-1}\left[\frac{2\pi N_U}{N_c}\right], \tag{15}$$

where N_U is the number of uncorrelated components and N_c

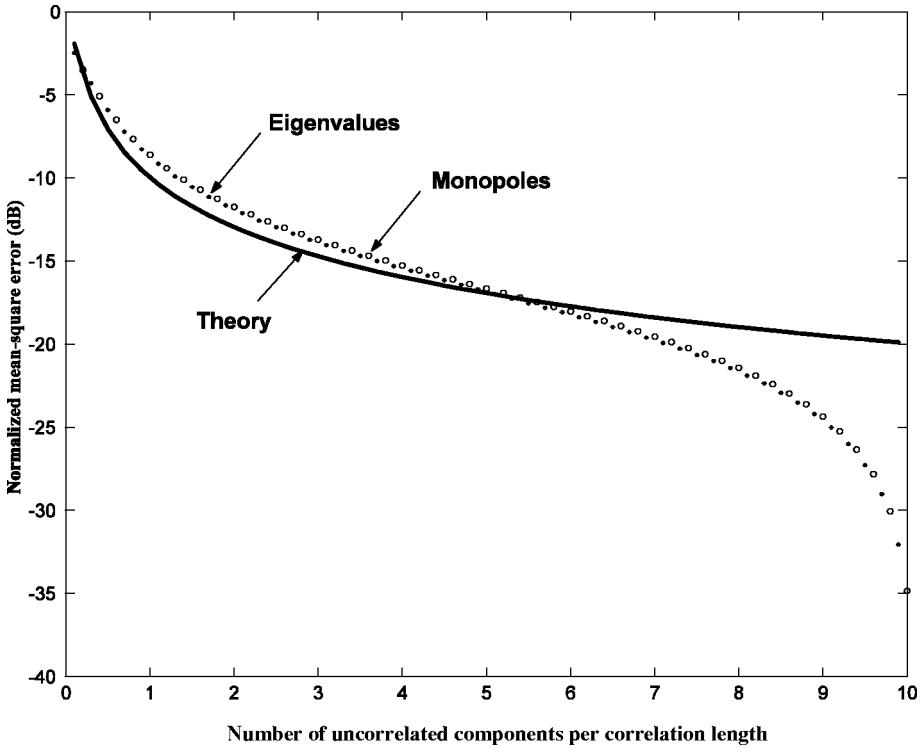


FIG. 8. The normalized mean square error associated with a reduced-rank approximation to the spectral density matrix for a TBL pressure field in the spanwise direction, calculated for the 100 microphone array, as a function of the number of uncorrelated components per correlation length (dots), and the theoretical form of this error for an infinite array (solid line). Also shown is the error when synthesising the pressure field with an equal number of acoustic monopoles (circles), as described in Sec. IV.

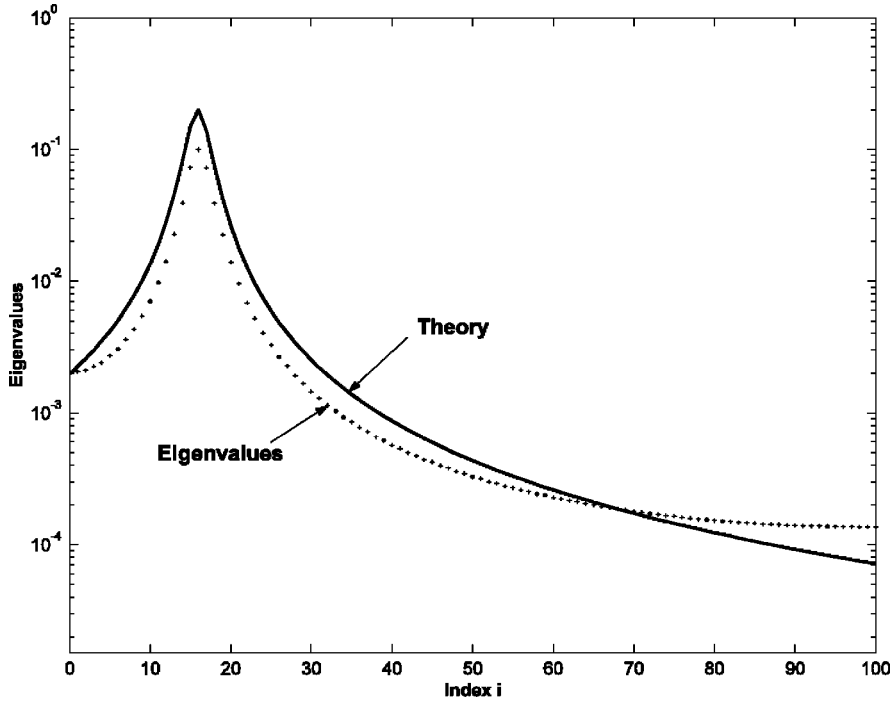


FIG. 9. The eigenvalues of the circulant equivalent of the spectral density matrix calculated for a linear array of 100 microphones exposed to a TBL pressure field in the streamwise direction (dots) together with the theoretical form of this function for an infinite array (solid).

is the number of correlation lengths within the window. This theory predicts that about one uncorrelated component per correlation length is required to obtain a 10 dB reduction in the mean square error, which is close to the simulation results.

Finally, the solid line in Fig. 9 shows the theoretical prediction for the eigenvalue spectrum obtained by taking the Fourier transform of the spatial correlation of the TBL pressure field in the streamwise direction, in Eq. (2), which is proportional to

$$F(k_y) = \frac{1}{1+(k_y L_y - \alpha_y)^2}, \tag{16}$$

where L_y is the correlation length and α_y is the constant in Eq. (4), and in this case $k_y L_y = 2\pi i/10$. The eigenvalues from the simulation cannot be directly compared with this theory because when they are calculated numerically, they are generally ordered from largest to smallest. Instead the eigenvalues of the asymptotically equivalent circulant matrix have been calculated, and are plotted in Fig. 9, which are again in reasonable agreement with the theoretical prediction. By integrating $F(k_y)$ over a finite range of k_y, the theoretical form for the normalized mean square error involved in approximating $\mathbf{S}_{dd}$ for a large array of sensors in the streamwise direction is obtained as

$$\frac{J_e(N_U)}{J_d} = \frac{1 - \frac{2}{\pi}\tan^{-1}\left[\frac{2\pi N_U}{N_c} - \alpha_y\right]}{1 + \frac{2}{\pi}\tan^{-1}[\alpha_y]}. \tag{17}$$

This is plotted in Fig. 10, together with the results of the numerical simulation over ten correlation lengths. It can be seen that the theory predicts that about two uncorrelated components per correlation length are required to obtain a 10 dB reduction in the mean square error, and this is in reasonable agreement with the simulations. Thus, about twice as many uncorrelated components per correlation length are required to reproduce the TBL pressure field in the streamwise direction compared with the spanwise direction. This is because of the presence of the convective peak in the wave number spectrum, as seen in Fig. 9.

IV. SIMULATION OF THE PRESSURE FIELD WITH ACOUSTIC SOURCES

Having presented a method for determining the number of uncorrelated components necessary to represent a random field at a number of sensors, in this section we will examine how such a random field can be reproduced using a finite number of acoustic sources. The number of uncorrelated components sets a lower bound on the number of acoustic sources required.

The vector of N_S output spectra from the sensor array, when excited by the set of acoustic sources shown on the right-hand part of Fig. 3, can be written as

$$\mathbf{y} = \mathbf{G}\mathbf{W}\mathbf{x}, \tag{18}$$

where $\mathbf{x}$ is the vector of N_R reference signal spectra used to generate the inputs to the N_A actuators via the matrix of filter frequency responses denoted $\mathbf{W}$, and $\mathbf{G}$ represents the matrix of physical frequency responses from the N_A sources to the N_S sensors, including the acoustic propagation effects. It will be assumed that the reference signals are white, have unit variance, and are mutually uncorrelated, in which case their spectral density matrix is equal to the identity matrix

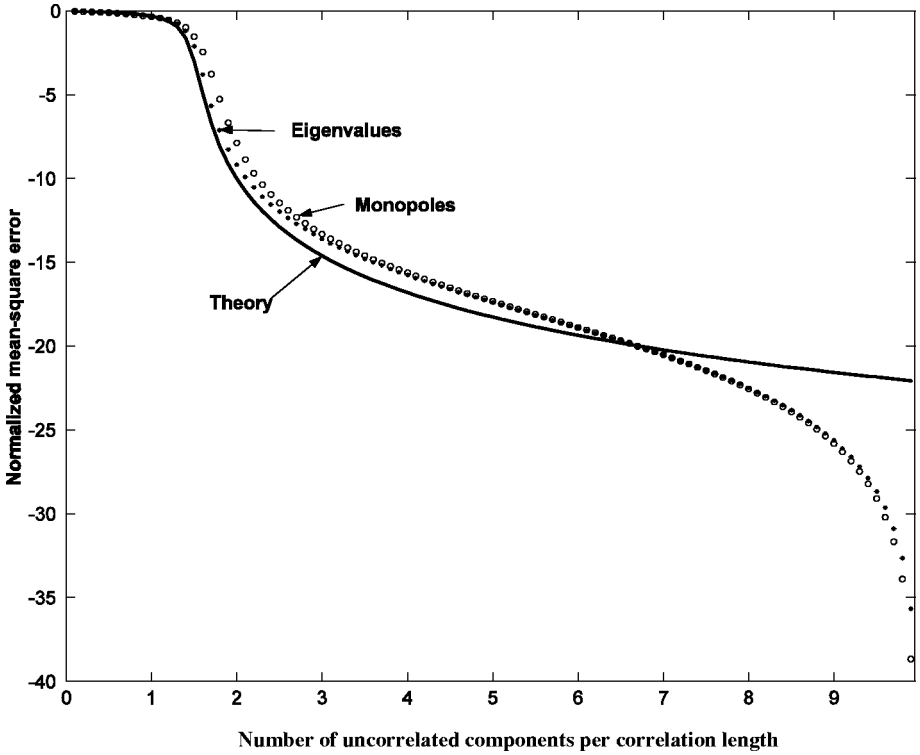


FIG. 10. The normalized mean square error associated with a reduced-rank approximation to the spectral density matrix for a TBL pressure field in the streamwise direction, calculated for the 100 microphone array, as a function of the number of uncorrelated components per correlation length (dots), and the theoretical form of this error for an infinite array (solid line). Also shown is the error when synthesizing the pressure field with an equal number of acoustic monopoles (circles), as described in Sec. III.

$$\mathbf{S}_{xx} = E[\mathbf{x}\mathbf{x}^{\mathrm{H}}] = \mathbf{I}. \tag{19}$$

The spectral density matrix of the microphone outputs when subject to the sensor array can thus be written as

$$\mathbf{S}_{yy} = \mathbf{G}\mathbf{W}\mathbf{W}^{\mathrm{H}}\mathbf{G}^{\mathrm{H}}. \tag{20}$$

Figure 4 shows the block diagram, used to design the responses of the filter matrix, **W**. It has been assumed that the set of desired signals at the sensors, **d**, has been generated by passing the same set of independent white reference signals discussed above, **x**, through the filter matrix **P**. Clearly the waveforms of the random pressures in the target pressure field are unknown, but if the spectral density matrix of the sensor outputs, $\mathbf{S}_{dd}$, can be made equal to that of the target pressure field, then the pressure fields are statistically equivalent.

In order to derive the properties of the filter matrix **P** in Fig. 4, we again consider the eigenvalue/eigenvector decomposition of the spectral density matrix for the measured pressures when exposed to the desired pressure field;

$$\mathbf{S}_{dd} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{\mathrm{H}}, \tag{21}$$

where **Q** is the matrix of eigenvectors and $\mathbf{\Lambda}$ the diagonal matrix of real, positive eigenvalues. If the vector of target pressures is generated from **x**, as in Fig. 4, then

$$\mathbf{d} = \mathbf{P}\mathbf{x}, \tag{22}$$

so that

$$\mathbf{S}_{dd} = \mathbf{P}\mathbf{P}^{\mathrm{H}}, \tag{23}$$

where Eq. (19) for $\mathbf{S}_{xx}$ has again been used.

Since $\mathbf{\Lambda}^{1/2} = (\mathbf{\Lambda}^{1/2})^{\mathrm{H}}$, then Eq. (21) can be written as

$$\mathbf{S}_{dd} = (\mathbf{Q}\mathbf{\Lambda}^{1/2})(\mathbf{Q}\mathbf{\Lambda}^{1/2})^{\mathrm{H}}, \tag{24}$$

and the statistical properties of the signals, **d**, generated in Fig. 4 are identical to those of the target pressure field if the number of reference signals is equal to the number of sensors, $N_R = N_S$, so that **P** is a square matrix, and **P** is given by

$$\mathbf{P} = \mathbf{Q}\mathbf{\Lambda}^{1/2}. \tag{25}$$

In principle, the number of reference signals could be reduced to the number of uncorrelated components, in which case **P** could be calculated from the eigenvalue/eigenvector decomposition of $\mathbf{S}_{\hat{d}\hat{d}}$ in Eq. (8), but we are interested below in determining the minimum number of sources independently of the minimum number of correlated components, and so we retain the full number of reference signals. The matrix of filter responses, **W**, is now calculated which minimizes the mean square difference between these desired signals and the sensor outputs due to the acoustic sources in Fig. 3. The vector of error signals, defined as the difference between the desired output signals, **d**, and the generated output signals, **y**, in Fig. 4, can be written as

$$\mathbf{e} = (\mathbf{P} - \mathbf{G}\mathbf{W})\mathbf{x}. \tag{26}$$

If these error signals can be made small, then the pressure signals generated by the actuator array must be similar to those of the target pressure field. The magnitude of these error signals can be quantified using a norm of their spectral density matrix, which, since $\mathbf{S}_{xx} = \mathbf{I}$, can be written as

$$\mathbf{S}_{ee} = \mathbf{G}\mathbf{W}\mathbf{W}^{\mathrm{H}}\mathbf{G}^{\mathrm{H}} - \mathbf{G}\mathbf{W}\mathbf{P}^{\mathrm{H}} - \mathbf{P}\mathbf{W}^{\mathrm{H}}\mathbf{G}^{\mathrm{H}} + \mathbf{P}\mathbf{P}^{\mathrm{H}}. \tag{27}$$

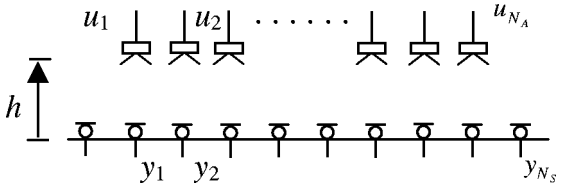


FIG. 11. A one-dimensional array of loudspeakers, modeled as acoustic monopoles, at a height h above a rigid surface on which is mounted an array of microphones, whose outputs are arranged to be as close as possible to the target random pressure field.

The matrix of control filters $\mathbf{W}$ is now calculated that minimizes the quadratic cost function, J_e, equal to the sum of the mean square value of the errors, which is equal to

$$J_e = \mathrm{trace}(\mathbf{S}_{ee}). \tag{28}$$

It is shown by Elliott[25] that the complex matrix containing the derivatives of J_e with respect to the real and imaginary parts of each element of $\mathbf{W}$, $\mathbf{W}_R$, and $\mathbf{W}_I$, is equal to

$$\frac{\partial J_e}{\partial \mathbf{W}_R} + j\frac{\partial J_e}{\partial \mathbf{W}_I} = 2\mathbf{G}^{\mathrm{H}}\mathbf{G}\mathbf{W} - 2\mathbf{G}^{\mathrm{H}}\mathbf{P}. \tag{29}$$

Assuming that $\mathbf{G}^{\mathrm{H}}\mathbf{G}$ is nonsingular and setting Eq. (29) to zero allows the optimum least-squares matrix of control filters to be calculated as

$$\mathbf{W}_{\mathrm{opt}} = [\mathbf{G}^{\mathrm{H}}\mathbf{G}]^{-1}\mathbf{G}^{\mathrm{H}}\mathbf{P} = \mathbf{G}^{\dagger}\mathbf{P}, \tag{30}$$

where $\mathbf{G}^{\dagger}$ denotes $[\mathbf{G}^{\mathrm{H}}\mathbf{G}]^{-1}\mathbf{G}^{\mathrm{H}}$, which is the pseudoinverse of the matrix $\mathbf{G}$. The frequency response matrix would probably not be causal, but could be made causal for a real-time implementation by including delays in all elements, which will not affect the statistical properties of the reproduced pressure field. In practice, the drive signals to the actuators could be generated off-line and just played out during synthesis, in which case the filter matrix would not need to be made causal. It is important, however, that the same geometric and physical conditions are maintained during reproduction as were present when the frequency responses contained in $\mathbf{G}$ were measured. The resulting minimum value of the norm of $\mathbf{S}_{ee}$, obtained by substituting Eq. (30) into Eq. (27), is

$$J_{e,\min} = \mathrm{trace}[(\mathbf{I} - \mathbf{G}\mathbf{G}^{\dagger})\mathbf{S}_{dd}]. \tag{31}$$

This can be divided by the trace of $\mathbf{S}_{dd}$, referred to above as J_d, to give a normalized value of the minimum mean square error involved in the simulation, which is a useful way of quantifying the degree to which the statistical properties of the target pressure field have been reproduced by the acoustic array.

V. EXAMPLES OF SIMULATION WITH MONOPOLE ACOUSTIC SOURCES

One practical actuator arrangement that could be used to simulate a given random pressure field is an array of loudspeakers, which can be modeled as monopole acoustic sources. The number of monopoles required to achieve a given accuracy of simulation will be investigated by again considering a one-dimensional array, as shown in Fig. 11. The simulation of an acoustic diffuse field is considered first.

In this computer simulation there are again assumed to be 100 microphones evenly spaced over a length scale, which includes 10 acoustic wavelengths, but now N_{A} loudspeakers are evenly spaced at a height h above the surface on which the microphones are positioned. The loudspeaker array is assumed to be driven by the outputs of a matrix of control filters, whose responses are adjusted to minimize the sum of the mean-square differences between the 100 microphone outputs and those due to a corresponding diffuse field excitation, as discussed in Sec. III. A series of simulations has been performed in which the loudspeakers have been modeled as acoustic monopole sources radiating over an infinite rigid surface on which the microphone array has been mounted, and their number, N_{A}, and their height above the microphone array, h, has been varied.

Figure 12 shows the variation of the normalized residual mean-square error as a function of the height of the loudspeaker array, divided by the acoustic wavelength λ, for different numbers of loudspeakers in the array, N_{A}. With $N_{\mathrm{A}} = 16$, the mean-square error falls with height, but little reduction in mean-square error is achieved. When a larger number of loudspeakers is used, $N_{\mathrm{A}} = 24$, which can achieve a more appreciable reduction in mean-square error, these reductions are relatively independent of the height of the loudspeaker array, provided it is greater than about 0.3λ, but have a minimum when $h \approx 0.4\lambda$. The control filters are clearly able to compensate for the matrix of acoustic responses between the loudspeakers and microphones, provided the loudspeakers are not too close to the microphones, in which case some microphone outputs are overwhelmingly dominated by the near-field responses of the adjacent loudspeakers. There are numerical conditioning problems if the loudspeakers are too far from the microphones, however, and a reasonable height, h, appears to be about half the separation distance between the loudspeakers.

Similar graphs are obtained when simulations are performed to simulate the random TBL pressure field along a one-dimensional array in either the spanwise or streamwise directions.[22] Although the number of loudspeakers required to achieve a given reduction in the error changes in these cases, near-optimal reductions are achieved when the distance of the sources from the microphone array is again about one-half the separation distance between the loudspeakers. The magnitude of the mean square error reduction, for this separation distance from the monopoles to the microphones, has been plotted as a function of the number of monopoles in the acoustic actuator arrays for the acoustic diffuse in Fig. 6 and for the TBL pressure fields in the span and streamwise directions in Figs. 8 and 10, respectively.

It is interesting to note that the normalized mean square error involved in simulating any of these one-dimensional random pressure fields with an array of suitably positioned monopole sources is only slightly larger than that when the reduced rank approximation to $\mathbf{S}_{dd}$ is considered with an equal number of uncorrelated components. The number of sources in a one-dimensional array of acoustic monopoles thus appears to approach the minimum possible number of uncorrelated sources required. The quality of the synthesis can also be assessed by plotting the correlation functions

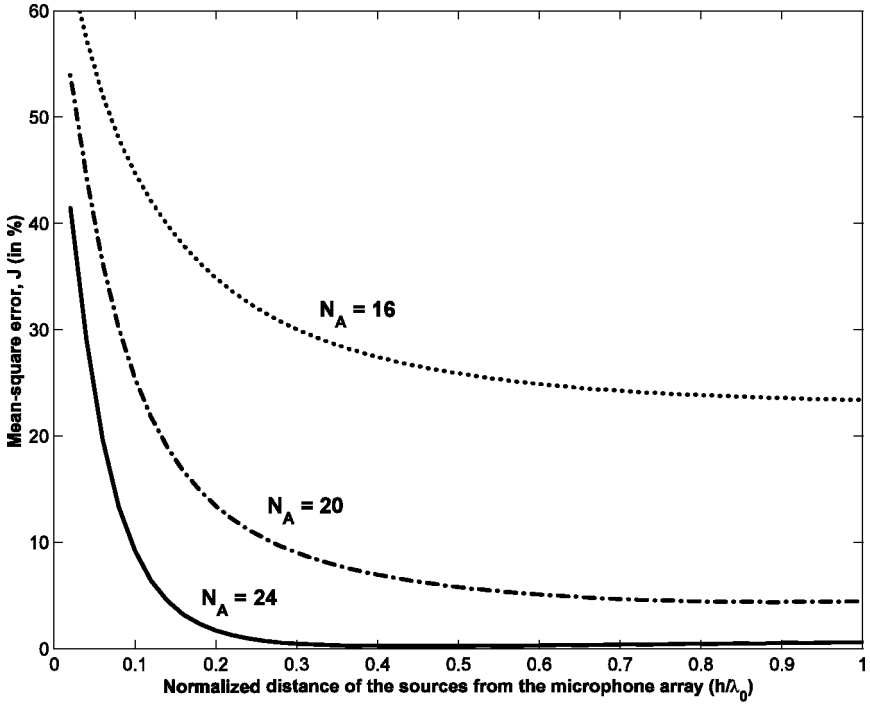


FIG. 12. The normalized mean-square error for simulations in which various numbers of acoustic sources, N_A, were used to reproduce a diffuse pressure field, as a function of the normalized distance of the sources from the microphone array, h/λ_0.

resulting from simulations with various numbers of monopoles per wavelength for the acoustic diffuse field, in Fig. 13, and with various numbers of monopoles per correlation length for the TBL pressure field in the spanwise and streamwise directions, in Figs. 14 and 15, respectively. The spatial correlation structures are reasonably well reproduced when the normalized mean square error is below -10 dB, although the peak value for the reproduced TBL pressure field still does not reproduce the sharp cusp of the Corcos model, which may not be physically realistic in any case, as discussed in Sec. II.

Although the current theory allows clear conclusions to be drawn about the number of sources required to simulate diffuse or TBL pressure fluctuations along a one-dimensional

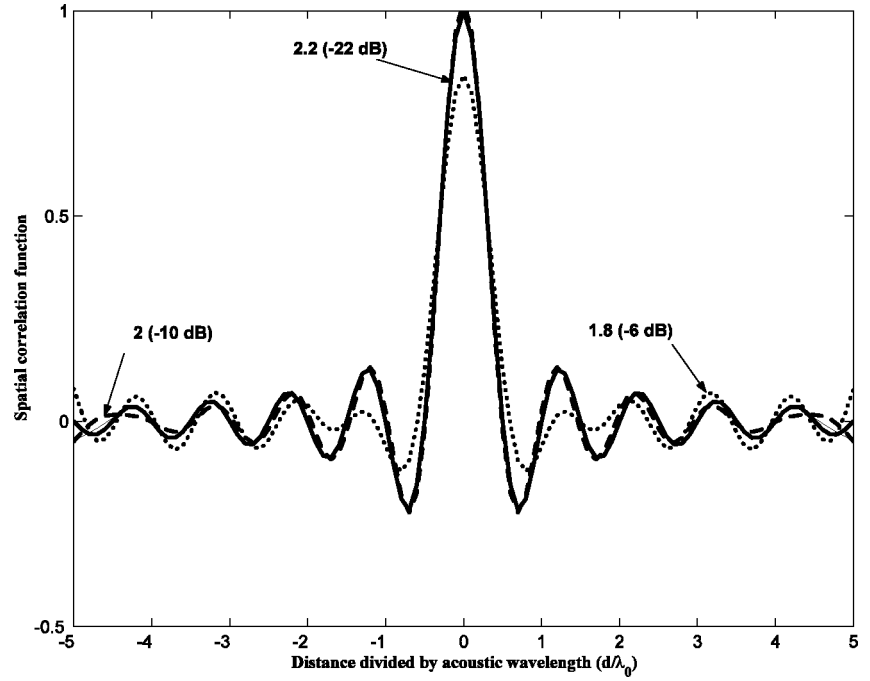


FIG. 13. The target spatial correlation function of the acoustic diffuse field (solid line), and that of the approximate pressure field generated using 1.8 monopoles per acoustic wavelength (dotted line), 2 monopoles per acoustic wavelength (dashed line), and 2.2 monopoles per acoustic wavelength (thin solid line), plotted as a function of normalized separation distance from the center microphone, divided by the acoustic wavelength. The numbers in brackets are the changes in mean square error associated with each number of monopoles.

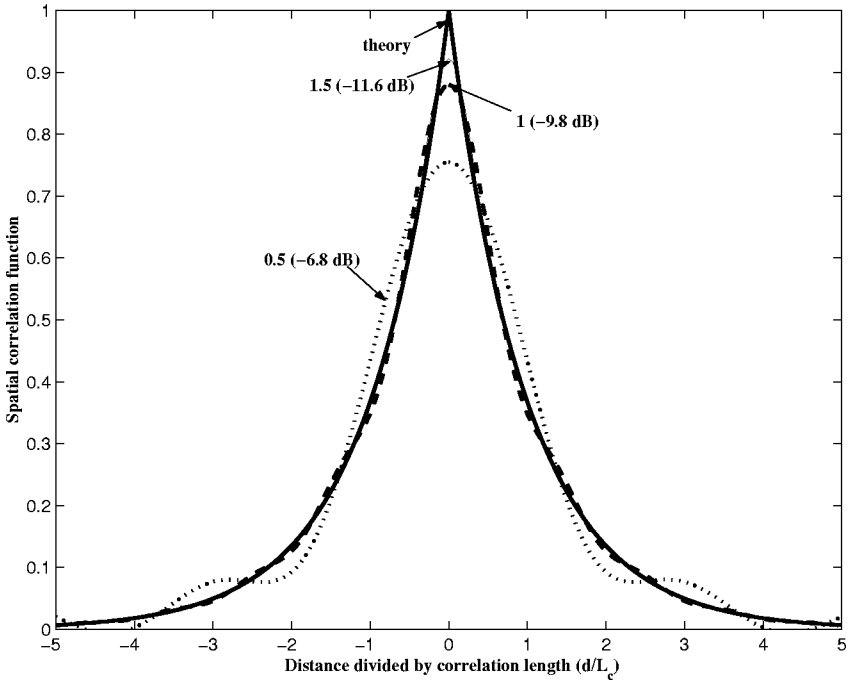


FIG. 14. The target spatial correlation function of the TBL pressure field in the spanwise direction (solid line), and that of the approximate pressure field generated using 0.5 monopoles per correlation length (dotted line), 1 monopole per correlation length (dashed line), and 1.5 monopoles per correlation length (thin solid line), plotted as a function of the normalized separation distance from the center microphone, divided by the correlation length. The numbers in brackets are the changes in mean square error associated with each number of monopoles.

sensor array, the theoretical results do not generalize easily to two-dimensional arrays. It is still possible, however, to use the theory outlined in Sec. III to calculate the best least-squares approximation to a given random pressure field over a two-dimensional array of microphones. Figure 16, for example, shows the results of a simulation in which a 5×5 array of monopoles has been used to generate a least squares approximation to a diffuse or a TBL pressure field over a 100×100 array of microphones on the surface of the 0.7 $m\times0.5$ $m\times0.003$ m panel, as in the simulation used to generate in Fig. 2. The overall kinetic energy of the panel is again plotted as a function of frequency, with that generated by the pressure field synthesized with the monopole array shown dotted, and that due to the theoretical pressure fields, whose spatial correlation functions are given by Eqs. (1) and (2), shown solid.

It can be seen that the panel's response to the diffuse pressure field is reproduced reasonably well by the monopole array up to about 800 Hz. At this frequency the largest spacing between the monopoles is about one-third of the acoustic wavelength, in reasonable agreement with the maximum separation of one-half wavelength predicted by the one-dimensional theory. In current sound transmission suites, the Schröder frequency of the reverberant source room is typically well below 800 Hz. These simulations suggest that a much smaller and more heavily damped source room could be employed in sound transmission suites if a near-field loudspeaker array was used to reproduce the diffuse incident pressure field up to several hundred Hertz, and the naturally diffuse nature of the reverberant sound field in the source room was used to generate the incident pressure field only above this frequency.

The panel's response to the simulated TBL excitation is very similar to that of the ideal theoretical model only up to about 200 Hz, however. At this frequency the separation distance between the monopoles is about four-fifths of a correlation length in the spanwise direction, whereas the one-dimensional theory above would suggest that the maximum separation would be equal to the correlation length. Once again the one-dimensional theory provides a reasonable guide to the number of acoustic sources required, although the small spanwise correlation lengths severely limit the frequency range over which a 5×5 array can adequately reproduce the TBL pressure field.

VI. INITIAL EXPERIMENTAL RESULTS

As a preliminary investigation of the practical problems associated with simulating various random pressure fields with an array of loudspeakers, an enclosure was constructed[26] that housed 16 loudspeakers and 32 microphones, as illustrated in Fig. 17. The top of the enclosure had internal dimensions of approximately 430×430 mm. The 74 mm diameter loudspeakers were mounted in a four by four array so that their cones were about 40 mm below a 1 mm thick aluminum panel, which could be mounted on top of the enclosure, but is not shown in Fig. 17, on whose lower surface the pressure field was to be generated. About 10 mm below the panel, a grid of brass rods was mounted, onto which were attached 32 miniature electret microphones. The frequency response from each loudspeaker to each microphone was measured up to about 1 kHz, with the panel in place, to form the experimental version of the matrix **G** in Eq. (18), both with and without acoustic damping material in

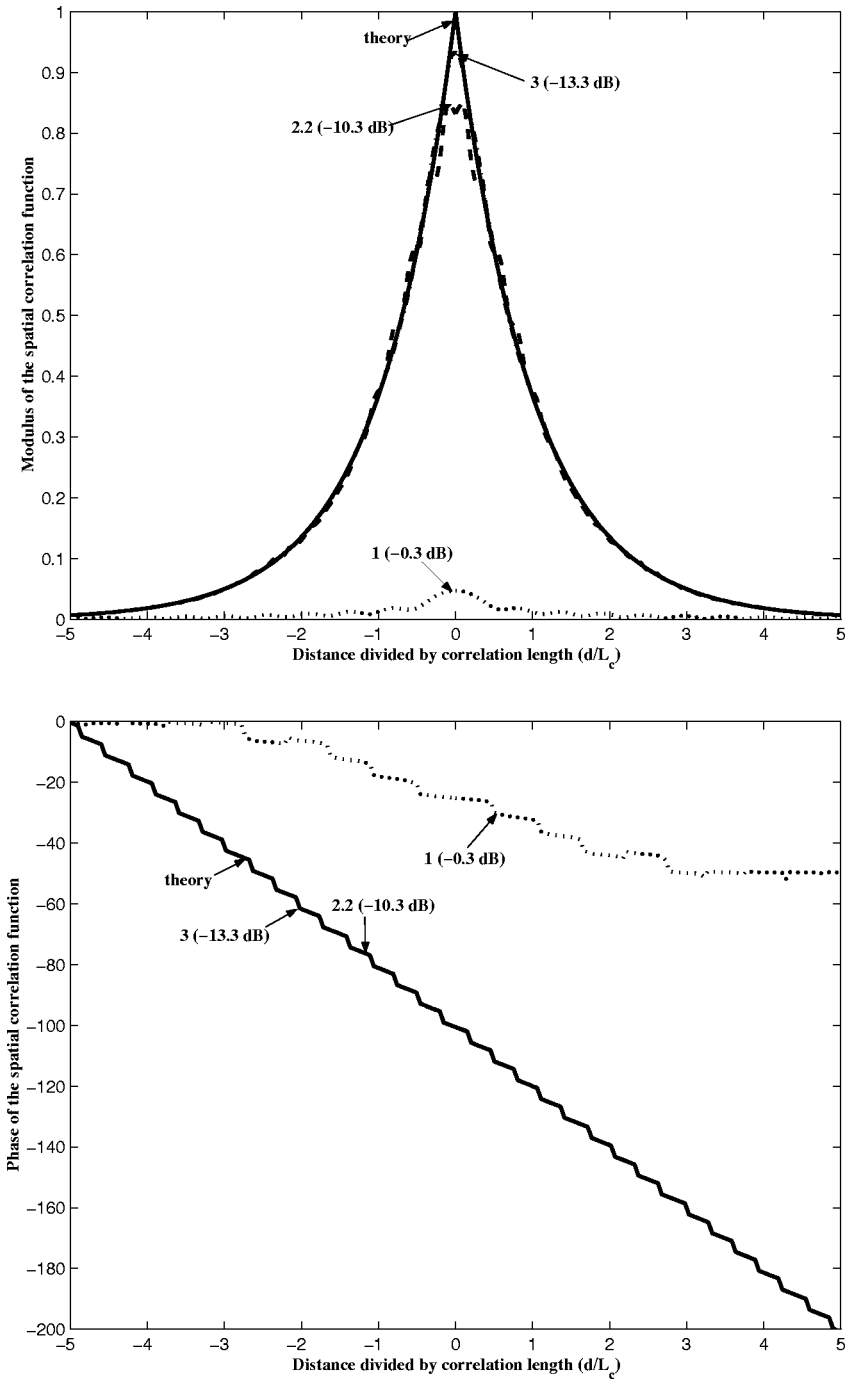


FIG. 15. The modulus and the phase of the target spatial correlation function of the TBL pressure field in the streamwise direction (solid line), and that of the approximate pressure field generated with an array of monopoles when using 1 monopole per correlation length (dotted line), 2.2 monopoles per correlation length (dashed line) and 3 monopoles per correlation length (thin solid line), plotted as a function of the normalized separation distance from the center microphone, divided by correlation length. The numbers in brackets are the changes in mean square error associated with each number of monopoles.

the enclosure and mechanical damping material on the plate. The ratio of the predicted minimum mean square error when simulating various random sound fields at the 32 microphones, as specified by the matrix $\mathbf{S}_{dd}$, was then calculated from Eq. (31) and divided by the trace of $\mathbf{S}_{dd}$ to give the normalized error, as above. For the undamped enclosure the normalized error increased significantly at frequencies corresponding to the coupled natural frequencies of the panel and acoustic enclosure. At these frequencies it is very difficult for the loudspeakers to generate anything but the pressure field corresponding to modal response. When the system is more heavily damped, these resonances are still apparent in the individual responses of **G**, but they are not evident in the graphs of normalized error, indicating that they have been equalized by the optimum controller in Eq. (30). Figure 18 shows the normalized error calculated from the measured **G** matrix and Eq. (31) for the damped enclosure when attempting to simulate a perfect acoustic diffuse field (thick line) and a pressure field corresponding to the TBL, whose properties are described in Sec. II (thin line).

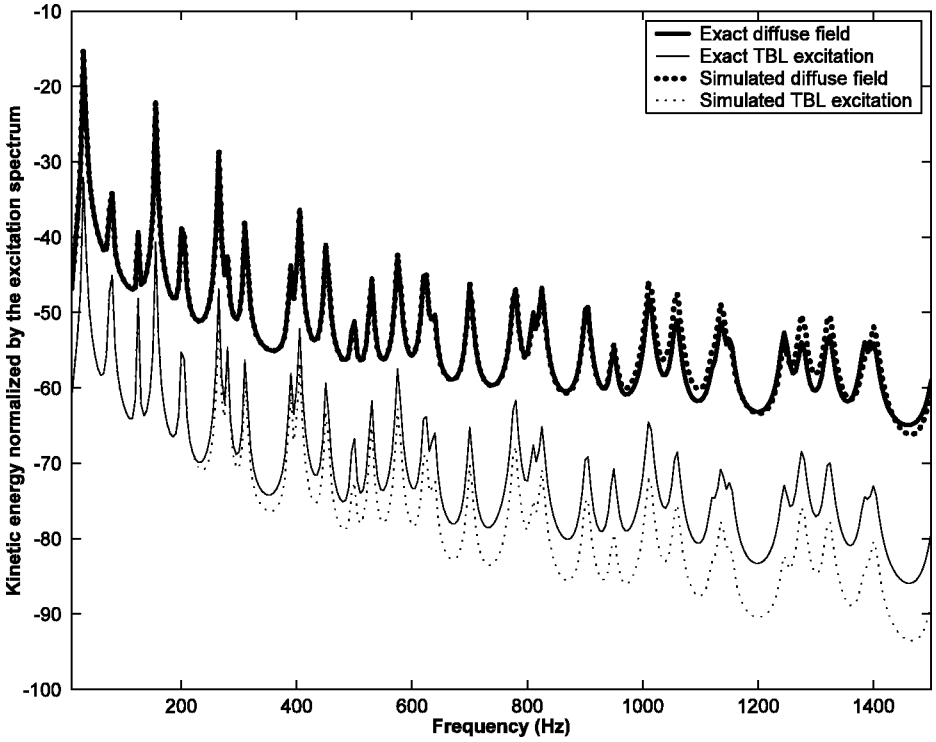


FIG. 16. Spectral density of a panel's kinetic energy, normalized by the excitation spectrum, for an acoustic diffuse field excitation (solid bold curve), for a TBL excitation (solid thin curve), and for a reproduced acoustic diffuse field generated with an array of 5×5 monopoles close to the panel (dotted bold curve) and for a reproduced TBL excitation generated with the same array of monopoles (dotted thin curve).

It is clear from Fig. 18 that in the case of the diffuse field, the simulation error is relatively low up to about 600 Hz, at which frequency the separation between the loudspeakers is about one-third of an acoustic wavelength, whereas in the case of the TBL pressure field, the simulation error is somewhat higher, and rises significantly at about 200 Hz, at which frequency the loudspeaker separation distance is about four-fifths of the correlation length in the spanwise direction. The experimental results are thus in broad agreement with the simulation results presented in Sec. V, provided lightly damped resonances are avoided.

VII. CONCLUSIONS

The framework presented allows the feasibility to be assessed of generating a pressure field with a given spatial

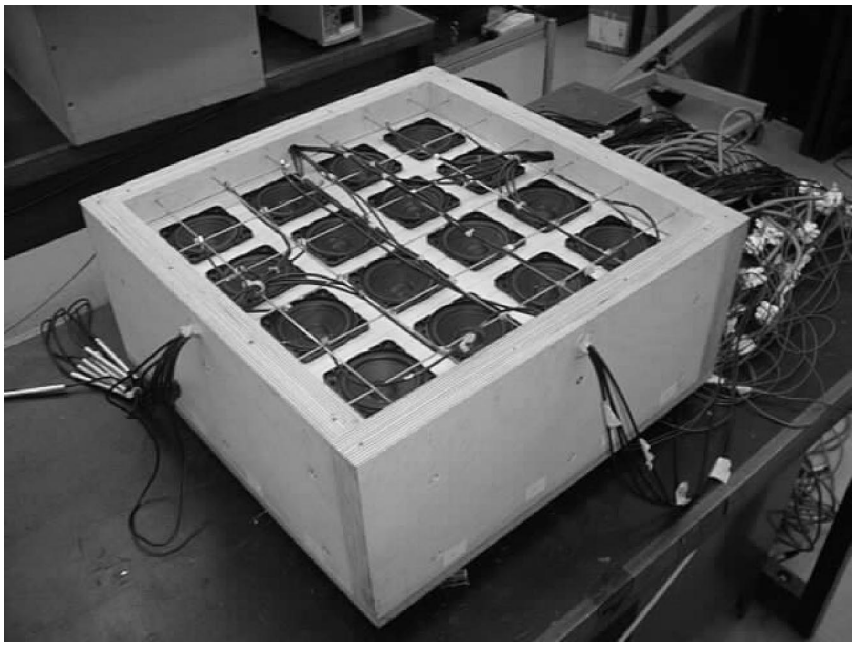


FIG. 17. Photograph of an experimental implementation using 16 loudspeakers and 32 microphones to simulate various random pressure fields on the lower face of a panel, which is normally attached to the top of the enclosure.

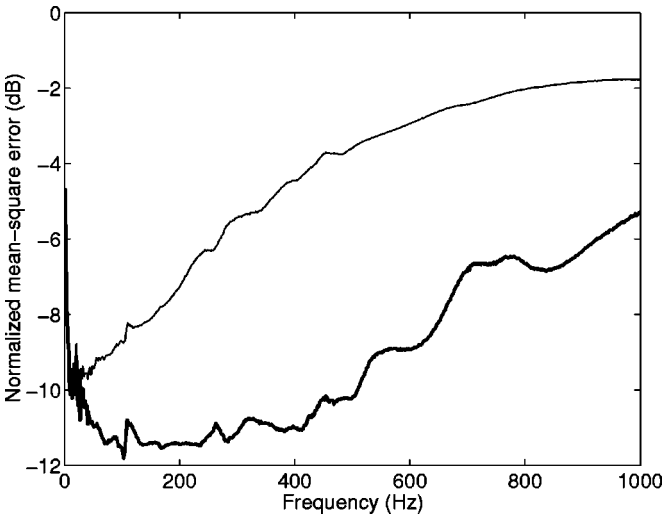


FIG. 18. Normalized error as a function of frequency calculated from the **G** matrix measured in the experimental enclosure when simulating an acoustic diffuse field (dark curve) and the pressure field due to a TBL (light curve).

correlation structure using an array of acoustic actuators. The pressure field is specified in terms of the spectral density matrix of the outputs of an array of sensors, which can be calculated from a theoretical model of the spatial correlation structure of the pressure field to be synthesized, or measured in an experimental application.

The formulation allows a lower bound to be calculated on the number of actuators required to accurately generate a given pressure field, which is equal to the number of uncorrelated components required to approximate the statistical properties of the field to a required degree of accuracy. A one-dimensional simulation suggests that two uncorrelated components per wavelength are required to approximate a diffuse acoustic field. This is in good agreement with the theoretical prediction obtained from a wave number decomposition of the diffuse field, which indicates that exactly two uncorrelated components per wavelength are required for an infinite array. For the Corcos model of the TBL pressure field, this number scales with the correlation length, and for a 10% mean square error in the simulation, about one uncorrelated component per correlation length is required in the spanwise direction and about two per correlation length are required in the streamwise direction. Although a mean-square error criterion has been used here, other error criteria could be used in this formulation, emphasizing different aspects of the spatial correlation structure. The level of error criterion that is acceptable will depend on the use to which the simulation is to be put. If the target pressure field is calculated from a model, the accuracy of this model must also be taken into account when assessing the physical significance of the error criterion.

The framework has been used to assess the feasibility of synthesizing a diffuse sound field with an array of acoustic sources close to a surface. Such an arrangement may be useful to replace the source room in sound transmission tests, or at least to significantly reduce its size. The number of appropriately placed sources required per wavelength to achieve a good reproduction of the spatial correlation function is almost the same as the number of uncorrelated components in the pressure field.

When simulations are performed to reproduce TBL pressure fields with acoustic monopoles, the number of monopoles required to reproduce the pressure field is again approximately equal to the number of uncorrelated components. The simple Corcos model predicts that the correlation length is inversely proportional to frequency, and so the number of sources required to reproduce such a TBL model rises rapidly with frequency. Simple calculations suggest that the number of loudspeakers required to synthesize pressure fields due to a TBL at high subsonic Mach numbers may be technologically feasible up to a few hundred Hertz. A number of practical issues need to be addressed before such a simulation technique can be used with any confidence in practice, but a preliminary experimental study provides results that are in broad agreement with the simulations, provided lightly damped resonances in the driving acoustic field are avoided.

ACKNOWLEDGMENT

Part of the work reported in this paper was supported by the EC under Contract No. GRD 1-1999-1087 ENABLE, ''environmental noise associated with turbulent boundary layer excitation.''

[1] T. J. Schultz, ''Diffusion in reverberation rooms,'' J. Sound Vib. **16**, 17–28 (1971).
[2] F. Jacobsen, ''The diffuse soundfield,'' Report No. 27, Acoustics Laboratory, Technical University of Denmark, 1979.
[3] R. K. Cook, R. V. Waterhouse, R. D. Berendt, S. Edelman, and M. C. Thompson, ''Measurements of correlation coefficients in reverberant soundfields,'' J. Acoust. Soc. Am. **27**, 1072–1077 (1955).
[4] M. R. Schroeder, ''Statistical parameters of the frequency response curves of large rooms,'' Acustica **4**, 594–600 (1954); English translation in J. Audio Eng. Soc. **35**, 299–316 (1987).
[5] A. D. Pierce, *Acoustics: An Introduction to its Physical Properties and Applications* (McGraw-Hill, New York, 1981).
[6] T. Bravo and S. J. Elliott, ''Variability of low frequency sound transmission measurements,'' J. Acoust. Soc. Am. **115**, 2986–2997 (2004).
[7] G. M. Corcos, ''The resolution of pressures in turbulence,'' J. Acoust. Soc. Am. **35**, 192–199 (1963).
[8] A. Preumont, *Random Vibration and Spectral Analysis* (Kluwer Academic, New York, 1990).
[9] F. J. Fahy, ''On simulating the transmission through structures of noise from turbulent boundary layer pressure fluctuations,'' J. Sound Vib. **3**, 57–81 (1966).
[10] G. Robert and J. Sabot, ''Use of random forces to simulate the vibro-acoustic response of a plate excited by a hydrodynamic turbulent boundary layer,'' ASME Winter Meeting, Symposium on Flow-Induced Vibrations, Vol. 5: Turbulence-Induced Noise and Vibration of Rigid and Compliant Surfaces, Conference Code 06196, 1984.
[11] C. J. Dodds, ''The laboratory simulation of vehicle service stress,'' ASME J. Eng. Ind. **96**, 391–398 (1974).
[12] M. Shinozuka, ''Simulation of multivariate and multidimensional random processes,'' J. Acoust. Soc. Am. **49**, 357–367 (1970).
[13] M. P. Mignolet and P. D. Spanos, ''Recursive simulation of stationary multivariate random processes: Parts I and II,'' Trans. ASME, J. Appl. Mech. **54**, 674–687 (1987).
[14] C. Soize and F. Poirion, ''Numerical simulation of homogeneous and inhomogeneous Gaussian stochastic vector fields,'' Rech. Aerosp. **1**, 41–61 (1989).
[15] R. L. Clark and K. D. Frampton, ''Aeroelastic structural acoustic coupling: Implications on the control of turbulent boundary-layer noise transmission,'' J. Acoust. Soc. Am. **102**, 1639–1647 (1997).
[16] B. M. Efimtsov, ''Characteristics of the field of turbulent wall pressure

fluctuations at large Reynolds numbers,'' Sov. Phys. Acoust. **28**, 289–292 (1982).
[17] W. R. Graham, ''A comparison of models for the wave number-frequency spectrum of turbulent boundary layer pressures,'' J. Sound Vib. **206**, 541–565 (1997).
[18] G. K. Batchelor, *The Theory of Homogenous Turbulence* (Cambridge University Press, Cambridge, 1953).
[19] C. Maury, P. Gardonio, and S. J. Elliott, ''A wave number approach for the response of aircraft sidewalls to random pressure fluctuations,'' *Proceedings of the 17th International Congress on Acoustics*, Rome, 2001 (on CD-Rom).
[20] C. Maury, P. Gardonio, and S. J. Elliott, ''A wave number approach to modelling the response of a randomly excited panel, Part I: General theory and Part II: Application to aircraft panels excited by a turbulent boundary layer,'' J. Sound Vib. **252**, 83–113 and 115–139 (2002).
[21] J. S. Bendat and A. G. Piersol, *Random Data*, 2nd ed. (Wiley, New York, 1986).
[22] C. Maury, S. J. Elliott, and P. Gardonio, ''A convergence study for the laboratory simulation of random pressure fields,'' *Proceedings of the Institute of Acoustics Spring Conference*, Salford, UK, 2002 (on CD-Rom).
[23] G. Berkooz, P. Holmes, and J. L. Lumley, ''The proper orthogonal decomposition in the analysis of turbulent flows,'' Annu. Rev. Fluid Mech. **25**, 539–575 (1993).
[24] R. M. Gray, ''On the asymptotic distribution of Toeplitz matrices,'' IEEE Trans. Inf. Theory **IT-18**, 725–730 (1972).
[25] S. J. Elliott, *Signal Processing for Active Control* (Academic, New York, 2001).
[26] C. Heal, ''Active noise simulation with loudspeaker array,'' M. Eng. Project Report, ISVR, University of Southampton, 2001.

Efficiency parameters in time reversal acoustics: Applications to dispersive media and multimode wave propagation

Ismael Núñez[a] and Carlos Negreira
Laboratorio de Acústica Ultrasonora, Instituto de Física, Facultad de Ciencias, Igua 4225. Montevideo 11400, Uruguay

(Received 28 February 2004; revised 2 December 2004; accepted 15 December 2004)

This paper develops a theoretical background for the interpretation of the time reversal (TR) processes applied to the refocalization of acoustic waves in dispersive media with multimode propagation of waves. Two parameters are introduced in order to measure the efficiency of signal recompression in TR processes: spatial and temporal efficiencies. It is demonstrated that the signal recompression enhances when the medium is highly dispersive, when there are several modes involved in the process, and when the angular aperture of the TR mirror is appreciable. These results are applied to Lamb waves recompression in thin plates and focalization efficiencies in longitudinal and transversal directions are deduced. Finally, these theoretical results are compared with the experimental ones obtained for different plate thicknesses.
[DOI: 10.1121/1.1856272]

PACS numbers: 43.60.Tj, 43.20.Mv, 43.20.Bi, 43.35.Cg, 43.35.Zc [RLW] Pages: 1202–1209

I. INTRODUCTION

Fink and collaborators have successfully developed time reversal processes in several branches of acoustics.[1–5] The technological progress in transducer manufacturing and in emission-reception electronic devices offers a very wide field of applications of the acoustical time reversal mirror (TRM). For example, we can improve the refocalization processes in multi-scattering media by using TRM, showing that the multiscattering media behave like a convergent lens after the time reversal process. Inside a multiscattering medium, the TRM seems to have a greater aperture than it actually has. In particular, previous works demonstrate that time reversal (TR) refocalization is very strong, even in presence of multiple scattering.[3,4]

By using a similar effect, important improvements on TR focalization were achieved in guided waves and closed cavities. The multiple reflections on the boundaries of the guide or cavity increase the apparent size of the TRM.[5–7]

There are experimental works about time reversal refocalization and time recompression in Lamb waves.[8,9] Lamb waves are strongly dispersive in thin plates and its wave packet propagates like a superimposing of several modes.[10]

In this work, we quantitatively study the TR process in dispersive media with multimode propagation. We apply our conclusions to Lamb waves propagation.

We use a transfer function for Lamb waves, which we found in a previous work,[11] and we apply it to define a TR transfer function for two-dimensional waves in dispersive media with multimode propagation.

In order to study the refocalization processes, we develop and apply the equations in the vicinity of the focal point. This allows us to define *efficiency parameters* to evaluate the performance of the TR refocalization process. Finally, we compare our theoretical results with the experimental ones obtained for different plate thicknesses.

II. TIME REVERSAL ACOUSTICS IN DISPERSIVE MEDIA

A. Transfer function for time reversal operation

The temporal Fourier transform of a linear acoustical signal coming from a source point $\mathbf{r}=\mathbf{0}$ is, in a given point $\mathbf{r}_n$,

$$S(\omega,|\mathbf{r}_n|)=H(\omega,|\mathbf{r}_n|)E(\omega), \tag{1}$$

where ω is the angular frequency, $H(\omega,|\mathbf{r}_n|)$ is the transfer function for the propagation from $\mathbf{0}$ to $\mathbf{r}_n$, and $E(\omega)$ is the Fourier transform of the signal emitted in the origin.

Let a small transducer (size$<\lambda$) of TR be placed at $\mathbf{r}_n$. This ideal transducer reads and emits acoustical signals like a punctual one. We have an array of transducers and we suppose that each transducer has the same acousto-electrical response, $h(\omega)$ being its transfer function. The electrical signal read by the nth transducer in the frequency domain is

$$S^{(e)}(\omega,|\mathbf{r}_n|)=h(\omega)H(\omega,|\mathbf{r}_n|)E(\omega). \tag{2}$$

In order to perform the TR process, we take the complex conjugate of the Fourier transform of the original signal in the frequency domain (which is delayed in a time τ to guarantee the causality[1]). Therefore, the electrical signal, time-reversed by the transducer at $\mathbf{r}_n$, is

$$\begin{aligned} S^{(e)}_{\mathrm{TR}}(\omega,|\mathbf{r}_n|)&=\exp(i\omega\tau)S^{(e)*}(\omega,|\mathbf{r}_n|)\\ &=\exp(i\omega\tau)h^*(\omega)H^*(\omega,|\mathbf{r}_n|)E^*(\omega), \end{aligned} \tag{3}$$

where the asterisk stands for the complex conjugate.

The transducer's outgoing acoustical signal at $\mathbf{r}_n$ is

$$\begin{aligned} S_{\mathrm{TR}}(\omega,|\mathbf{r}_n|)&=h(\omega)S^{(e)}_{\mathrm{TR}}(\omega,|\mathbf{r}_n|)\\ &=\exp(i\omega\tau)|h(\omega)|^2H^*(\omega,|\mathbf{r}_n|)E^*(\omega). \end{aligned} \tag{4}$$

[a]Electronic mail: ismael@fisica.edu.uy

The returned signal in a generic point $\mathbf{r}$ requires another propagation $H(\omega,|\mathbf{r}-\mathbf{r}_n|)$ from the transducer placed at $\mathbf{r}_n$ to the observation point at $\mathbf{r}$. Applying Eq. (4), we obtain the Fourier transform of the acoustical signal in the point $\mathbf{r}$ as

$$S_{\mathrm{TR}}^{(n)}(\omega,|\mathbf{r}|)=H(\omega,|\mathbf{r}-\mathbf{r}_n|)S_{\mathrm{TR}}(\omega,|\mathbf{r}_n|)$$
$$=\exp(i\omega\tau)|h|^2H(\omega,|\mathbf{r}-\mathbf{r}_n|)$$
$$\times H^*(\omega,|\mathbf{r}_n|)E^*(\omega). \tag{5}$$

Because of the linearity of the involved equations, if we have an array of small TR transducers in several points $\mathbf{r}_n$, then, the total signal at $\mathbf{r}$ is

$$S(\omega,\mathbf{r})=\sum_n S_{\mathrm{TR}}^{(n)}(\omega,|\mathbf{r}|)$$
$$=\exp(i\omega\tau)|h|^2\sum_n H(\omega,|\mathbf{r}-\mathbf{r}_n|)$$
$$\times H^*(\omega,|\mathbf{r}_n|)E^*(\omega). \tag{6}$$

The sum in Eq. (6) depends on the location of the original source point (the origin $\mathbf{0}$), the location of the observation point ($\mathbf{r}$), and the distribution of the TR transducers in the array (the TR mirror). We define the *time reversal transfer function* as

$$H_{\mathrm{TR}}(\omega,\mathbf{r})=\exp(i\omega\tau)|h|^2\sum_n H(\omega,|\mathbf{r}-\mathbf{r}_n|)H^*(\omega,|\mathbf{r}_n|). \tag{7}$$

Using definition (7), we can write Eq. (6) as

$$S(\omega,\mathbf{r})=H_{\mathrm{TR}}(\omega,\mathbf{r})E^*(\omega). \tag{8}$$

B. Maximum amplitude of TR refocalization process

The signal for which Eq. (6) gives its respective Fourier is

$$s(\mathbf{r},t)=\int_{-\infty}^{\infty}S(\omega,\mathbf{r})\exp(-i\omega t)d\omega$$
$$=\int_{-\infty}^{\infty}|S(\omega,\mathbf{r})|\exp[i\varphi_S(\omega,\mathbf{r})-i\omega t]d\omega, \tag{9}$$

$\varphi_S(\omega,\mathbf{r})$ being the phase of S.

If the signal s has a central frequency ω_0, we rewrite Eq. (9) by multiplying and dividing its second term by $\exp[i\varphi_S(\omega_0,\mathbf{r})-i\omega_0 t]$. By doing so, it becomes

$$s(\mathbf{r},t)=a(\mathbf{r},t)\exp[i\varphi_S(\omega_0,\mathbf{r})-i\omega_0 t], \tag{10}$$

$a(\mathbf{r},t)$ being the modulated amplitude, which results in

$$a(\mathbf{r},t)=\int_{-\infty}^{\infty}|S(\omega,\mathbf{r})|\exp[i\varphi_S(\omega,\mathbf{r})-i\varphi_S(\omega_0,\mathbf{r})]$$
$$\times\exp[-i(\omega-\omega_0)]d\omega. \tag{11}$$

Then, we introduce the shifted frequency $\Omega=\omega-\omega_0$ in Eq. (11), writing it as

$$a(\mathbf{r},t)=\int_{-\infty}^{\infty}|S(\omega_0+\Omega,\mathbf{r})|\exp[i\varphi_S(\omega_0+\Omega,\mathbf{r})$$
$$-i\varphi_S(\omega_0,\mathbf{r})]\exp[-i\Omega]d\Omega. \tag{12}$$

If we consider that the signal's bandwidth is not large enough, we can use the development of $\varphi_S(\omega_0+\Omega,\mathbf{r})$ until the second order in Ω. This is

$$\varphi_S(\omega_0+\Omega,\mathbf{r})\approx\varphi_S(\omega_0,\mathbf{r})+\left(\frac{\partial\varphi_S}{\partial\omega}\right)_{\omega_0}\Omega+\frac{1}{2}\left(\frac{\partial^2\varphi_S}{\partial\omega^2}\right)_{\omega_0}\Omega^2. \tag{13}$$

After substituting Eq. (13) in Eq. (12), we can write it as

$$a(\mathbf{r},t)=\int_{-\infty}^{\infty}|S(\omega_0+\Omega,\mathbf{r})|\exp\left\{i\left[\left(\frac{\partial\varphi_S}{\partial\omega}\right)_{\omega_0}-t\right]\Omega\right\}$$
$$\times\exp\left[\frac{i}{2}\left(\frac{\partial^2\varphi_S}{\partial\omega^2}\right)_{\omega_0}\Omega^2\right]d\Omega. \tag{14}$$

Equation (14) gives the maximum amplitude $a_m(\mathbf{r})$ for each point $\mathbf{r}$, in the exact same instant in which the components of the superimposing are in phase (stationary phase method[12,13]). In other words, when the chosen time t satisfies

$$\left(\frac{\partial\varphi_S}{\partial\omega}\right)_{\omega_0}-t=0. \tag{15}$$

Substituting condition (15) in Eq. (14), we obtain the signal's maximum amplitude for each point $\mathbf{r}$ as

$$a_m(\mathbf{r})=\int_{-\infty}^{\infty}|S(\omega_0+\Omega,\mathbf{r})|\exp\left[\frac{i}{2}\left(\frac{\partial^2\varphi_S}{\partial\omega^2}\right)_{\omega_0}\Omega^2\right]d\Omega. \tag{16}$$

The integration of Eq. (16) can be performed if we approach the spectrum $|S|$ of the signal as a Gaussian distribution around ω_0. This is,

$$|S(\omega_0+\Omega,\mathbf{r})|=|S(\omega_0,\mathbf{r})|\exp\left(-\frac{\Omega^2}{2\Delta^2}\right), \tag{17}$$

Δ being the half bandwidth.

Inserting Eq. (17) in Eq. (16), performing the integration, and taking the modulus of a_m squared, we obtain

$$|a_m(\mathbf{r})|^2=2\pi\frac{\Delta^4|S(\omega_0,\mathbf{r})|^2}{\sqrt{1+\Delta^4\left(\frac{\partial^2\varphi_S}{\partial\omega^2}\right)_{\omega_0}^2}}. \tag{18}$$

At this point, we write the TR transfer function H_{TR} and the Fourier transform E of the emitted signal as

$$H_{\mathrm{TR}}(\omega,\mathbf{r})=|H_{\mathrm{TR}}(\omega,\mathbf{r})|\exp[i\varphi_{TR}(\omega,\mathbf{r})],$$
$$E(\omega)=|E(\omega)|\exp[i\varphi_E(\omega)]. \tag{19}$$

Then, the modulus and the phase of S in Eq. (8) are, respectively,

$$|S(\omega,\mathbf{r})|=|H_{\mathrm{TR}}(\omega,\mathbf{r})||E(\omega)|,$$

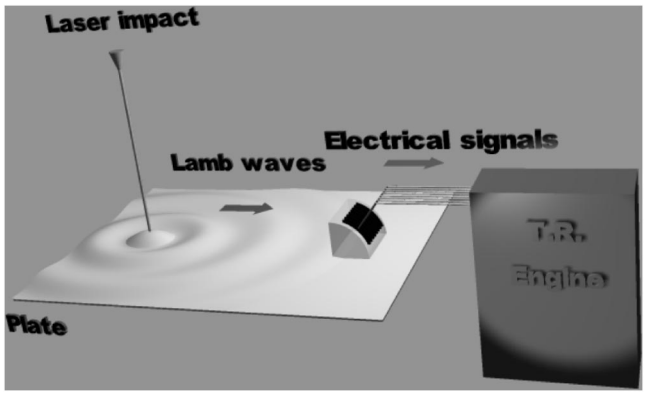


FIG. 1. The time reversal mirror (TRM) is an array of transducers driven by a time reversal engine. It performs the time reversal process of an acoustical signal emitted in a point of a plate. The acoustical source could be, for example, a focalized laser impact.

$$\varphi_S(\omega,\mathbf{r})=\varphi_{\mathrm{TR}}(\omega,\mathbf{r})-\varphi_E(\omega). \tag{20}$$

The spectrum of $|H_{\mathrm{TR}}(\omega,\mathbf{r})||E(\omega)|$ has a maximum in the central frequency ω_0, and its half bandwidth is Δ.

Substituting the modulus of S from Eq. (20) in Eq. (18), the maximum amplitude results in

$$|a_m(\mathbf{r})|^2=2\pi\frac{\Delta^4|H_{\mathrm{TR}}(\omega_0,\mathbf{r})|^2|E(\omega_0)|^2}{\sqrt{1+\Delta^4\left(\dfrac{\partial^2\varphi_S}{\partial\omega^2}\right)^2_{\omega_0}}}. \tag{21}$$

Equation (21) features the dependence of the TR refocalization process on two parameters: the modulus $|H_{\mathrm{TR}}|$ of the TR transfer function and its phase φ_{TR}. The calculation of the explicit expressions for these parameters allows the analysis and improvement of the efficiency in the refocalization process.

C. Time reversal mirror for two-dimensional acoustic waves

In order to apply Eq. (21) to surface waves or Lamb waves in thin plates, we must work with two-dimensional waves propagation. For these kinds of waves, the array of transducers can be considered as linear (Figs. 1 and 2). This array of transducers generates a time reversal mirror that is placed both perpendicular and symmetrical respect to the z axis. The punctual acoustic source is in the origin $\mathbf{r}=\mathbf{0}$.

The transfer function of propagation for two-dimensional waves in the free space is[11]

$$H(\omega,\mathbf{r})=\Theta(\omega)\frac{\exp[ik(\omega)|\mathbf{r}|]}{\sqrt{|\mathbf{r}|}}, \tag{22}$$

where $\Theta(\omega)$ is a filter factor that depends on the excited mode, in multimode propagation. $k(\omega)$ is the wave number at frequency ω.

Substituting Eq. (22) in Eq. (7), we obtain

$$H_{\mathrm{TR}}(\omega,\mathbf{r})=\exp(i\omega\tau)|h|^2|\Theta(\omega)|^2 \times\sum_n\frac{\exp[ik(\omega)(|\mathbf{r}-\mathbf{r}_n|-|\mathbf{r}_n|)]}{\sqrt{|\mathbf{r}-\mathbf{r}_n||\mathbf{r}_n|}}. \tag{23}$$

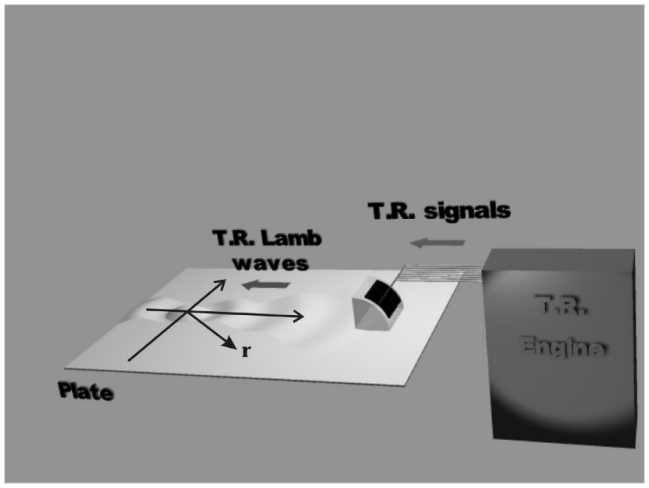


FIG. 2. The time reversed signal detected in the neighborhood of the source point (origin of coordinates). The vector $\mathbf{r}$ is a general position where the returned signal is evaluated.

In order to study the refocalization properties of the TR process, the vicinity of the original source point ($\mathbf{r}=\mathbf{0}$) results of special interest. Then, in Fig. 3 we observe that the condition $|\mathbf{r}|\ll|\mathbf{r}-\mathbf{r}_n|$ is satisfied and hence we can approximate the results

$$|\mathbf{r}-\mathbf{r}_n|-|\mathbf{r}_n|\approx-|\mathbf{r}|\cos\theta_n, \quad \frac{1}{\sqrt{|\mathbf{r}-\mathbf{r}_n||\mathbf{r}_n|}}\approx\frac{1}{|\mathbf{r}_n|}. \tag{24}$$

Considering the approximation resulting from Eq. (24), the TR transfer function in Eq. (23) is

$$H_{\mathrm{TR}}(\omega,\mathbf{r})\approx\exp(i\omega\tau)|h|^2|\Theta(\omega)|^2 \times\sum_n\frac{\exp[-ik(\omega)|\mathbf{r}|\cos\theta_n]}{|\mathbf{r}_n|}. \tag{25}$$

The sum in Eq. (25) extends over all N elemental transducers of the array in Fig. 3.

If the size of each elemental transducer is small enough (strictly, smaller than half a wavelength), we can approximate the sum in Eq. (25) by an integral over the total angular aperture γ of the time reversal mirror (Fig. 4).

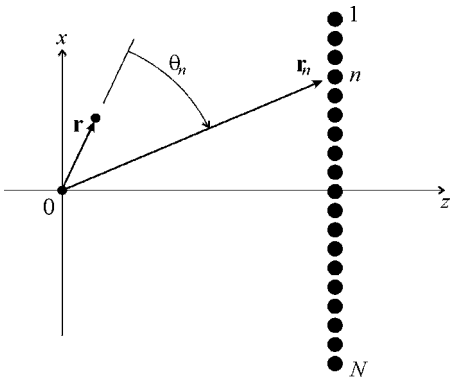


FIG. 3. Schematic representation of the experimental setup. The acoustical source is placed at the origin of coordinates. The linear array of N small transducers performs the two-dimensional time reversal process of an acoustical signal emitted in the point **0**.

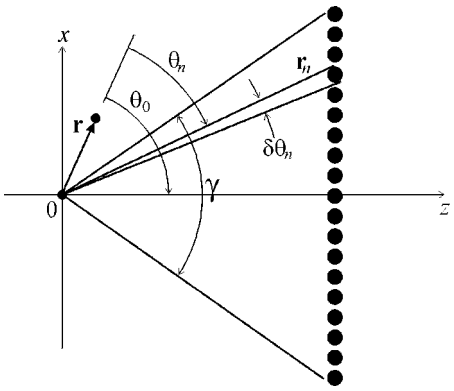


FIG. 4. Angular parameters of the TR mirror in order to calculate the total signal returned in the point **r**.

Let δL be the size of each transducer. In Fig. 4 we can see

$$\delta\theta_n = \frac{\delta L}{|\mathbf{r}_n|}. \tag{26}$$

Substituting $|\mathbf{r}_n|$ from Eq. (26) in Eq. (25), we approximate the time reversal transfer function as

$$H_{\mathrm{TR}}(\omega,\mathbf{r}) \approx \frac{|h|^2}{\delta L}\exp(i\omega\tau)|\Theta(\omega)|^2 \times \int_{\theta_0-\gamma/2}^{\theta_0+\gamma/2} \exp[-ik(\omega)|\mathbf{r}|\cos\theta]d\theta, \tag{27}$$

where θ_0 is the angular position with respect to the z axis of the observation point **r** in Fig. 4. γ is the angular aperture with respect to the original source point of the time reversal mirror.

III. MULTIMODE WAVES PROPAGATION

A. Lamb waves

Lamb waves, as guided waves in thin plates, present multimode propagation.[10] Equation (27), for the time reversal transfer function, remains for each mode m as

$$H_{\mathrm{TR}}^{(m)}(\omega,\mathbf{r}) \approx \frac{|h|^2}{\delta L}\exp(i\omega\tau)|\Theta_m(\omega)|^2 \times \int_{\theta_0-\gamma/2}^{\theta_0+\gamma/2} \exp[-ik_m(\omega)|\mathbf{r}|\cos\theta]d\theta. \tag{28}$$

Due to the linearity of the processes, the time reversal transfer function for all propagating modes is

$$H_{\mathrm{TR}}(\omega,\mathbf{r}) = \sum_m H_{\mathrm{TR}}^{(m)}(\omega,\mathbf{r}). \tag{29}$$

Substituting Eq. (28) in Eq. (29), we obtain

$$H_{\mathrm{TR}}(\omega,|\mathbf{r}|) \approx \frac{|h|^2}{\delta L}\exp(i\omega\tau)\int_{\theta_0-\gamma/2}^{\theta_0+\gamma/2}\left\{\sum_m |\Theta_m(\omega)|^2 \times \exp[-ik_m(\omega)|\mathbf{r}|\cos\theta]\right\}d\theta, \tag{30}$$

where summation extends over all modes m, with their respective weight factors $|\Theta_m(\omega)|^2$.

These filter factors $\Theta_m(\omega)$ arise from the Green functions for Lamb waves[11] called $H_{xx}^{(m)}(\omega)$ in the referred paper. They can be calculated in function of the amplitude of the normal displacement $W^{(m)}(\omega)$ on the plate surface, and in function of the average flow power per unit length[14] $P_m(\omega)$ for each mode m, as

$$\Theta_m(\omega) = \frac{i\pi\omega}{2P_m(\omega)}|W^{(m)}(\omega)|^2. \tag{31}$$

The sum in Eq. (30) is a complex number. We can write it as

$$\sum_m |\Theta_m(\omega)|^2 \exp[-ik_m(\omega)|\mathbf{r}|\cos\theta] = Q(\omega,\mathbf{r})\exp[-i\phi(\omega,\mathbf{r})]. \tag{32}$$

In the vicinity of the origin we verify the condition $k_m(\omega)|\mathbf{r}| \ll 1$. Therefore, we insert the second-order approximations in Eq. (32)

$$\exp[-ik_m|\mathbf{r}|\cos\theta] \approx 1 - ik_m|\mathbf{r}|\cos\theta - \tfrac{1}{2}k_m^2|\mathbf{r}|^2\cos^2\theta,$$
$$\exp[-i\phi] \approx 1 - i\phi - \tfrac{1}{2}\phi^2. \tag{33}$$

Equating the imaginary parts in Eq. (32) to the second approximation order, we have, for the phase ϕ,

$$Q\phi = \sum_m |\Theta_m|^2 k_m|\mathbf{r}|\cos\theta. \tag{34}$$

We define the *average wave number* for multimode propagation as

$$\overline{k}(\omega) = \frac{1}{\Sigma_m|\Theta_m(\omega)|^2}\sum_m |\Theta_m(\omega)|^2 k_m(\omega), \tag{35}$$

and the *average square wave number* as

$$\overline{k^2}(\omega) = \frac{1}{\Sigma_m|\Theta_m(\omega)|^2}\sum_m |\Theta_m(\omega)|^2 k_m^2(\omega). \tag{36}$$

Then, Eq. (34) can be expressed as

$$Q\phi = \sum_m |\Theta_m|^2 \overline{k}(\omega)|\mathbf{r}|\cos\theta. \tag{37}$$

On the other hand, equating the real parts in Eq. (32), and keeping the development up to the second order, we obtain

$$Q = \sum_m |\Theta_m|^2\left(1 - \frac{1}{2}\sigma_k^2\left|\mathbf{r}\right|^2\cos^2\theta\right), \tag{38}$$

σ_k being the standard deviation of the k_m values. This is,

$$\sigma_k^2(\omega) = \overline{k^2}(\omega) - \overline{k}^2(\omega). \tag{39}$$

We substitute Q from Eq. (38) in Eq. (37), and we keep the development up to the second order. The resulting phase becomes

$$\phi \approx \overline{k}(\omega)|\mathbf{r}|\cos\theta. \tag{40}$$

The final expression for summation (32) over all propagating modes is

$$\sum_m |\Theta_m(\omega)|^2 \exp[-ik_m(\omega)|\mathbf{r}|\cos\theta] = \sum_m |\Theta_m|^2 \left(1-\frac{1}{2}\sigma_k^2 \left|\mathbf{r}\right|^2 \cos^2\theta\right) \exp(-i\overline{k}|\mathbf{r}|\cos\theta). \tag{41}$$

In order to be coherent, we must express the second member of Eq. (41) up to the second order in $|\mathbf{r}|$. This is

$$\sum_m |\Theta_m(\omega)|^2 \exp[-ik_m(\omega)|\mathbf{r}|\cos\theta] \approx \Omega\left(1-\frac{\overline{k^2}+\sigma_k^2}{2}|\mathbf{r}|^2\cos^2\theta - i\overline{k}|\mathbf{r}|\cos\theta\right), \tag{42}$$

where we have abbreviated the positive amount

$$\Omega = \sum_m |\Theta_m(\omega)|^2. \tag{43}$$

Inserting result (42) in Eq. (30) and performing the integration, we obtain an expression for the time reversal transfer function as

$$H_{\mathrm{TR}}(\omega,|\mathbf{r}|) \approx \frac{|h|^2 \exp(i\omega\tau)}{\delta L}\Omega\gamma\left[1-i\overline{k}|\mathbf{r}|\cos\theta_0\frac{\sin(\gamma/2)}{\gamma/2} - \frac{\overline{k^2}+\sigma_k^2}{4}|\mathbf{r}|^2\left(1+\cos 2\theta_0\frac{\sin\gamma}{\gamma}\right)\right]. \tag{44}$$

We want to use Eq. (21) in order to calculate the maximum amplitude in the time reversal refocalization process. So, we must express Eq. (44) as

$$H_{\mathrm{TR}} = |H_{\mathrm{TR}}|\exp(i\varphi_{\mathrm{TR}}). \tag{45}$$

According to Eq. (44), and preserving the development up to the second order in $|\mathbf{r}|$, we obtain

$$\varphi_{\mathrm{TR}}(\omega,\mathbf{r}) \approx \omega\tau - \overline{k}(\omega)|\mathbf{r}|\cos\theta_0\frac{\sin(\gamma/2)}{\gamma/2}, \tag{46}$$

and

$$|H_{\mathrm{TR}}|^2 \approx |h|^4\left(\frac{\Omega\gamma}{\delta L}\right)^2(1-e_S|\mathbf{r}|^2), \tag{47}$$

e_S being

$$e_S = \frac{\overline{k^2}+\sigma_k^2}{2}\left(1+\cos 2\theta_0\frac{\sin\gamma}{\gamma}\right) - \overline{k^2}\cos^2\theta_0\left[\frac{\sin(\gamma/2)}{\gamma/2}\right]^2. \tag{48}$$

B. Focalization efficiencies

Using Eq. (46), the second derivative of the total phase φ_S of Eq. (20) with respect to ω is

$$\frac{\partial^2\varphi_S}{\partial\omega^2} \approx \frac{d^2\overline{k}}{d\omega^2}\cos\theta_0\frac{\sin(\gamma/2)}{\gamma/2} + \frac{d^2\varphi_E}{d\omega^2}. \tag{49}$$

At this point, we are in a good position to evaluate the maximum amplitude of the time reversal focalization process for surface waves, in dispersive media with multimode propagation. Substituting Eqs. (49) and (47) in Eq. (21), this last one results in

$$|a_m(\mathbf{r})|^2 = |a_m(\mathbf{0})|^2\frac{1-e_S|\mathbf{r}|^2}{\sqrt{1+e_T|\mathbf{r}|^2}}, \tag{50}$$

where

$$|a_m(0)|^2 = 2\pi|h|^4\left(\frac{\Omega\gamma}{\delta L}\right)^2\Delta^4|E(\omega_0)|^2, \tag{51}$$

and

$$e_T = \Delta^4\left[\left(\frac{d^2\overline{k}}{d\omega^2}\right)_{\omega_0}\cos\theta_0\frac{\sin(\gamma/2)}{\gamma/2} + \left(\frac{d^2\varphi_E}{d\omega^2}\right)_{\omega_0}\right]^2. \tag{52}$$

ω_0 is the central frequency of the wave packet, and Δ is its half bandwidth.

The positive parameters e_S and e_T defined in Eqs. (48) and (52), respectively, determine the efficiency of the refocalization process. According to Eq. (50), the greater these parameters are, the sharpest the focalization peak centered in $\mathbf{r}=\mathbf{0}$ is.

We call the parameter e_S defined in Eq. (48) *spatial efficiency of focalization*, and we call the parameter e_T defined in Eq. (52) *temporal efficiency of focalization.*

The name for the parameter e_S, *spatial efficiency*, arises from its strong dependence on diffraction effects due to the finite size of the time reversal mirror. If we approximate the angular aperture to $\gamma\approx 0$ in Eq. (48), then we have

$$e_S = \sigma_k^2\cos^2\theta_0. \tag{53}$$

For a one-mode propagation only, we have $\sigma_k = 0$ in Eq. (53), and there is no chance of focalization by this means.

The name for the parameter e_T, *temporal efficiency*, arises from its dependence on the dispersive properties of the medium. If it is nondispersive, the dispersion curves $k_m(\omega)$ are straight lines and the second derivative is zero in Eq. (52). There is no temporal focalization if the phase φ_E approaches to a linear function in the vicinity of ω_0.

C. Longitudinal and transversal focalizations

We call the analysis of temporal and spatial efficiencies along the z axis in Fig. 1 *longitudinal focalization* (see Fig. 5). This is obtained by inserting $\theta_0 = 0, \pi$ in Eq. (48) for the spatial efficiency, and in Eq. (52) for the temporal efficiency.

Hence, Eq. (48) for spatial and longitudinal focalization efficiency becomes

$$e_{S,\mathrm{long}} = \frac{\overline{k^2}+\sigma_k^2}{2}\left(1+\frac{\sin\gamma}{\gamma}\right) - \overline{k^2}\left[\frac{\sin(\gamma/2)}{\gamma/2}\right]^2. \tag{54}$$

In order to obtain the temporal and longitudinal focalization efficiency, let us point out that if the central frequency ω_0 of the wave packet is also the peak of the power spectrum of the emitted signal $E(\omega)$, then its phase $\varphi_E(\omega)$ has an inflexion point in ω_0. This means that $(d^2\varphi_E/d\omega^2)_{\omega_0} = 0$ in Eq. (52). Substituting there $\theta_0 = 0, \pi$,

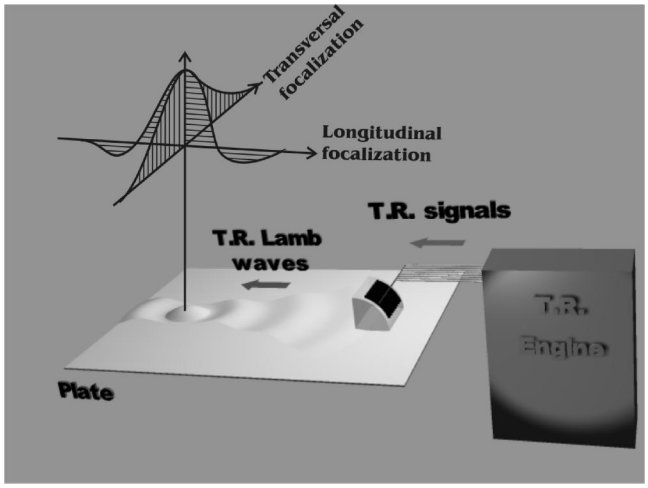


FIG. 5. Transversal and longitudinal focalizations in two-dimensional TR processes.

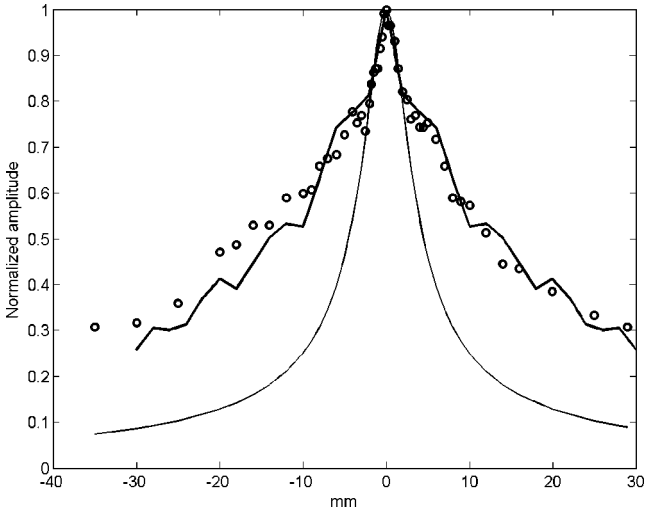


FIG. 6. Longitudinal time reversal focalization in an aluminum plate of 1.1 mm thickness. Experimental measures from Ing and Fink—Refs. 8 and 9 (circles), our theoretical result (thick continuous line), and the near focal point approximation (thin continuous line).

the temporal and longitudinal focalization efficiency becomes

$$e_{T,\mathrm{long}}=\Delta^4\left(\frac{d^2\overline{k}}{d\omega^2}\right)^2_{\omega_0}\left[\frac{\sin(\gamma/2)}{\gamma/2}\right]^2. \tag{55}$$

We call the analysis of temporal and spatial efficiencies along the x axis in Fig. 1 *transversal focalization* (see Fig. 5). We obtain this by substituting $\theta_0=\pm\pi/2$ in Eq. (48) for the spatial efficiency, and in Eq. (52) for the temporal efficiency. In the same conditions in which $(d^2\varphi_E/d\omega^2)_{\omega_0}=0$, these equations become

$$e_{S,\mathrm{trans}}=\frac{\overline{k^2}+\sigma_k^2}{2}\left(1-\frac{\sin\gamma}{\gamma}\right), \tag{56}$$

for the spatial and transversal focalization efficiency, and

$$e_{T,\mathrm{trans}}=0, \tag{57}$$

for the temporal and transversal focalization efficiency.

Even though in the most common cases, the angular aperture γ of the TR mirror with respect to the focal point is small, it is not negligible. We can use a second-order approximation for the developments

$$\frac{\sin\gamma}{\gamma}\approx 1-\frac{\gamma^2}{6},\quad \left[\frac{\sin(\gamma/2)}{\gamma/2}\right]^2\approx 1-\frac{\gamma^2}{12}. \tag{58}$$

With these approximations, Eqs. (54) and (55) for longitudinal efficiencies become

$$e_{S,\mathrm{long}}=\sigma_k^2\left(1-\frac{\gamma^2}{12}\right)\approx\sigma_k^2,$$

$$e_{T,\mathrm{long}}=\Delta^4\left(\frac{d^2\overline{k}}{d\omega^2}\right)^2_{\omega_0}\left(1-\frac{\gamma^2}{12}\right)\approx\Delta^4\left(\frac{d^2\overline{k}}{d\omega^2}\right)^2_{\omega_0}. \tag{59}$$

Approximations (58) for small aperture inserted in Eqs. (56) and (57) yield the transversal efficiencies

$$e_{S,\mathrm{trans}}=\frac{\overline{k^2}+\sigma_k^2}{12}\gamma^2,\quad e_{T,\mathrm{trans}}=0. \tag{60}$$

Relationships (59) and (60) are the most useful in order to evaluate the performance of the TR refocalization process for two-dimensional and dispersive waves with multimode propagation. These parameters determine the sharpness of the focalization peak in Eq. (50). Figure 5 shows those peaks (qualitatively) in longitudinal and transversal cases.

When there is only one efficiently returned mode, then $\overline{k(\omega)}=k(\omega)$ and $\sigma_k=0$. Equation (59) for longitudinal focalization efficiencies becomes

$$e_{S,\mathrm{long}}=0,\quad e_{T,\mathrm{long}}(\omega_0)=\Delta^4\left(\frac{d^2\overline{k}}{d\omega^2}\right)^2_{\omega_0}. \tag{61}$$

Equation (60) for transversal focalization efficiencies becomes

$$e_{S,\mathrm{trans}}=\frac{k^2(\omega_0)\gamma^2}{12},\quad e_{T,\mathrm{trans}}=0. \tag{62}$$

The possibility of longitudinal refocalization by a TR process depends on the dispersion of the returned mode at the central frequency, as Eq. (61) determines. If such mode is not dispersive enough at the frequency ω_0, the longitudinal focalization is very poor. On the other hand, the transversal focalization depends on the angular aperture of the TR mirror and on the wave number k, as Eq. (62) shows.

IV. EXPERIMENTAL RESULTS

Maximum amplitude experimental values of time reversed Lamb waves in thin plates were obtained by Ing and Fink.[8,9] Lamb waves are generated on aluminum plates of three different thicknesses (1.1, 1.5, and 2.9 mm) using a short-time-duration (80 ns) laser impact in ablation regime. We apply our results to the 1.1 mm plate. The device used to detect Lamb waves is a transducer array of 32 elements coupled to a Lucite wedge. The array has a central frequency of 4 MHz and a half bandwidth nearly 1 MHz.

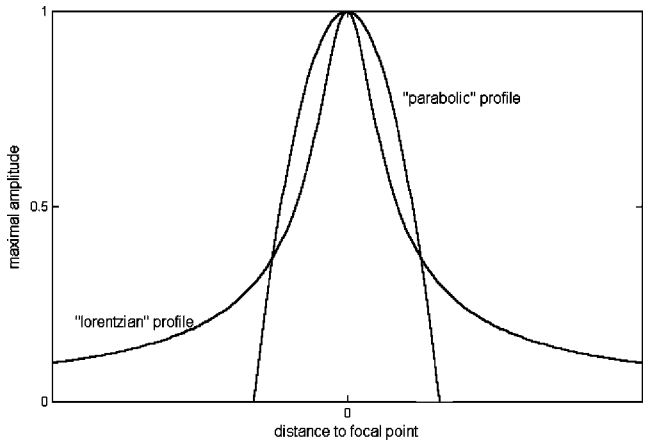


FIG. 7. Maximal amplitude in TR focalization around the focal point. ''Lorentzian'' profile (when the temporal focalization is dominant), and ''parabolic'' profile (when the spatial focalization is dominant).

Each elemental transducer measures 10 mm in height and 1.085 mm in pitch. The total width of the array is 33.7 mm and the laser impact point is placed 255 mm away from the transducer. Then, the angular aperture γ of the TR mirror is

$$\gamma \approx \frac{33.7}{255} \approx 0.13. \tag{63}$$

In the referred papers, the authors verify that the mainly returned mode in each plate is the first symmetrical S_0. So, we can evaluate numerically Eq. (30) in order to obtain the time reversal transfer function H_{TR} as a function of ω, for longitudinal focalization ($\theta_0 = 0$) and for the only mode S_0. We also assume that the acousto-electrical transfer function h is a constant.

If we suppose that the laser impact excitation behaves like a δ-Dirac, then we consider its spectrum $E^*(\omega)$ as a constant. We apply Eq. (8) in order to evaluate the spectra $S(\omega, \mathbf{r})$ of the time-reversed signals for each point $\mathbf{r}$ (in longitudinal direction), we perform an inverse Fourier transform to obtain temporal signals $s(\mathbf{r}, t)$ and we pick up the maximal amplitudes $a(\mathbf{r})$.

Figure 6 shows theoretical results (thick continuous line) and experimental results (circles). The near focal point approximation (thin line) is valid only in the neighborhood of the focus (± 3 mm), as we should expect. Nevertheless, a more accurate calculation using Eqs. (30) and (31) gives the thick line. In order to apply such equations, we suppose that only the first symmetrical and first antisymmetrical mode are returned.

As the mainly returned mode in the plate is the first symmetrical S_0, we are under the conditions expressed by Eqs. (61) and (62). First, we observe that the only longitudinal focalization in Eq. (61) is the temporal $e_{T,\mathrm{long}}$. Equation (50), with $e_S = 0$, shows that the maximum amplitude has a ''Lorentzian'' profile in the vicinity of the focal point, as shown in Fig. 7. This should be the characteristic behavior when $e_T \gg e_S$.

On the other hand, Eq. (62) shows that spatial focalization is present only in the transversal direction. Substituting $e_T = 0$ in Eq. (50), we observe that the maximum amplitude has a ''parabolic'' profile in the vicinity of the focal point. This should be the characteristic behavior when $e_T \ll e_S$.

The second derivative $k''(\omega)$ for the mode S_0 is shown in Fig. 8 for each one of the three plates of different thicknesses: 1.1, 1.5, and 2.9 mm. In the central frequency of 4 MHz, the absolute magnitude of k'' is greater for the thinnest plate, and it almost vanishes for the thickest plate. From what is implied in Eq. (61), we can predict that the longitudinal focalization will be better for the thinnest plate and poorer for the thickest one.

We can see in Fig. 8 that the second derivative of the S_0 wave number is nearly $d^2k/d\omega^2 \approx 0.01\ \mu\mathrm{s}^2/\mathrm{mm}$ for the 1.1 mm plate in the central frequency of 4 MHz. The half bandwidth is nearly 1 MHz, then $\Delta \approx 2\pi \times 1\ \mathrm{MHz} \approx 6\ \mu\mathrm{s}^{-1}$ in Eq. (61). From these equations, the longitudinal focalization efficiencies are $e_S = 0$, $e_T \approx 0.16\ \mathrm{l/mm}^2$. We plot in Fig. 6 maximal amplitudes obtained with Eq. (50) (thin continuous line). Obviously, this equation is only valid in the neighborhood of the origin (Fig. 9).

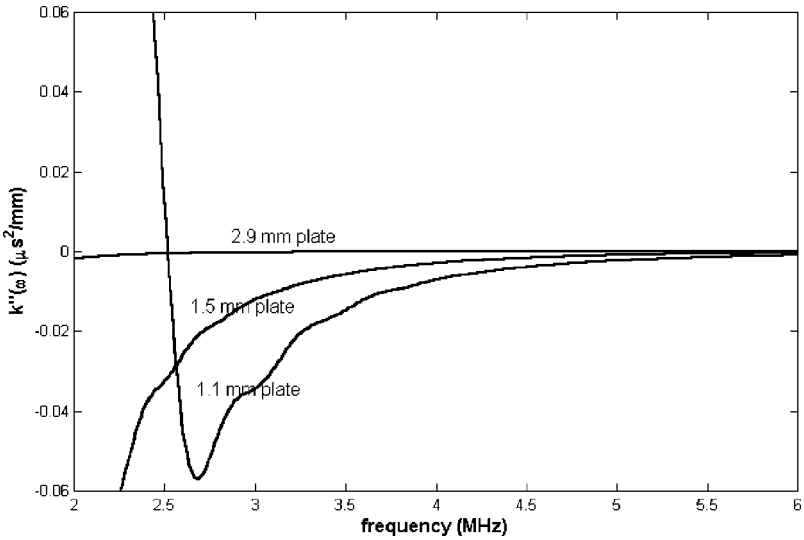


FIG. 8. Second derivative of the S_0 wave number in function of the frequency, for three aluminum plates of different thickness.

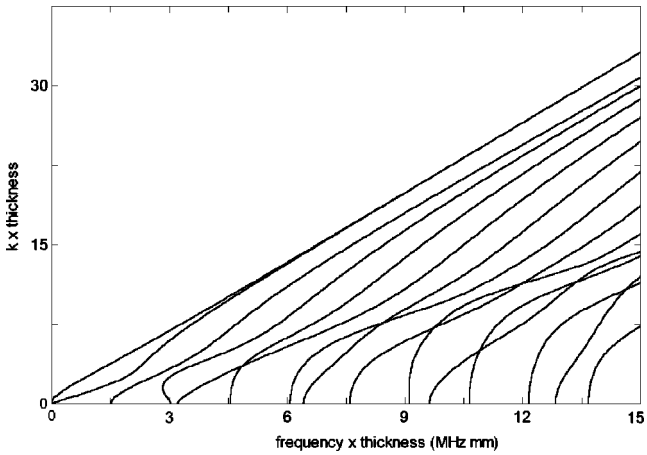


FIG. 9. Dispersion curves for the first symmetrical and antisymmetrical Lamb modes in an aluminum plate of arbitrary thickness.

V. CONCLUSIONS

This work aims at generating a procedure to anticipate the efficiency of the time reversal refocalization process in dispersive media with multimode propagation. The main results are the equations to calculate the longitudinal and transversal focalization efficiencies, with small aperture TR mirrors, since these are the common experimental conditions.

In the longitudinal focalization (along the propagation axis), we find that the time reversal refocalization is better as both the standard deviation σ_k among exciting modes and the dispersion properties of the medium (through the second derivative of k) are higher. The amount σ_k increases with the frequency because of the generation of new modes. Nevertheless, we must point out that the behavior of the dispersion curves tends to be linear for lower modes, when the product frequency×thickness is high, as Fig. 8 shows. Therefore, the second derivative of the mean value k tends to vanish, and the longitudinal focalization efficiency mainly comes from the amount σ_k.

We can improve the longitudinal focalization efficiency by sending a short pulse (large bandwidth Δ). One of the most important contributions of this work is demonstrating that there is another way to perform the improvement. This can be done by returning as many modes as possible in the TR process, generating an increase in the amount σ_k.

On a different front, we can deduce from the equations for longitudinal focalization efficiencies that the angular aperture γ of the TR mirror is not significant.

Another important contribution of this work is showing that, in transversal focalization efficiencies, the greater the amounts of k and σ_k are, the better the time reversal refocalization process is. Both said amounts increase when the central frequency of the returned signal increases (Fig. 8). The first one increases because the k values for each of the different modes increase with the frequency. The second one increases because of the generation of new modes.

We also observe that the role of the angular aperture γ of the TR mirror is significant for the transversal focalization efficiency. This is an expected result since the transversal focalization strongly depends on the diffraction effects, for all wave propagation phenomena. The new feature of our results is that the other way to enhance the transversal focalization is to increase the frequency.

In the common cases in which the TR process can return only one Lamb mode, the standard deviation σ_k is zero. In these cases we must choose an adequate central frequency for the TR mirror operation. In order to enhance the longitudinal focalization, we must work at the frequency in which the graph of the dispersion curve $k(\omega)$ has the greatest curvature (highly dispersive region) for the used mode. In order to enhance the transversal focalization, we must choose the greatest possible angular aperture γ of the TR mirror.

[1] M. Fink, "Time reversal of ultrasonic fields. I. Basic Principles," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**, 555–566 (1992).
[2] M. Fink *et al.*, "Time reversed acoustics," Rep. Prog. Phys. **63**, 1933–1995 (2000).
[3] A. Derode, P. Roux, and M. Fink, "Robust acoustics time reversal with high order multiple scattering," Phys. Rev. Lett. **75**, 4206–4209 (1995).
[4] A. Derode, A. Tourin, and M. Fink, "Ultrasonic pulse compression with one-bit time reversal through multiple scattering," J. Appl. Phys. **85**, 6343–6344 (1999).
[5] P. Roux, B. Roman, and M. Fink, "Time reversal in ultrasonic wave guides," Appl. Phys. Lett. **70**, 1811–1813 (1997).
[6] H. Song, W. Kuperman, W. Hodgkiss, T. Akal, and C. Ferla, "Iterative time reversal in the ocean," J. Acoust. Soc. Am. **105**, 3176–3184 (1999).
[7] C. Draeger and M. Fink, "One channel time reversal of elastic waves in a chaotic 2D-silicon cavity," Phys. Rev. Lett. **79**, 407–410 (1997).
[8] R. K. Ing and M. Fink, "Self-focusing and time recompression of Lamb waves using a time reversal mirror," Ultrasonics **36**, 179–186 (1998).
[9] R. K. Ing and M. Fink, "Time-reversed Lamb Waves," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**, 1032–1043 (1998).
[10] I. A. Viktorov, *Rayleigh and Lamb Waves* (Plenum, New York, 1967).
[11] I. Núñez, R. K. Ing, C. Negreira, and M. Fink, "Transfer and Green functions based in modal analysis for Lamb waves generation," J. Acoust. Soc. Am. **107**, 2370–2378 (2000).
[12] G. S. Kino, *Acoustic Waves: Devices, Imaging, and Analog Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1987).
[13] B. A. Auld, *Acoustic Fields and Waves in Solids* (Krieger, 1973).
[14] X. Jia, "Modal analysis of Lamb wave generation in elastic plates by liquid wedge transducers," J. Acoust. Soc. Am. **101**, 834–842 (1997).

Encoding of vowel-like sounds in the auditory nerve: Model predictions of discrimination performance

Qing Tan
Boston University Hearing Research Center, Department of Biomedical Engineering, Boston University, 44 Cummington Street, Boston, Massachusetts 02215

Laurel H. Carney[a)]
Boston University Hearing Research Center, Department of Biomedical Engineering, Boston University, 44 Cummington Street, Boston, Massachusetts 02215, and Department of Bioengineering & Neuroscience, Institute for Sensory Research, 621 Skytop Road, Syracuse University, Syracuse, New York 13244

(Received 2 May 2004; revised 15 December 2004; accepted 16 December 2004)

The sensitivity of listeners to changes in the center frequency of vowel-like harmonic complexes as a function of the center frequency of the complex cannot be explained by changes in the level of the stimulus [Lyzenga and Horst, J. Acoust. Soc. Am. **98**, 1943–1955 (1995)]. Rather, a complex pattern of sensitivity is seen; for a spectrum with a triangular envelope, the greatest sensitivity occurs when the center frequency falls between harmonics, whereas for a spectrum with a trapezoidal envelope, greatest sensitivity occurs when the center frequency is aligned with a harmonic. In this study, the thresholds of a population model of auditory-nerve (AN) fibers were quantitatively compared to these trends in psychophysical thresholds. Single-fiber and population model responses were evaluated in terms of both average discharge rate and the combination of rate and timing information. Results indicate that phase-locked responses of AN fibers encode phase transitions associated with minima in these amplitude-modulated stimuli. The temporal response properties of a single AN fiber, tuned to a frequency slightly above the center frequency of the harmonic complex, were able to explain the trends in thresholds for both triangular- and trapezoidal-shaped spectra.
 [DOI: 10.1121/1.1856391]

PACS numbers: 43.64.Bt [WPS] Pages: 1210–1222

I. INTRODUCTION

The cues used by listeners to detect spectral changes in vowels have been studied for many years. However, the cues embedded in vowel signals and the mechanisms used by the auditory system to encode and process these cues are still not completely clear. Formant frequencies characterize the basic shape of the speech spectrum and are important for phonetic identification (Rabiner and Schafer, 1978). Estimating the ability of the auditory system to resolve changes in formant frequency is a first step in understanding speech processing in the auditory system. Psychophysical experiments have estimated formant-frequency discrimination ability (Flanagan, 1955; Mermelstein, 1978; Sinnott and Kreiter, 1991; Kewley-Port and Watson, 1994); however, reported thresholds of the formant-frequency discrimination tasks have differed among studies because of the complexity of the stimuli and differences in experimental procedures. For example, Mermelstein (1978) found that the threshold for discriminating changes in the first formant at 350 Hz was 50 Hz, which is much higher than the result of Flanagan (1955), who reported discrimination thresholds for the first formant (at 300 Hz) of 12 to 17 Hz.

Lyzenga and Horst (1995) conducted an interesting set of experiments concerning the ability to discriminate changes in the center frequency of bandlimited harmonic complexes (Fig. 1), which are a convenient simplification of synthetic vowel signals. Figure 2(a) shows Lyzenga and Horst's (1995) results for triangular spectra with different spectral slopes; the highest thresholds for discrimination of center frequency are near the center frequencies of 2000 and 2100 Hz, when the peak of the spectral envelope is near a harmonic frequency [e.g., Fig. 1(b)]. The center-frequency discrimination threshold is lowest [Fig. 2(a), center frequency=2050 Hz] when the peak of the spectral envelope is between two harmonic components [Fig. 1(a)]. In contrast, the thresholds were lowest for the discrimination task with a trapezoidal spectral envelope [Figs. 1(c), (d)] when the center frequency was near a harmonic frequency [Fig. 2(b), center frequency=2000, 2100, or 2200 Hz; Fig. 1(d)].

In the same study, just noticeable differences (jnd's) for the center frequency of the spectral envelope were measured with a randomly varied signal level (Lyzenga and Horst, 1995). The roving-level paradigm makes signal level less reliable as a cue to detect frequency change. Thresholds with and without the roving signal level show similar trends across frequency (Fig. 2, dotted and dashed lines), with slightly elevated thresholds for the roving condition. The ratio of roving versus nonroving jnd's (keeping all the other parameters the same) is about 1.5 in most cases (Lyzenga and Horst, 1995). This result suggests that the auditory system does not rely on level cues to encode the center frequency of harmonic complexes.

[a)]Author to whom correspondence should be addressed. Electronic mail: lacarney@syr.edu

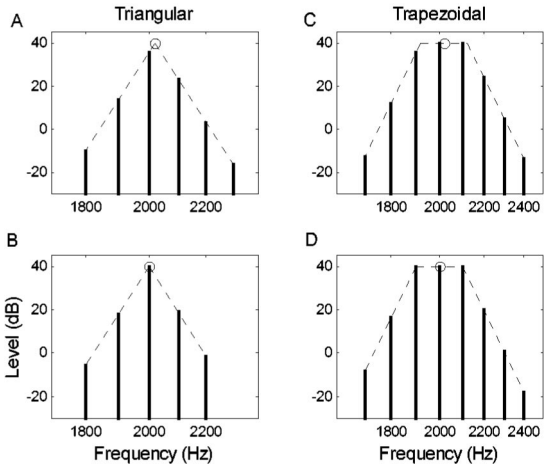


FIG. 1. Examples of the harmonic-complex spectra, with a triangular envelope (a), (b) or a trapezoidal envelope (c), (d). All spectra have a fundamental frequency of 100 Hz. The bold circles on top of each spectrum indicate the center frequency. The spectra on the bottom row (b), (d) are centered at 2000 Hz, while the spectra on the top row (a), (c) are centered at 2020 Hz (shifted 20 Hz away from 2000 Hz).

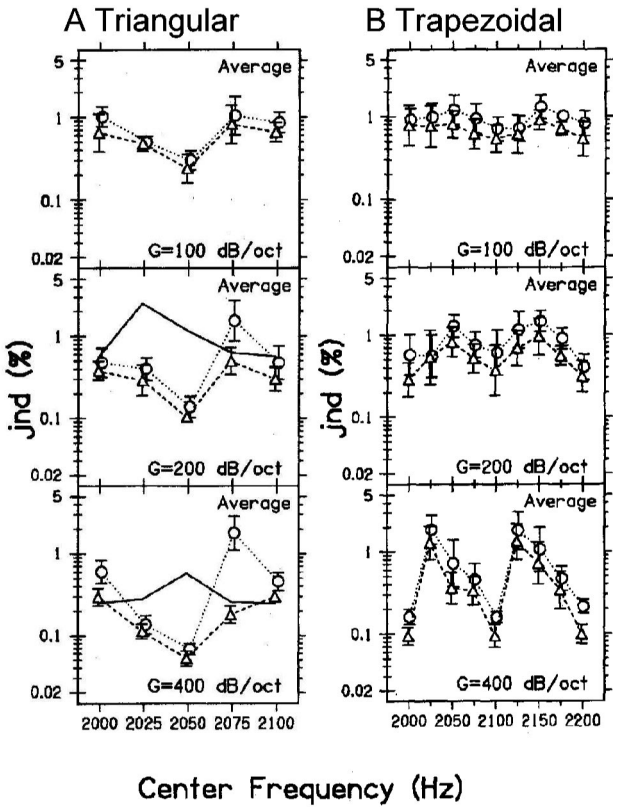


FIG. 2. Human thresholds for center-frequency discrimination of the harmonic-complexes (from Lyzenga and Horst, 1995; reprinted with permission) with (a) triangular spectrum envelope and (b) trapezoidal spectrum envelope. Each row corresponds to the thresholds for a different slope of the harmonic envelope: $G=100$, 200, and 400 dB/oct from top to bottom. The solid lines [(a), lower two panels] are predictions based on the change in the overall level of the stimuli. The dashed lines with triangle signs are human thresholds for the experiments described in the text. The dotted lines with circles are human thresholds with a random within-trial rove of the stimulus level (Lyzenga and Horst, 1995).

In the current study, thresholds for center-frequency discrimination were estimated based on the response patterns of a computational model for a population of auditory-nerve (AN) fibers. A general approach to quantifying the ability of AN population responses to explain psychophysical thresholds was proposed by Siebert (1965), who combined an analytical model of the peripheral auditory system with an ideal central processor to predict performance limits in psychophysical tasks. The discrimination ability of the ideal central processor can be estimated with methods from the theory of statistical hypothesis testing. Heinz *et al.* (2001a) adopted Siebert's ideal-processor mechanism and combined it with a detailed computational AN model in a study of monaural level and frequency discrimination. In this study, the Heinz *et al.* (2001a) approach was applied to the problem of center-frequency discrimination of harmonic complexes, and model predictions were compared with the psychophysical results of Lyzenga and Horst (1995). Predictions based on average rate of the AN responses were compared with predictions based on both average rate and the fine structure of the AN response patterns (i.e., the timing information). The Tan and Carney (2003) computational AN model was used to simulate responses of the population of AN fibers to the harmonic-complex signals.

A study of the coding mechanisms used by the peripheral auditory system is important to understand how speech signals are encoded. The purpose of this project is to explore the cues used by the auditory system in formant-frequency discrimination tasks. The study was not designed to identify the neural processing mechanism that achieved the best performance (i.e., lowest threshold); rather, the goal was to identify neural cues and mechanisms that can explain the performance of listeners. Thus, predicting the trends in the psychophysical results was the focus, not the absolute values of the thresholds. In general, the model thresholds were better than psychophysical thresholds, but model thresholds could be modified by the addition of internal noise (i.e., randomness of the neural discharges) or by assuming that fewer AN model fibers were engaged in the task.

II. METHODS

A. Stimuli

Two center-frequency discrimination experiments by Lyzenga and Horst (1995) were simulated using bandlimited harmonic complexes with a fundamental frequency of 100 Hz (Fig. 1). Stimulus parameters were the shape (triangle or trapezoid), the slope ($G=100$, 200, or 400 dB/oct), and the center frequency of the spectral envelope (from 2000 to 2100 Hz for the triangular envelope and from 2000 to 2200 Hz for the trapezoidal envelope). In the first experiment, the spectral envelope was triangular on a log–log scale [Figs. 1(a), (b)]. In the second experiment, the spectral envelope was trapezoidal on a log–log scale, with a 200-Hz-wide constant-level plateau [Figs. 1(c), (d)]. The fundamental frequency was always 100 Hz, and all frequency components of the complexes had a starting phase angle of zero degrees. Signal duration in each trial was 250 ms, including 25-ms onset and offset ramps shaped by a raised cosine.

As in the physiological experiments, the frequencies of the harmonic components (Fig. 1, vertical lines) were held

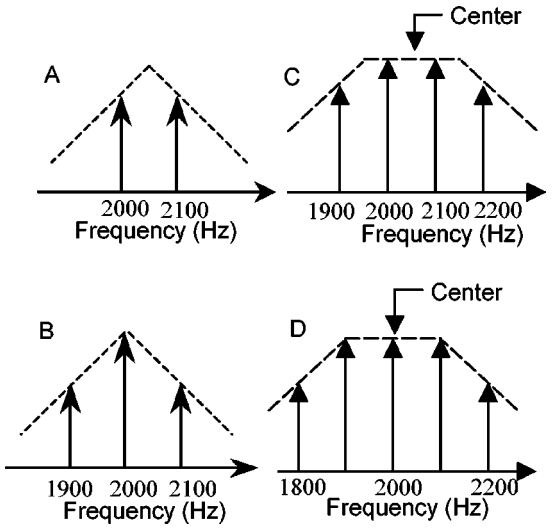


FIG. 3. Simplified harmonic-complex signals: (a) triangular envelope with center frequency at 2050 Hz; (b) triangular envelope with center frequency at 2000 Hz; (c) trapezoidal envelope with center frequency at 2050 Hz; (d) trapezoidal envelope with center frequency at 2000 Hz. The component amplitudes correspond to those in the 400-dB/oct slope condition.

constant throughout all simulations. The task was to discriminate changes in the center frequency (Fig. 1, circles) of the spectral envelope (Fig. 1, dashed lines). The magnitudes of the harmonic components changed as the center frequency of the spectrum envelope shifted to lower or higher frequencies. For example, the center frequency of the harmonic complex in Fig. 1(b) was 2000 Hz. When this center frequency shifted to a frequency slightly higher than 2000 Hz (e.g., 2005 Hz), the magnitude of all the components with frequencies higher than 2000 Hz increased, and the magnitude of all the components with frequency lower than 2000 Hz decreased. When the center frequency decreased slightly, the magnitudes of the components with frequencies lower than 2000 Hz increased, and the magnitudes of the components with frequencies higher than 2000 Hz decreased.

To better understand the features of the harmonic complexes and the performance predicted by the AN population model, it was useful to consider simpler signals with fewer components in addition to the harmonic complexes described above. The simplified signal also made mathematical analysis more tractable. We will illustrate stimuli with center frequencies of 2000 and 2050 Hz because there are large differences in psychophysical thresholds for these two center frequencies (Fig. 2; Lyzenga and Horst, 1995). Figure 3 demonstrates simplified versions of the stimuli in Fig. 1 with triangular (left) and trapezoidal spectra (right). For the triangular spectrum with center frequency at 2050 Hz [Fig. 3(a)], only the two harmonic components closest to the center of the envelope were included. In this case, the simplified signal combined two sinusoids with the same amplitude. This combination of signals can be represented as a sinusoidal signal modulated by a cosine

$$\sin(2\pi f_1 t)+\sin(2\pi f_2 t) = 2\underbrace{\sin\left(\frac{2\pi(f_1+f_2)t}{2}\right)}_{\text{Carrier}}\underbrace{\cos\left(\frac{2\pi(f_1-f_2)t}{2}\right)}_{\text{Modulator}}. \quad (1)$$

The cosine modulator serves as the envelope of the signal. An interesting feature of this simplified signal is that at the zero-crossing point of the cosine signal (when the cosine signal changes from positive to negative or from negative to positive), there is a 180-deg phase change in the fine structure of the harmonic complex's temporal waveform.

Figure 4 shows the simplified signals in the time domain. In Fig. 4(a), the thick solid line is the simplified two-component signal with center frequency at 2050 Hz [corresponding to the spectrum in Fig. 3(a)], and the thin solid line is the simplified signal with center frequency at 2060 Hz

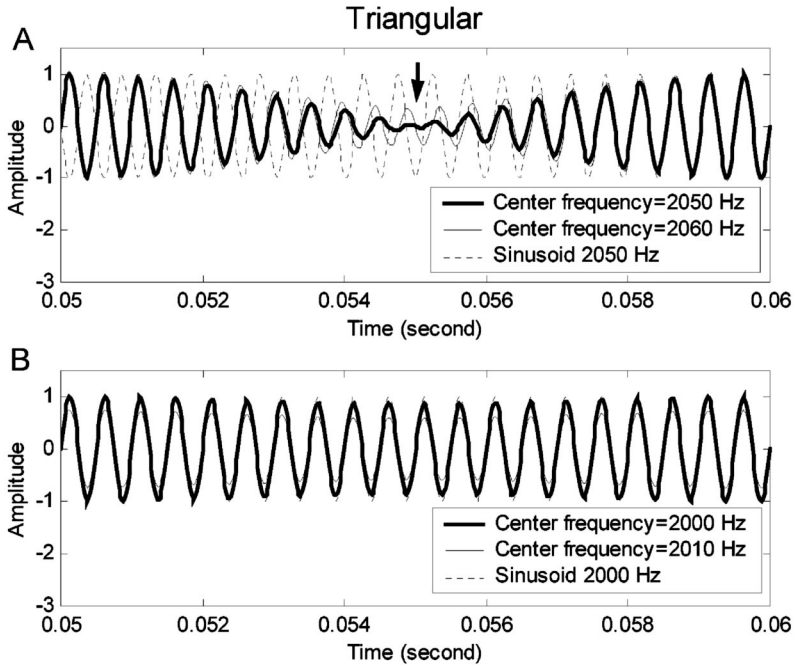


FIG. 4. Time-domain waveforms for the simplified harmonic-complex signals with triangular spectra. (a) Stimulus with center frequency of 2050 Hz; (b) stimulus with center frequency of 2000 Hz. In each panel, the thick solid line illustrates a signal without the center-frequency shift and the thin solid line illustrates a signal with a 10-Hz center-frequency shift. The dotted lines are reference sinusoidal signals of 2050 Hz (a) and 2000 Hz (b).

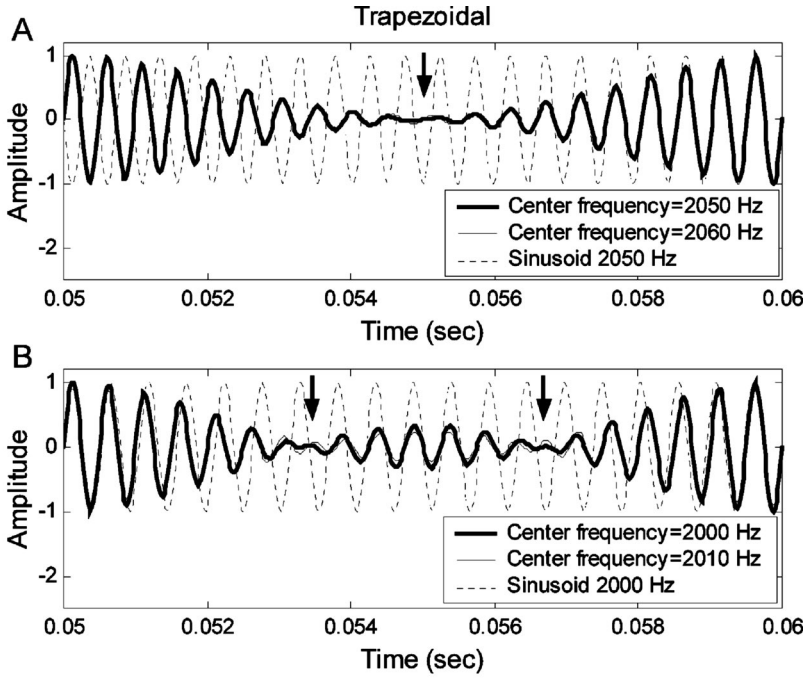


FIG. 5. Time-domain waveforms for the simplified harmonic-complex signals with trapezoidal spectra. (a) Stimulus with center frequency of 2050 Hz; (b) stimulus with center frequency of 2000 Hz. In each panel, the thick solid line illustrates a signal without the center-frequency shift and the thin solid line illustrates a signal with a 10-Hz center-frequency shift. The dotted lines are reference sinusoidal signals of 2050 Hz (a) and 2000 Hz (b).

(shifted 10 Hz from 2050 Hz). The dotted line in Fig. 4(a) is a pure sinusoidal signal at 2050 Hz, inserted to provide a visual reference. By comparing the thick and the thin solid lines with the dotted reference line in Fig. 4(a), the 180-deg phase transition that occurs at zero crossings of the envelope can be observed. On the right side of the marker (the downward arrow), the thick and the thin solid lines have the same phase as the dotted sinusoidal reference signal. On the left side of the marker, the thick and the thin solid lines have a 180-deg phase difference from the dotted reference line. The phase transition in the thick solid line differs slightly from that in the thin solid line. The thin solid line has a relatively slower phase shift; that is, the phase shift in the thin solid line starts earlier and ends later than in the thick solid line. This difference in the phase transient provides information for center-frequency discrimination, assuming that the AN response phase locks to the fine structure of the sound stimulus.

Figure 3(b) shows the simplified triangular spectrum used for a center frequency of 2000 Hz; three harmonic components were kept in this case. Because of the existence of the center component, the combination of these three harmonic components does not have the 180-deg phase change in the time domain. As described below, the presence or absence of this 180-deg phase shift can explain threshold differences between these stimulus conditions.

In Fig. 4(b), the thick solid line is the simplified three-component signal with center frequency at 2000 Hz [corresponding to the spectrum in Fig. 3(b)], and the thin solid line is the simplified signal with center frequency at 2010 Hz (shifted 10 Hz from 2000 Hz). The dotted line is a pure sinusoidal signal (2000 Hz) which is included to provide a visual reference. It is clear that the result of the 10-Hz shift of the triangular spectral envelope is primarily a magnitude change in the time domain.

The same simplification strategy was applied to the stimuli with trapezoidal spectra [Figs. 3(c), (d)], except that a larger number of components was required in the simplified signals for the trapezoidal spectrum. The central four components were kept for the stimulus with center frequency of 2050 Hz [Fig. 3(c)], and the central five components were kept for center frequency of 2000 Hz [Fig. 3(d)]. Figure 5(a) (thick solid line) shows the simplified four-component signals with a center frequency of 2050 Hz [i.e., between two harmonic components, Fig. 3(c)] and 2060 Hz (thin solid line, a 10-Hz deviation from 2050 Hz) in the time domain. The reference sinusoid at 2050 Hz (dotted line) is included to illustrate the 180-deg phase reversals (arrow) in both signals; the time courses of phase reversals differ slightly between the two stimuli. Figure 5(b) shows simplified five-component signals with center frequency at 2000 Hz (thick solid line) and 2010 Hz (thin solid line, with a 10-Hz deviation from 2000 Hz) in the time domain. The arrows in Fig. 5(b) indicate two abrupt 180-deg phase reversals in the thick solid line, whereas the phase reversals in the thin solid line are relatively smooth. Over the same period of time, there are more phase reversals in the five-component stimulus (twice between 0.05 and 0.06 s) than in the four-component stimulus (once between 0.05 and 0.06 s). For the same difference in center frequency (10 Hz), the time over which the phase shift takes place is longer for the stimulus with a center frequency at a harmonic frequency [Figs. 3(d), 5(b)] than for the stimulus with a center frequency between harmonic frequencies [Figs. 3(c), 5(a)]. These differences are potential explanations for the relatively lower discrimination threshold for center frequency at 2000 Hz as compared to 2050 Hz for the trapezoidal spectra [Fig. 2(b); Lyzenga and Horst, 1995]. The simulations below allowed quantification of the information in these differences and direct comparison between predicted thresholds based on a physiological model and actual thresholds.

This description of the stimuli with simplified spectra

was included to facilitate the description and discussion of the possible cues in the stimuli. The threshold predictions shown in the figures below were based on the original signals, unless specifically stated otherwise.

B. Nonlinear AN population model

The simulations of AN responses in this project were based on a nonlinear computational AN model (Tan and Carney, 2003) designed to simulate the time-varying discharge rate of AN fiber responses in cat to arbitrary sound stimuli. This AN model has compression, two-tone suppression, and an instantaneous frequency (IF) glide in its reverse-correlation function (e.g., Carney *et al.*, 1999). This model was selected for this study to allow investigation of the potential contributions to the results of compression and the frequency glide, which interact in a nonlinear fashion. Both the compressive nonlinearity and the IF glide can be "turned on and off" by manipulating the parameters of the model. Threshold predictions based on a model without these features were not significantly different from those reported here; thus, these model features were not critical for the predictions described (see Tan [2003], Chap. 5, for more detail). In addition, the simulations presented here were repeated using another nonlinear AN model (Heinz *et al.*, 2001c), which has sharper tuning that is based on estimates of human auditory filters. Threshold predictions based on the Heinz *et al.* (2001c) model only differed from those presented here in one case (discussed below), despite the sharper tuning of the AN fibers. This result was expected because the information in the rates of high-spontaneous AN fibers to wideband harmonic complexes presented at mid to high levels are not greatly effected by the AN filter bandwidth, nor are the temporal response properties (such as phase locking) that are critical for the temporal representations. Thus, the trends in the threshold predictions presented here were robust across different versions and configurations (e.g., linear versus nonlinear) of the AN model.

The AN model population was based on a subset of the total 30 000 AN fibers in human (Rasmussen, 1940), which were assumed to have characteristic frequencies (CFs) evenly distributed on a log scale from 20 to 20 000 Hz using a simplified version of the human cochlear map of Greenwood (1990). Calculations presented here were based on 50 AN models with CFs evenly distributed on a log scale from 1500 to 3000 Hz (CFs beyond this range were not considered for efficiency in computation), or on subsets of these fibers. The 50 model fibers represented approximately 10% of the 30 000 AN fibers ($[\log(3000/1500)/\log(20\,000/20)] \times 100\% = 10\%$), which corresponded to a subpopulation of 3000 fibers. Thus, each of the 50 AN models represented about 60 AN fibers, for a total of 3000 fibers in the 1500–3000-Hz range.

C. Statistical methods

The predictions of the jnd's in center frequency were made based on the assumption that the observations of the population AN-model response was a set of independent nonstationary Poisson processes (Siebert, 1968). An ideal central processor was assumed to optimally use the information encoded in the response pattern of each AN model fiber, and the threshold of this central processor was estimated. The bound on the variance of the estimate of a variable can be described by the Cramér–Rao bound (Cramér, 1951; van Trees, 1968). The variance σ_i of the estimate of any signal parameter (e.g., F_c, the center frequency of the harmonic complex) based on the observation from the ith AN fiber is bounded by (Siebert, 1965, 1968)

$$\frac{1}{\sigma_i^2} \leqslant \int_0^T \frac{1}{r_i(t)} \left[\frac{\partial r_i(t)}{\partial F_c} \right]^2 dt, \tag{2}$$

where $r_i(t)$ is the ith AN fiber's instantaneous discharge rate, T is the duration of the stimulus, and F_c is the center frequency of the harmonic complex. Note that Siebert's strategy for estimating the jnd is based on descriptions of the instantaneous rate, $r_i(t)$, for each model AN fiber in response to each stimulus. Simulations of individual AN discharge times are not required; the randomness of AN responses from trial to trial is incorporated in the assumption that the AN responses are Poisson in nature.

Equation (2) represents the normalized sensitivity of the ith AN fiber to a change in the center frequency change of the signal. By assuming that the discharge patterns of all AN fibers are statistically independent (Johnson and Kiang, 1976), the bound of the variance of the observation based on the AN population's response pattern can be found by summing the bounds for each single AN fiber; i.e., $1/\sigma_{\text{all}}^2 = \Sigma_i 1/\sigma_i^2$. The jnd of the ideal central processor corresponding to $d' = F_{c\text{jnd}}/\sqrt{\sigma_{\text{all}}^2} = 1$ can then be found (Siebert, 1965) as follows:

$$F_{c\text{jnd}} = \left[\frac{1}{\sum_i \frac{1}{\sigma_i^2}} \right]^{1/2} = \left[\frac{1}{\sum_i \int_0^T \frac{1}{r_i(t)} \left[\frac{\partial r_i(t)}{\partial F_c} \right]^2 dt} \right]^{1/2}. \tag{3}$$

Equation (3) describes the jnd of an ideal processor that uses both rate and timing information (i.e., "all information," Heinz *et al.*, 2001a). If only the average-rate information of the AN model responses is used, Eq. (3) can be simplified to

$$F_{c\text{jnd}} = \left[\frac{1}{\sum_i \frac{1}{\sigma_i^2}} \right]^{1/2} = \left[\frac{1}{\sum_i \frac{1}{Y_i} \left[\frac{\partial Y_i}{\partial F_c} \right]^2} \right]^{1/2}, \tag{4}$$

where $Y_i = \int_0^T r_i(t)dt$ is the expected number of spikes (representing the average-rate information) from the ith model AN fiber in one trial.

The calculation of the partial derivative was approximated by calculating the ratio between the change in the response due to a small change in the center frequency of the signal and the small change in the center frequency (e.g., Heinz *et al.*, 2001a)

$$\frac{\partial r_i(t)}{\partial F_c} \cong \frac{r_i(t|F_c + \Delta F_c) - r_i(t|F_c)}{\Delta F_c}. \tag{5}$$

In this study, the approximation of the partial derivative was computed using $\Delta F_c = 1$ Hz.

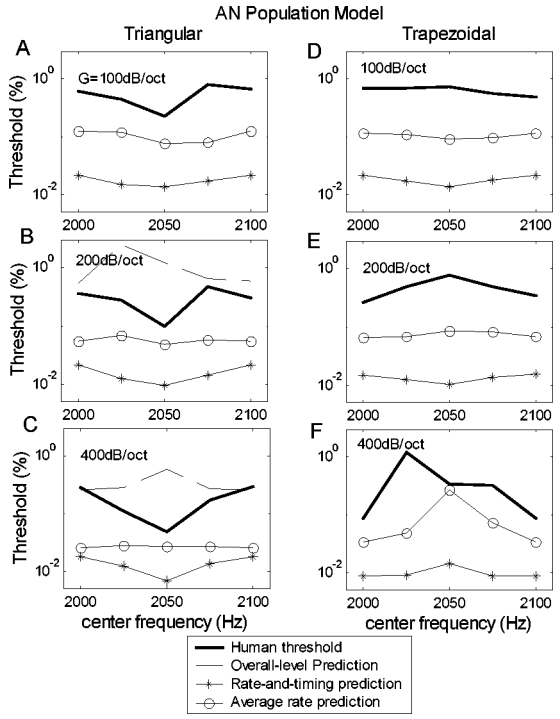


FIG. 6. Thresholds of the AN population model for discrimination of the center frequency of harmonic-complex signals with triangular spectra (a), (b), (c) and trapezoidal spectra (d), (e), (f). Each panel corresponds to one slope of the spectrum envelope ($G = 100$, 200, and 400 dB/oct from top to bottom). The lines with circles are predictions based on only average-rate information of the AN population responses. The lines with asterisks are predictions based on both rate and temporal information. The dashed lines (b), (c) are threshold predictions based on the change in the overall level of the stimuli with triangular spectra; these predictions clearly have incorrect trends in threshold as a function of center frequency as compared to the psychophysical results (Lyzenga and Horst, 1995). The predictions for triangular spectra based on the combination of rate and timing information (asterisks) showed the desired trends (i.e., lowest thresholds for center frequencies between harmonics) for all three spectral slopes.

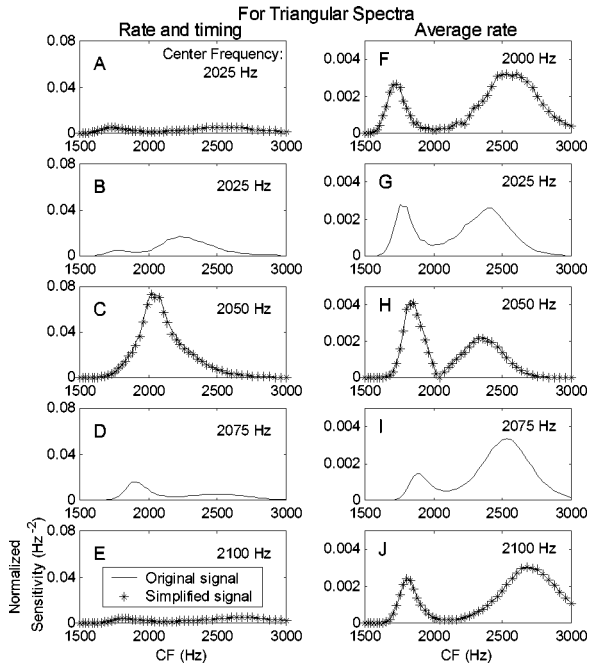


FIG. 7. Sensitivity of model fibers to the changes of the triangular harmonic-complex center frequency as a function of model fiber CF. The left column (a)–(e) shows predictions based on rate information; the right column (f)–(j) shows predictions based on both rate and timing information. Each row corresponds to one harmonic-complex center frequency (2000 to 2100 Hz with a step size of 25 Hz). The lines are based on the original stimuli (for all center frequencies) and the asterisks are based on the simplified signals (shown for 2000, 2050, and 2100 Hz only).

The computer programs used for the simulations presented here are available at http://web.syr.edu/~lacarney

III. RESULTS

A. Predictions for signals with triangular spectra

Figures 6(a)–(c) shows the predictions of the AN population model thresholds for center-frequency discrimination of the harmonic complexes with triangular spectra. Center-frequency discrimination thresholds (jnd's) are plotted as a function of the center frequency of the spectral envelope. Each panel corresponds to predictions for one value of the spectral slope; psychophysical thresholds from Lyzenga and Horst (1995) are replotted (thick lines), along with their predictions based on changes in overall stimulus level [Figs. 6(b), (c) dashed lines]. Model predictions were based on either the combined rate and timing information of the AN model population response (asterisks) or only on rate information (circles). The predictions based on the combination of rate and timing information for all three spectral slopes showed the general trend of those observed for human listeners. That is, rate-and-timing-based thresholds plotted as a function of center frequency showed a ''trough,'' or were lowest when the center frequency was between two harmonics (2050 Hz) and highest when the center frequency fell on a harmonic (2000 and 2100 Hz). Model threshold predictions based only on rate information were relatively flat as a function of center frequency (circles). In contrast, Lyzenga and Horst's (1995) prediction based on the level of the signals (dashed lines)[1] showed a *peak* in the thresholds at 2050 Hz.

Rate-based predictions for an AN model with sharper tuning (Heinz *et al.*, 2001c) also had a small peak at 2050 Hz (not shown). The sharper tuning in the Heinz *et al.* AN model enhanced the energy-based information in the stimulus, which resulted in predictions that had the wrong trend (i.e., a peak rather than a trough in threshold plotted as a function of center frequency.) Predictions for the rate-and-timing-based predictions of the model with sharper tuning had trends that agreed with the human data (i.e., a trough rather than a peak in the predicted thresholds as a function of center frequency) and were not significantly different from predictions for the Tan and Carney (2003) model used in this study. This was expected because the temporal response properties of AN fibers are not strongly affected by reasonable differences in bandwidth of tuning for this type of stimulus.

Explanations for the trends in the model thresholds are provided by examining the sensitivity of different members of the model AN population. Figure 7 shows the normalized sensitivity[2] (in units of $1/\text{Hz}^2$) as a function of model-AN CF for the triangular spectrum with a slope of 400 dB/oct. Each row corresponds to one spectral-envelope center frequency. The normalized sensitivity for each AN model fiber based on both rate and timing information [left column, Eq. (2)] is

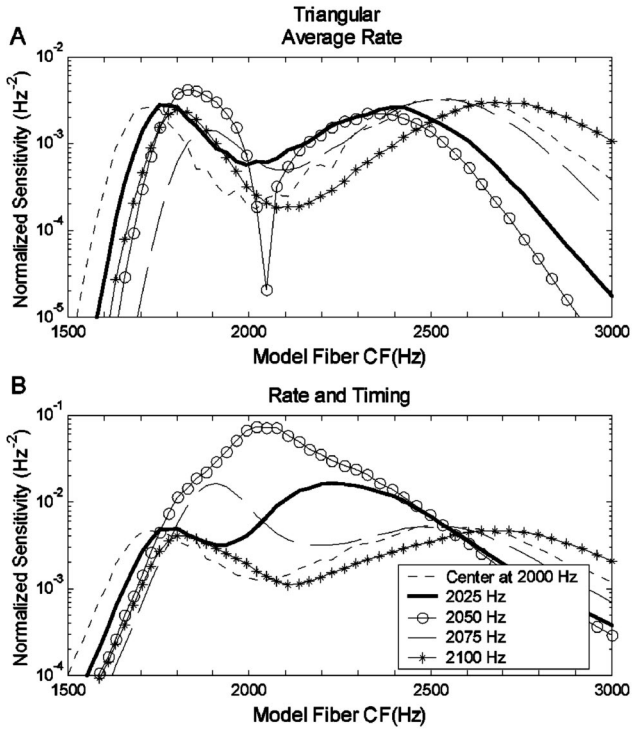


FIG. 8. (a) Normalized sensitivity patterns for triangular spectra with a slope of 400 dB/oct at various center frequencies based on average-rate information of AN model responses. (b) Normalized sensitivity patterns for triangular spectra at various center frequencies based on rate and timing information of AN model responses. This figure illustrates how information used by the model to discriminate the triangular spectrum stimuli is distributed across fibers with different CFs. Rate-only and rate-plus-timing information is distributed differently, especially for fibers tuned near the center frequency of the spectrum.

defined as $\int_0^T [1/r_i(t)][[\partial r_i(t)]/\partial F_c]^2 dt$. The normalized sensitivity based on average-rate information (right column) is defined as $1/Y_i[\partial Y_i/\partial F_c]^2$ [see Eq. (4)]. Predictions based only on average rate ignore timing information; therefore, as expected, the normalized sensitivity based on average rate is always lower than the normalized sensitivity based on both rate and timing information (note the different ordinate scales in Fig. 7).

The solid lines in all ten panels of Fig. 7 were computed with the stimuli that were used in the psychophysical study (Fig. 1); the asterisks are results based on the simplified harmonic complex signals (Fig. 3). The results for the simplified signals were nearly identical to those for the original signals, suggesting that the simplified signals contained the cues that dominated the predicted thresholds. To compare the results across different center frequencies, the sensitivity profiles of the AN population are plotted together in Fig. 8(a) (rate-based sensitivity) and Fig. 8(b) (rate and timing). There is always a drop in sensitivity based on average-rate information for fibers with CFs near 2050 Hz [Fig. 8(a)]. In contrast, the sensitivity based on rate plus temporal information of fibers with CFs near 2050 Hz varies depending upon the stimulus center frequency [Fig. 8(b)]; these fibers have relatively low sensitivities for some center frequencies yet are the most sensitive fibers in the population for other center frequencies.

When only rate information is used, the overall sensitivities (combined across CFs) for center frequency at 2000 Hz [Fig. 8(a); dotted line] and 2100 Hz (asterisks) are approximately the same as that for center frequency of 2050 Hz (circles). Thus, the threshold based only on rate information is relatively flat as a function of center frequency [cf. Figs. 6(a)–(c)]. When both rate and timing information are included, the overall sensitivity for the stimulus with center frequency at a harmonic frequency (2000 or 2100 Hz) is lower than the overall sensitivity for center frequency at 2050 Hz, where a peak is observed in the sensitivity pattern (circles). Thus, the population threshold based on both rate and timing information is higher for center frequencies of 2000 and 2100 Hz than for 2050 Hz [cf. Figs. 6(a)–(c)].

The general trends for the AN population model predictions based on rate and timing information qualitatively match those in the psychophysical results [Figs. 6(a)–(c)]. That is, the presence of peaks or troughs in the predictions is in agreement with the experimental results. However, there are more detailed trends in the psychophysical results that were explored further using the responses of subpopulations of AN fibers. For example, human thresholds are not symmetric around the lowest point on the threshold versus center frequency curve [Figs. 6(a)–(c), thick lines]; thresholds at 2025 Hz are always lower than those at 2075 Hz. The population model results based on both rate and timing information [Figs. 6(a)–(c), asterisks] show a slight trend that agrees with this aspect of the psychophysical data. Figure 8(b) shows a substantial difference in model sensitivity profiles between results for stimulus center frequencies of 2025 and 2075 Hz. This difference is effectively reduced in the overall population sensitivity due to the presence of the sidebands in the profiles. These profiles suggested that predictions based on a smaller population of AN model fibers, centered at about 2050 Hz, would have a larger difference in threshold across center frequencies. Predictions based on a restricted population of AN fibers are also interesting to consider because it is reasonable to assume that the brain encodes information in a specific frequency region (i.e., near one formant frequency) based on information from a subset of AN fibers rather than from the entire population.

Thresholds for different subsets of model AN fibers are further explored in Fig. 9. Predictions based on model fibers with CFs between 1500 and 3000 Hz (asterisks) are compared to those for CFs limited to 1900–2200 Hz (squares). Predictions are also shown for two single-fiber models: in one case, the model fiber used for each center frequency was the CF with the highest sensitivity to changes in that center frequency (circles). The other single-fiber model was based on the response of the model AN fibers with CF at 2106 Hz (diamonds). This CF was chosen as the member of the logarithmically spaced population that showed trends in sensitivity, when both rate and timing information were used, that most closely matched the trends in the psychophysical results. Predicted thresholds for the neighboring fiber in the AN population (CF=2077 Hz) were very similar (not shown), indicating that the predicted thresholds were not highly sensitive to the precise choice of CF.

Figures 9(a)–(c) show results based only on rate infor-

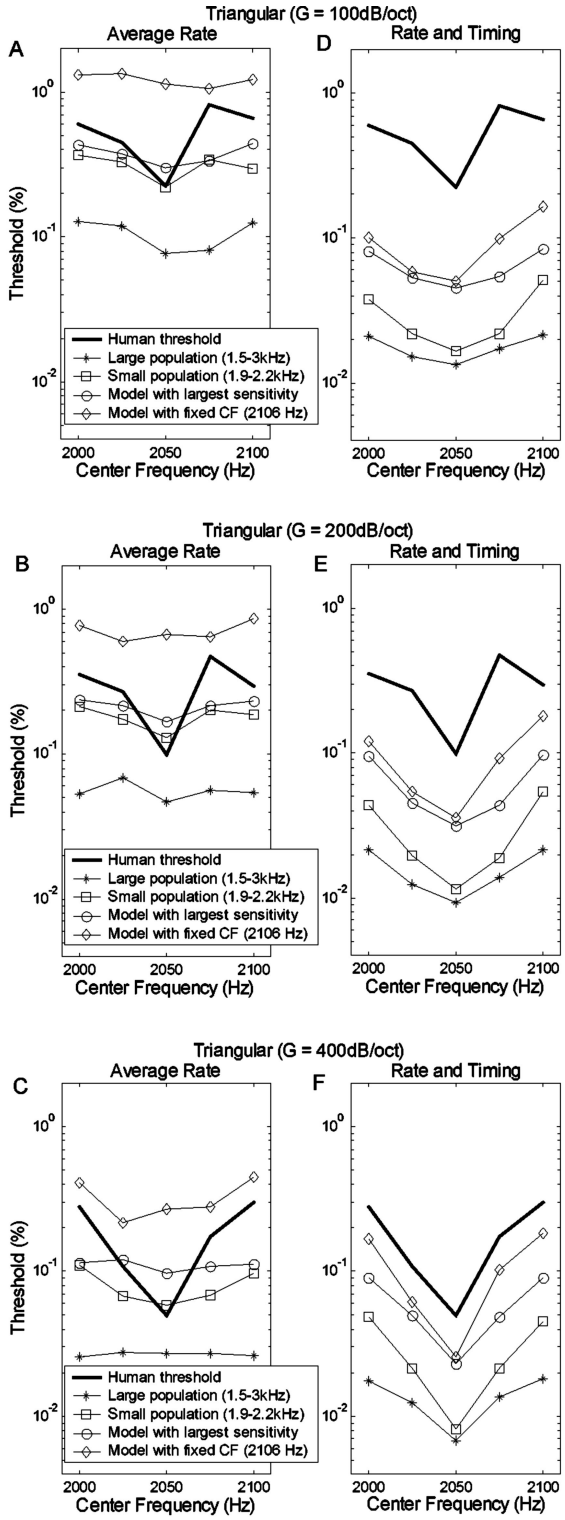


FIG. 9. Thresholds for triangular spectra with various subsets of AN model fibers based on average-rate information (a)–(c) or both rate and timing information (d)–(f). Each row corresponds to spectral envelopes with a given slope ($G=100$, 200, and 400 dB/oct from top to bottom). Human thresholds (Lyzenga and Horst, 1995) are shown as the solid line with no symbols. The selections of the CFs of AN model fibers for the small-population predictions are distinguished by different symbols. The line with circles is based on the model fiber with the highest sensitivity at each center frequency. (The selection of this best model fiber could change for different center frequencies.) The line with diamonds is based on a single AN model fiber with CF equal to 2106 Hz.

mation. The predictions based on the small population with CFs between 1900 and 2200 Hz (marked by squares) show the correct trend (i.e., lowest threshold at 2050 Hz and highest thresholds at 2000 and 2100 Hz). However, the threshold difference in this prediction (e.g., about a factor of 2 for $G=400$ dB/oct.) is much smaller than the difference in the psychophysical results (e.g., about a factor of 10 for $G=400$ dB/oct.). The two single-fiber model predictions based only on the rate information fail to predict the general trend of the psychophysical results.

Figures 9(d)–(f) show predictions based on both rate and timing information. The predictions based only on the set of model fibers with CF at 2106 Hz (diamonds) show a shape that is most similar to the detailed trends seen in the psychophysical results, with the lowest threshold at 2050 Hz, the lowest threshold and the highest threshold differing by approximately a factor of 10 (for $G=400$ dB/oct), and an asymmetrical threshold function (the threshold at 2075 Hz is higher than the threshold at 2025 Hz). The other predictions also show trends similar to those in the psychophysical data; however, they either do not have the asymmetry (squares and circles) or they have a relatively small difference between the lowest and highest thresholds (asterisks) as compared to the psychophysical results.

B. Predictions for signals with trapezoidal spectra

Figures 6(d)–(f) compare model predictions with psychophysical results for center-frequency discrimination of stimuli with trapezoidal spectral envelopes of different slopes. The experiments using trapezoidal stimuli (Lyzenga and Horst, 1995) included a center-frequency range equal to two times the fundamental frequency, with the thresholds showing the same patterns in the frequency range from 2100 to 2200 Hz as in the range from 2000 to 2100 Hz. The model predictions are illustrated only for the range from 2000 to 2100 Hz; by illustrating this frequency range, the contrast between the results for the triangular and trapezoidal spectra is clearer. The changes in threshold across center frequency are relatively small for trapezoidal stimuli that have low spectral slopes [$G=100$ and 200 dB/oct, Figs. 10(a), (b), (d), (e)]. For these slope conditions, predictions based on both rate-alone and rate-and-timing are also relatively flat as a function of center frequency; neither model captures the small changes in threshold across center frequency for these slope conditions. For the condition that resulted in relatively large changes in threshold at different center frequencies [$G=400$ dB/oct, Figs. 10(c), (f)] both the rate-based and rate-and-timing-based predictions have the general trends seen in the psychophysical results, with the highest thresholds for center frequency 2050 Hz and lower thresholds for center frequencies of 2000 and 2100 Hz. As was the case for the triangular spectra (Fig. 7), predictions based on the simplified versions of the trapezoidal spectra were similar to those for the complete stimuli (not shown), suggesting that the cues contained in the simplified stimuli were responsible for the model thresholds.

The results for the stimuli with 400-dB/oct slopes [Figs. 10(c), (f)] were further examined by again looking at profiles of sensitivity versus model AN CF (Fig. 11). Figure 11(a)

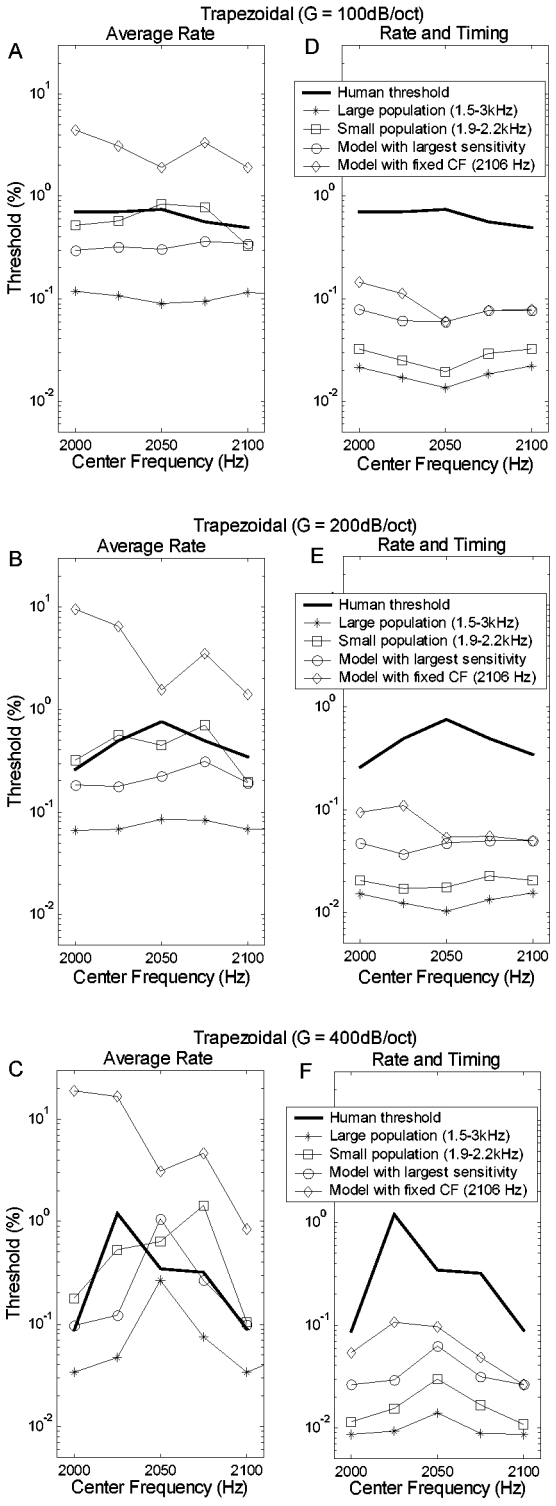


FIG. 10. Thresholds for the trapezoidal spectra based on average rate information only (a)–(c) and based on both rate and timing information (d)–(f). Each row corresponds to spectral envelopes with a given slope ($G = 100$, 200, and 400 dB/oct from top to bottom). The thick solid lines illustrate human thresholds. Different symbols distinguish the model predictions with different selections of model CF range.

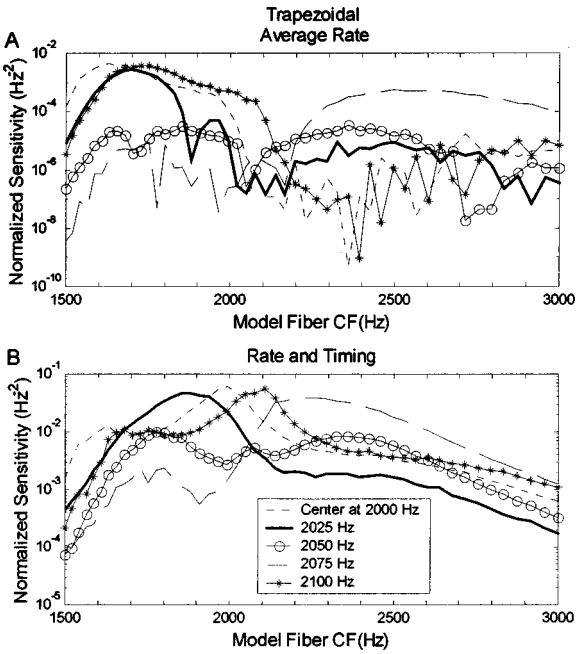


FIG. 11. Sensitivity patterns as a function of model-fiber CF for the trapezoidal spectra with a spectral slope of 400 dB/oct. (a) Sensitivity based on only average-rate information and (b) based on both rate and timing information. Different harmonic-complex center frequencies are distinguished by different symbols. This figure illustrates how the information used by the model to discriminate changes in center frequency of stimuli with trapezoidal spectra is distributed across fibers tuned to different frequencies. As was true for triangular spectra (Fig. 8), the differences between the rate and rate-and-timing based models is largest for fibers with CFs near the center frequency of the stimulus spectrum.

shows sensitivities based on the average-rate information; the integral of the sensitivity over model CF for the stimulus with center frequency of 2050 Hz [Fig. 11(a), line with circles] is lower than for center frequencies of 2000 (dashed line) and 2100 Hz (line with asterisks). Therefore, the threshold at 2050 Hz is highest based on the rate information of the large population of AN model fibers [CFs of 1500–3000 Hz; asterisks in Fig. 10(c)].

Figure 11(b) shows the sensitivity patterns based on both rate and timing information; the overall sensitivity for the stimulus with center frequency of 2050 Hz is lower than that for the other center frequencies, and thus the highest threshold appears at 2050 Hz for the prediction based on all the model fibers with CFs between 1500–3000 Hz in Fig. 10(f) (asterisks).

Figure 10 also shows predictions based on smaller AN populations. For both the rate-only and the rate-and-timing predictions for the 400-dB/oct slope condition, the trends of the prediction based on the model fibers with CF from 1900–2200 Hz (squares) and the prediction based on the single-model CF with highest sensitivity for each stimulus (circles) were most similar to the general trend of the prediction based on the larger population of model fibers [Figs. 10(c), (f)]. The thresholds based on a single-model fiber with a CF equal to 2106 Hz was also calculated (diamonds). For the rate-only prediction, the trend of this prediction is wrong (i.e., the highest threshold occurs at 2000 Hz). The prediction based on a single CF at 2106 Hz, using both rate and timing information, is the best match to the trends in the psychophysical results [Fig. 10(b)], including the asymmetry in thresholds across stimulus center frequency. It is interesting that the same CF channel (the model fiber with CF equal to 2106 Hz) resulted in the best match to psychophysical results for both

the triangular and trapezoidal spectra for the rate-and-timing-based predictions.

IV. DISCUSSION

In this study, harmonic-complex frequency-discrimination experiments were simulated with a computational AN model, and the thresholds of an optimal detector for the frequency-discrimination tasks were evaluated. The model performance was quantified using only average-rate information or using both rate and timing information.

Lyzenga and Horst (1995) showed predictions based on the overall level change in the stimuli for the harmonic-complex frequency discrimination. Their threshold predictions based on stimulus level and the threshold predictions here based only on model AN rate information both disagreed with the trends in human thresholds for harmonic-frequency discrimination. Predictions based on combined rate and timing information generally agreed with the trends in psychophysical thresholds.

A method of simplifying the harmonic-complex spectrum was useful for identifying potential timing cues encoded in the harmonic complexes. The simplified signals had phase-transition cues that qualitatively explained the general trends in the thresholds. For the triangular spectrum, when the center frequency (2050 Hz) was between two harmonic components, the speed of the 180-deg phase transition provided timing information that distinguished this stimulus from one with a center frequency at a harmonic component (2000 or 2100 Hz). For the trapezoidal spectrum, the phase transients occurred more often in stimuli with center frequency at 2000 or 2100 Hz than in the stimulus with center frequency at 2050 Hz. The rate-and-timing predictions apparently take advantage of this phase-transition cue and show the same trends as in human thresholds, for both the simplified stimuli and for the full harmonic complexes.

Figure 12 illustrates the representation of the phase-transition cue in the response of a model AN fiber with a CF of 2106 Hz, which was the fiber used for the single-channel model predictions [Figs. 9, 10, diamonds]. In Fig. 12(a), the responses of this AN model fiber to harmonic complexes (triangular spectrum, $G=400$ dB/oct) with center frequencies of 2050 Hz (thick line) and 2060 Hz (thin line) are compared to a sinusoid signal (dashed line). The 180-deg phase reversal that was illustrated for the simplified stimulus [Fig. 4(a)] was also observed in the responses of the model AN fiber. Figure 12 shows the normalized changes in the response of the AN model fiber due to a 10-Hz center-frequency change in the harmonic complex; i.e., the difference in the thin and the thick solid lines in Fig. 12(a) normalized by the thick solid line

$$R_{\mathrm{diff}}(t)=\frac{1}{r_{\mathrm{CF}=2106}(t|f=2050)}\times\left[\frac{r_{\mathrm{CF}=2106}(t|f=2060)-r_{\mathrm{CF}=2106}(t|f=2050)}{2060-2050}\right]^2. \quad (6)$$

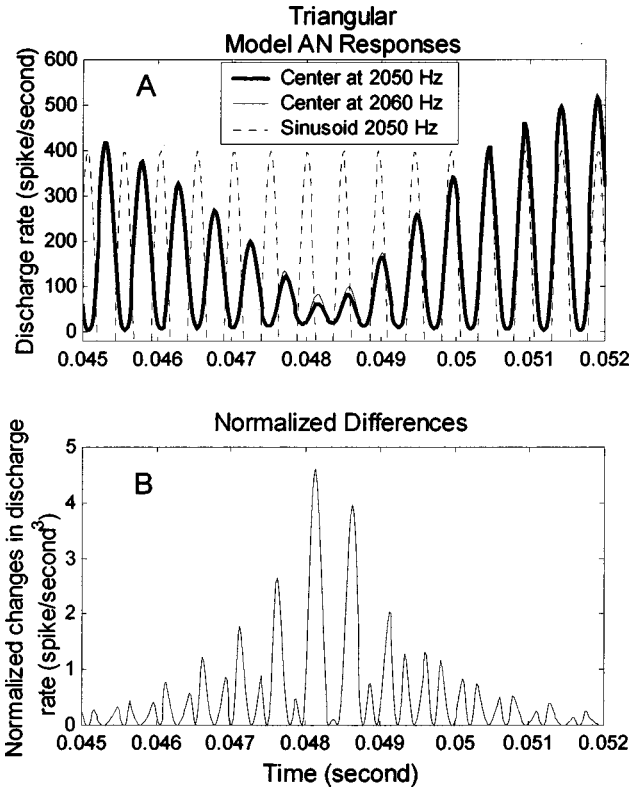


FIG. 12. The phase transition in the triangular spectrum stimulus [see Fig. 4(a)] is preserved in the response of an AN model fiber. In panel (a), the dashed line is a reference signal (sinusoid, 2050 Hz). The thick and the thin solid lines are responses of an AN model fiber (CF=2106 Hz) to harmonic complexes with triangular spectra with slopes of 400 dB/oct. and center frequencies at 2050 and 2060 Hz, respectively. Panel (b) shows the difference between the responses to the harmonic complexes with and without the 10-Hz center-frequency change normalized by the response to the harmonic complex without the 10-Hz frequency shift. The largest differences in the AN model response between stimuli with different center frequencies occur near the time of the phase transition, when the stimulus envelope has the smallest amplitude. See the text for more detail.

$R_{\mathrm{diff}}(t)$ illustrates the change in sensitivity as a function of time to the 10-Hz shift in center frequency due to this model AN fiber's response. The integral of $R_{\mathrm{diff}}(t)$ over time is the normalized sensitivity of this AN model fiber to the center frequency change. Figure 12(b) shows that $R_{\mathrm{diff}}(t)$ has relatively high values during the 180-deg phase reversal. This observation, along with the results in Figs. 9 and 10, supports the suggestion that the phase transitions provide temporal information that is consistent with the sensitivity of listeners in the harmonic-complex center-frequency discrimination task.

This study showed that fibers from a single-frequency channel provided the best prediction of the trends of threshold across center frequency for both the triangular spectrum and the trapezoidal spectrum for the slope condition that resulted in the most significant threshold changes. This single-frequency channel is on the high-frequency side of the harmonic-complex envelope at the lowest center frequency. This is in agreement with a suggestion of Van Zanten (1980) that the temporal modulation transfer function (TMTF) for noise stimuli with various bandwidths and center frequencies is governed by the signal contents within the highest frequency bands of the stimuli.

Model results based on a small number of AN fibers with CFs near the signal frequency are intuitively more realistic because these models require less neural processing than models based on large populations of AN responses, such as spread-of-excitation models. If the stimuli are narrow-band signals, most of the AN fibers outside the small population centered at the target frequency generally would have reduced sensitivity to changes in the stimulus, especially at low sound-pressure levels. If the stimuli are broadband signals (such as speech, or narrow-band signals in the presence of noise), the responses of AN fibers outside the small population are likely to be dominated by stimulus components other than the target component. In addition, when the task in a psychophysical experiment is to discriminate changes of more than one frequency component (e.g., at several formant frequencies), it is reasonable to assume that the auditory system discriminates the change at each formant frequency based on the information from AN fibers tuned near that formant frequency. Previous models based on temporal information in the form of first-order intervals also concluded that a suboptimal model based on relatively few AN fiber responses over limited time windows provided better predictions of psychophysical data for frequency discrimination of pure tones than did a model based on the complete responses of a large population of fibers (Goldstein and Srulovicz, 1977; Srulovicz and Goldstein, 1983).

Our predictions based on rate and timing in the responses of a few model fibers had trends more similar to those in human thresholds than did predictions based on a larger population for the triangular spectrum [Figs. 9(d)–(f)], but the improvement was not as significant for the trapezoidal spectrum [Figs. 10(d)–(f)]. The reason for this difference may be that the trapezoidal spectrum had a 200-Hz plateau and thus had a larger bandwidth than the triangular spectrum. More AN fibers are likely to be involved in the discrimination task for this type of spectrum.

The assumption of an ideal central processor is not physiologically realistic, as it requires the central nervous system to have a perfect memory for the response patterns to each stimulus. A more realistic temporal processing strategy, across-CF coincidence detection, was also investigated (Tan, 2003). Across-CF coincidence detection did not effectively explain trends in thresholds. Across-CF coincidence detection is most effective for across-channel temporal cues (e.g., Heinz *et al.*, 2001b) and would not be expected to be effective for within-channel temporal cues, such as the phase-transition cues in the harmonic complexes (Figs. 4, 5). Other mechanisms for extracting temporal cues, such as tuning in the modulation-frequency domain or interval-based codes, should be further explored in future studies. A recent physiologically based model for extraction of envelope cues provides one possible mechanism that should be tested in future studies for its potential to explain the psychophysical results studied here (Nelson and Carney, 2004).

Included in Lyzenga and Horst's study (1995) were the first steps in a series of studies that examined models to test the adequacy of various cues and decoding mechanisms to explain their data. These models included a profile comparison model and one based on amplitude-modulation detection thresholds. They concluded that an excitation-profile comparison model, which roughly predicted the lowest threshold for both spectra shapes, best predicted their data. However, their explanation required the assumption that the excitation difference included only negative values for responses to the trapezoidal envelope, whereas both positive and negative values had to be included for the triangular envelope (Fig. 10 of Lyzenga and Horst, 1995). It is interesting that they were able to explain their data with this model; however, it is difficult to envision a simple, physiologically based model designed to explain results for both types of stimuli that would respond as their profile-comparison model requires. They also suggested that the results for the triangular spectrum can be partially explained by sensitivity to amplitude-modulation depth (Fig. 9 of Lyzenga and Horst, 1995); however, this theory cannot explain why the threshold trends differ as a function of center frequency between the triangular and trapezoidal spectra.

Lyzenga and Horst (1997) extended their earlier study and concluded that phase cues influence the threshold in the frequency region near 2000 Hz. They observed that when the fundamental frequency is 100 Hz, three harmonic components fall into one critical band (roughly 250-Hz wide) and thus the excitation-profile model cannot explain the data because it is insensitive to the relative phase relations of the harmonic components. They also calculated the envelope-weighted or intensity-weighted averaged instantaneous frequency (EWAIF or IWAIF; Feth, 1974), and concluded that EWAIF and IWAIF showed little correspondence with the psychophysical data. Additionally, Lyzenga and Horst (1997) pointed out that the occurrence of peaks in the second-order derivative of the triangular-spectrum signal's temporal envelope clearly depended on the center frequency of the harmonic complex, as well as on the phase relation of the harmonic components. These results indicate the potential importance of temporal cues in explaining discrimination thresholds.

As an extension of Lyzenga and Horst's (1997) analysis of envelope-based cues, the *unweighted* change in the averaged instantaneous frequency (AIF) was calculated for the harmonic complexes

$$\Delta\mathrm{AIF}=\left|\int_0^T f_1(t)dt-\int_0^T f_2(t)dt\right|, \tag{7}$$

where T is the duration of the stimulus, $f_1(t)$ is the instantaneous frequency of the stimulus at a particular center frequency (2000, 2025, 2050, 2075, or 2100 Hz), and $f_2(t)$ is the instantaneous frequency of the stimulus at a center frequency that has a 10-Hz shift from the center frequency for $f_1(t)$. If the auditory system used ΔAIF to decode the center-frequency change, then the size of ΔAIF should be proportional to the relative sensitivity of the auditory system to the center-frequency change, and the reciprocal of ΔAIF should be proportional to threshold. The ΔAIF and its reciprocal are shown in Fig. 13. The reciprocal of ΔAIF shows general trends as a function of center frequency that are similar to human thresholds, and thus the changes in the mean value of the instantaneous frequency could roughly account for the trends of the performance. Because ΔAIF is defined as the

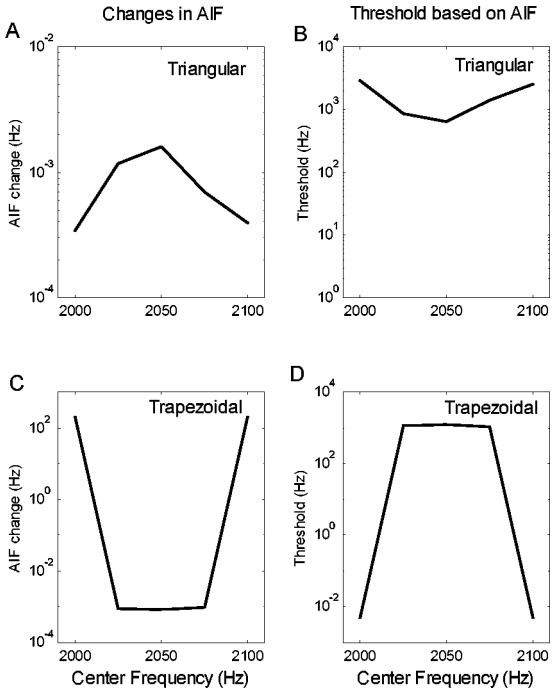


FIG. 13. Predictions of center-frequency discrimination results for the triangular spectrum (a), (b) and the trapezoidal spectrum (c), (d) with $G = 400$ dB/oct based on the unweighted averaged instantaneous frequency (AIF). The left column (a), (c) shows ΔAIF as a function of center frequency. The right column (b), (d) shows the reciprocal of ΔAIF as a function of the center frequency (solid line) and human thresholds (dashed line).

instantaneous-frequency difference between two stimuli averaged over the duration of the stimulus, it is ''averaged timing information'' and thus is a suboptimal decoding mechanism for timing information.

As described above, the AN model discharge rate, $r_i(t)$, preserved the 180-deg phase transition (Fig. 12), which is closely related to the signal's instantaneous frequency. Thus, the predicted threshold trend might be able to match that of human thresholds better if a decoding mechanism that can extract this ΔAIF information from the AN model output is adopted.

The results above also suggest that the timing information related to phase transitions was greatest near the minima of the signal envelope [Fig. 12(b)]. However, both EWAIF and IWAIF assign the greatest weights to the instantaneous frequencies at the maxima of the signal envelope. This difference explains why EWAIF and IWAIF do not account for the center-frequency discrimination results, whereas the unweighted AIF better replicates the general trends in the results.

The simulations of AN fiber responses in this study were all based on a nonlinear AN model (Tan and Carney, 2003). Only AN fibers with high spontaneous rate were considered, and the parameters of the AN model were based on physiological data of cat and gerbil. This peripheral model probably does not provide an accurate representation of the response properties of human AN fibers. However, replacing the peripheral filters with more sharply tuned fibers [e.g., using the AN model of Heinz *et al.* (2001c) which has filters based on Glasberg and Moore's (1990) estimates of auditory filters] did not change the trends illustrated in any of the rate-and-timing predictions and only introduced a slight elevation in threshold for triangular spectra with center frequency of 2050 Hz, which worsened the agreement between model and psychophysical results. In addition, the trends in the results presented here were not affected by either the compressive nonlinearity or the glide in the instantaneous frequency of the impulse response that are included in the AN model used in this study (Tan, 2003).

One goal of this study was to improve our understanding of speech processing in the auditory periphery. The harmonic complex is a convenient, but highly simplified version of a vowel signal. Future work should pursue quantitative studies of neural coding with stimuli more similar to natural speech.

ACKNOWLEDGMENTS

We acknowledge helpful comments from and discussions with Dr. Steve Colburn, Dr. David Mountain, Dr. Allyn Hubbard, Dr. Barbara Shinn-Cunningham, Dr. Oded Ghitza, Paul Nelson, and two anonymous reviewers. Susan Early provided editorial assistance. This work was supported by NIH-NIDCD R01 01641.

[1]Absolute thresholds for Lyzenga and Horst's (1995) level-based predictions were based on a level jnd of 1.5 dB, which was empirically estimated as part of their study. The thresholds of the ideal processor used in this study are lower because they were not limited by an independently imposed jnd, but rather they were limited by the ideal processor's sensitivity.

[2]The squared normalized sensitivity [i.e. $(\delta')^2$ from Heinz *et al.*, 2001a] is convenient to use in illustrating the sensitivity of a population of fibers to changes in a stimulus parameter. The squared sensitivity of a population of independent fibers with different CFs is simply the sum of the individual fibers' squared sensitivities, i.e. $(\delta')^2$ can be handled similar to $(d')^2$; thus, the use of this metric allows one to visually estimate the sensitivity of the population by ''integrating'' across the entire population, or across subsets of the population. The normalized sensitivity is defined as sensitivity per unit frequency; therefore, the squared sensitivity has units of $1/\text{Hz}^2$.

Carney, L. H., McDuffy, M. J., and Shekhter, I. (**1999**). ''Frequency glides in the impulse responses of auditory-nerve fibers,'' J. Acoust. Soc. Am. **105**, 2384–2391.

Cramér, H. (**1951**). *Mathematical Methods of Statistics* (Princeton University Press, Princeton, NJ), Chap. 32.

Feth, L. L. (**1974**). ''Frequency discrimination of complex periodic tones,'' Percept. Psychophys. **15**, 375–379.

Flanagan, J. L. (**1955**). ''A difference limen for vowel formant frequency,'' J. Acoust. Soc. Am. **27**, 613–617.

Glasberg, B. R., and Moore, B. C. J. (**1990**). ''Derivation of auditory filter shapes from notched-noise data,'' Hear. Res. **47**, 103–138.

Goldstein, J. L., and Srulovicz, P. (**1977**). ''Auditory-nerve spike intervals as an adequate basis for aural spectrum analysis,'' in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, New York), pp. 337–347.

Greenwood, D. D. (**1990**). ''A cochlear frequency-position function for several species—29 years later,'' J. Acoust. Soc. Am. **87**, 2592–2605.

Heinz, M. G., Colburn, H. S., and Carney, L. H. (**2001a**). ''Evaluating auditory performance limits. I. One-parameter discrimination using a computational model for the auditory nerve,'' Neural Comput. **13**, 2273–2316.

Heinz, M. G., Colburn, H. S., and Carney, L. H. (**2001b**). ''Rate and timing cues associated with the cochlear amplifier: Level discrimination based on monaural cross-frequency coincidence detection,'' J. Acoust. Soc. Am. **110**, 2065–2084.

Heinz, M. G., Zhang, X., Bruce, I. C., and Carney, L. H. (**2001c**). ''Auditory-nerve model for predicting performance limits of normal and impaired listeners,'' J. Assoc. Res. Otolaryngol **2**, 91–96.

Johnson, D. H., and Kiang, N. Y. S. (**1976**). ''Analysis of discharges recorded simultaneously from pairs of auditory-nerve fibers,'' Biophys. J. **16**, 719–734.

Kewley-Port, D., and Watson, C. S. (**1994**). ''Formant-frequency discrimination for isolated English vowels,'' J. Acoust. Soc. Am. **95**, 485–496.
Lyzenga, J., and Horst, J. W. (**1995**). ''Frequency discrimination of band-limited harmonic complexes related to vowel formants,'' J. Acoust. Soc. Am. **98**, 1943–1955.
Lyzenga, J., and Horst, J. W. (**1997**). ''Frequency discrimination of stylized synthetic vowels with a single formant,'' J. Acoust. Soc. Am. **102**, 1755–1767.
Mermelstein, P. (**1978**). ''Difference limens for formant frequencies of steady-state and consonant-bound vowels,'' J. Acoust. Soc. Am. **63**, 572–580.
Nelson, P. C., and Carney, L. H. (**2004**). ''A phenomenological model of peripheral and central neural responses to amplitude-modulated tones,'' J. Acoust. Soc. Am. **116**, 2173–2196.
Rabiner, L. R., and Schafer, R. W. (**1978**). *Digital Processing of Speech Signals* (Prentice-Hall, Upper Saddle River, NJ).
Rasmussen, G. L. (**1940**). ''Studies of the VIIIth cranial nerve in man,'' Laryngoscope **50**, 67–83.
Siebert, W. M. (**1965**). ''Some implication of the stochastic behavior of primary auditory neurons,'' Kybernetik **2**, 206–215.
Siebert, W. M. (**1968**). ''Stimulus transformation in the peripheral auditory system,'' in *Recognizing Patterns*, edited by P. A. Kolers and M. Eden (MIT Press, Cambridge, MA), pp. 104–133.
Sinnott, J. M., and Kreiter, N. A. (**1991**). ''Differential sensitivity to vowel continua in Old World monkeys (*Macaca*) and humans,'' J. Acoust. Soc. Am. **89**, 2421–2429.
Srulovicz, P., and Goldstein, J. L. (**1983**). ''The central spectrum: A synthesis of auditory-nerve timing and place cues in monaural communication of frequency spectrum,'' J. Acoust. Soc. Am. **73**, 1266–1276.
Tan, Q. (**2003**). ''Computational and statistical analysis of auditory peripheral processing for vowel-like signals,'' Dissertation, Boston University.
Tan, Q., and Carney, L. H. (**2003**). ''A phenomenological model for the responses of auditory-nerve fibers. II. Nonlinear tuning with a frequency glide,'' J. Acoust. Soc. Am. **114**, 2007–2020. Erratum (**2004).** J. Acoust. Soc. Am. **116**, 3224–3225.
van Trees, H. L. (**1968**). *Detection, Estimation, and Modulation Theory: Part I* (Wiley, New York), Chap. 2.
Van Zanten, G. A. (**1980**). ''Temporal modulation transfer functions for intensity modulated noise bands,'' in *Psychophysical, Physiological and Behavioural Studies in Hearing* (Delft University Press, Delft), pp. 206–209.

Cochlear gain control

Marcel van der Heijden[a)]
Laboratory of Auditory Neurophysiology, Medical School, K.U.Leuven, Herestraat 49—bus 801, B-3000 Leuven, Belgium

(Received 9 August 2004; revised 15 November 2004; accepted 16 December 2004)

The nonlinear auditory phenomena of compression, suppression, and distortion are known to have a cochlear-mechanical origin. An instantaneous nonlinear transfer function is often assumed to underlie these phenomena, but there are experimental indications that auditory nonlinearity is sluggish rather than instantaneous. This study analyzes the consequences of such sluggishness, using automatic gain control (AGC) as a model noninstantaneous nonlinearity. The distinctive characteristic of AGC, its delayed action, is shown to produce a number of observable and measurable effects that distinguish AGC from instantaneous nonlinearities. A major class of such AGC-specific effects concerns the phase of aural distortion products. For example, the phase of the cancellation tone in the classical psychoacoustic cancellation paradigm is linearly related to the frequency spacing of the primary tones in an AGC, as opposed to the square-law relationship produced by an instantaneous nonlinearity. These and other predictions are confronted with experimental data from the literature. The impact of putative AGC-related delays on the interpretation of distortion product otoacoustic emissions (DPOAEs) is discussed. Detailed suggestions are made for experiments specifically aimed at determining whether cochlear nonlinearity is instantaneous or delayed. [DOI: 10.1121/1.1856375]

PACS numbers: 43.64.Bt, 43.66.Ki, 43.64.Kc, 43.64.Jb [AK] Pages: 1223–1233

I. INTRODUCTION

In order to deal with a dynamic range of 100 dB or more, the mammalian auditory system utilizes compressive transformations at several stages of processing (for a review, see Cooper, 2004). Since the pioneering work of Rhode (1971) it has become clear that a significant degree of compression is already achieved at the stage of mechanical vibration in the cochlea. Basilar membrane (BM) vibrations in response to tones grow at compressive rates that can be as low as 0.2 dB/dB and the dynamic range over which compression is significant can be as large as 80 dB (for a review, see Robles and Ruggero, 2001). Closely related to the compressive response to single tones are the suppression of the response to one tone by another (two-tone suppression) and the creation of distortion products (DPs). These different manifestations of cochlear nonlinearity have been extensively studied in psychophysics, neurophysiology, and cochlear-mechanics. In addition, DPs can be measured in the ear canal as distortion-product otoacoustic emissions (DPOAEs).

Cochlear nonlinearity is commonly identified with an instantaneous compressive transfer of the waveform (*e.g.*, Patuzzi, 1996). Accordingly, the key element of many numerical models of auditory nonlinearity is a sigmoidally shaped transfer function (e.g., Weiss and Leong, 1985) and various attempts have been made to derive the shape of this transfer function from mechanical measurements (Geisler and Nutall, 1997) or from otoacoustic emissions (Bian *et al.*, 2002). Several facts, however, are not easily reconciled with the purported instantaneous character of cochlear nonlinearity. For example, instantaneous nonlinearities produce harmonic distortion whereas BM responses to sounds at low and moderate intensities show little harmonic distortion compared to intermodulation distortion of the same order (Cooper, 1998).

Particularly in the domain of suppression there are several indications that auditory nonlinearity is sluggish rather than instantaneous. Suppressors reduce the BM response to characteristic-frequency tones, but the temporal pattern of BM motion shows that suppression cannot follow the suppressor waveform in a cycle-by-cycle fashion for suppressors above 7 kHz (Cooper, 1996). Stated differently, there is a dc component of suppression that is not matched by an ac component (Geisler and Nutall, 1997). Consistent with such sluggishness is the nonzero ''reaction time'' of suppression observed in a recent study of auditory nerve responses of the cat (van der Heijden and Joris, 2005). That study showed that two-tone suppression below 10 kHz does not operate instantaneously, but instead lags excitation by several hundred microseconds. The suppression of otoacoustic emissions in humans also lags the onset of the suppressor, although the interpretation of this latency is hampered by uncertainties on the travel times of emissions. Ruggero and Rich (1983) found a latency of 1.2 ms between the suppressor onset and its effect on a spontaneous emission at 7.5 kHz. Bian *et al.* (2004) used a low-frequency bias tone to suppress DPOAEs around 4 kHz and found that suppression lags the bias cycle by 1.2–1.4 ms. These findings are consistent with a sluggish type of nonlinearity, although Bian *et al.* preferred an interpretation of their data in terms of hysteresis. Unfortunately, either interpretation is uncertain in the absence of detailed knowledge on other sources of delay such as travel times.

a)Electronic mail: marcel.vanderheyden[at]med.kuleuven.ac.be

While these indications against the instantaneous character of auditory nonlinearity may not be decisive, they do encourage the study of alternative views. One of those alternatives is automatic gain control (AGC). The potential occurrence of AGC in cochlear mechanics has previously been discussed by Lyon (1990) and by Zwislocki *et al.* (1996).

In the present study I do *not* analyze how AGC might be realized in the mechanics of the cochlea: no specific ''model'' of cochlear AGC is proposed. Instead, I simply assume that AGC plays a role in cochlear mechanics and analyze the implications for a number of nonlinear auditory phenomena. The aim of this approach is to produce specific predictions that can be tested empirically. I start by showing that AGC accounts for suppression and distortion products equally well as does an instantaneous nonlinearity, and that AGC reproduces many of the seemingly complex features of cochlear nonlinearity such as the nonmonotonic behavior of combination tones with primary level. I then analyze a number of measurable effects that distinguish the operation of AGC from that of instantaneous compression. Each of these effects stems from the delayed action of AGC.

II. THEORY AND DISCUSSION

A. Basic concepts and terminology

This paper deals with the dynamic character of auditory nonlinearity and, more specifically, with the contrast between instantaneous compression and ''slower'' types of compression. For the slower types of compression I will use the generic term AGC, in keeping with Zwislocki *et al.* (1996, p. 621), who write: ''AGC is used by us according to the classical concept that designates a dynamical process with a finite time constant, usually longer than the period of oscillation of the controlled signal. Accordingly, instantaneous nonlinearity is excluded.''

The distinctive feature of AGC is the adjustment of the gain based on an *integrated* measure of stimulus intensity, i.e., one that depends on a finite interval of recent stimulus history. The integration process will usually cause a delay between the fluctuations in stimulus intensity and the ensuing gain adjustments.[1] The mere existence of this delay or ''reaction time'' of the AGC causes measurable effects that are analyzed presently. Though the exact implementation of the AGC is irrelevant for that analysis, it is instructive to discuss an example.

A typical AGC system contains an amplifier whose output is rectified and lowpass filtered (leaky integrator), and fed back to adjust the gain the amplifier. The time constant(s) of the lowpass filter determine the net delay between stimulus fluctuations and the adjustment of the gain. Clearly, a very high cutoff of the filter (small time constant) results in an ''almost instantaneous'' control of the gain. Hence the difference between instantaneous nonlinearity and AGC is gradual rather than categorical. On the other hand, time constants that exceed the period of the stimulus cause a dynamic behavior that is qualitatively different from that of instantaneous nonlinearities (Zwislocki *et al.*, 1996).

The mathematical analyses of this study are based on an AGC scheme that is idealized in two respects. First, the gain is controlled on the basis of the envelope of the input; second, the delayed action of the AGC is realized by a simple retardation of the envelope. The choice of envelope as the measure of momentary stimulus intensity has a distinct, restrictive, effect on the distortion spectrum because the envelope lacks all waveform fine structure. The gain-controlling signals of more realistic AGC systems (e.g., the one based on a leaky integrator) may contain a fine-structure residue (ripple), yielding less restrictive distortion spectra. Thus the absence or presence of certain components in the distortion spectrum do not directly translate into evidence for or against AGC.

The second idealization, the representation of the AGC-induced delay by a single quantity, ignores the potential interplay of multiple time constants such as ''attack time'' and ''release time.'' It also ignores any dynamical aspects that may cause the delay to depend on the rate of intensity fluctuations. This simplification is justified to the extent that the gain adjustments suffers *some* nonzero delay re the stimulus envelope, which is sufficient to distinguish the AGC from instantaneous nonlinearity.

Before plunging into mathematical detail, it is useful to sketch how AGC can account for the well-known nonlinear phenomena of compression, suppression, and distortion. An AGC system lowers its gain with increasing stimulus intensity. When the input consists of a single tone, this leads to a compressive I/O function. An idealized, envelope-based, AGC produces no harmonic distortion because the envelope does not contain the stimulus frequency (for a single tone, the envelope is constant, leading to a constant gain).

When multiple components are present, the gain will be dominated by the strongest one(s), obviously leading to suppression. To understand how AGC produces odd-order combination tones, it helps to view the adjustment of the gain as a form of amplitude modulation. A two-tone input with angular frequencies ω_1 and ω_2 has an envelope that oscillates at the beat frequency $\omega_B=\omega_2-\omega_1$. A sufficiently fast AGC will adjust the gain based on this periodic envelope. Thus the two-tone stimulus is amplitude-modulated by a modulator containing components at ω_B and its harmonics $k\omega_B$. This modulation produces sidebands at frequencies $\omega_1\pm k\omega_B$ and $\omega_2\pm k\omega_B$. In a more familiar notation, the DPs occur at frequencies $(k+1)\omega_1-k\omega_2$ and $(k+1)\omega_2-k\omega_1$. These are precisely the odd-order distortions that play a dominant role in the cochlea. Their privileged role stems from the idealized, envelope-based, character of the AGC. In nonidealized systems, stimulus frequencies may pervade the gain-controlling signal, producing a wider class of distortions including harmonics, difference tones, and other intermodulation products.

In order to produce the odd-order distortions, the gain control mechanism must be able to follow the beating between the primaries. Interestingly, any temporal limitation of the AGC will result in an upper limit to the amplitude fluctuations it can follow. Such limitations would provide a natural explanation for the decline of DP generation with increasing primary spacing. The interpretation of this decline is problematic in the context of instantaneous nonlinearities because suppression does not decline at large primary spacing,

showing that a wide frequency separation in itself does not prevent the nonlinear interaction between the primaries.[2]

B. Effect of AGC on a two-tone stimulus

Mathematically, AGC is described by the stimulus transformation:

$$y(t)=G(t)x(t), \tag{1}$$

where $x(t)$ and $y(t)$ denote the input and output waveforms, and $G(t)$ is the time-varying gain. The gain is adjusted according to the "momentary intensity" of the stimulus, which is ideally represented by the envelope $E(t)$ of the input. Thus:

$$y(t)=G(E)x(t), \tag{2}$$

in which the time-dependence of E is suppressed to avoid cluttering of parentheses. For the transformation to be compressive, G must be a monotonically decreasing function of E, and G must not fall off faster than $1/E$ (or else the output level would decrease with increasing input level):

$$dG/dE<0; \tag{3a}$$

$$G(E)>1/E. \tag{3b}$$

A typical example is a power-law transformation:

$$G(E)=E^{\beta-1}, \tag{4}$$

where $0<\beta<1$ (in practical calculations, the singularity at $E=0$ must be removed, see the Appendix). For a single-tone input, the envelope is flat and equal to the amplitude of the tone. Equation (4) then produces an I/O function with a slope of β dB/dB. Thus β is the index of compression. For a two-tone input, on the other hand, the envelope is fluctuating. The envelope of an input $A\cos\omega_1 t+B\cos\omega_2 t$ is

$$E(t)=(A^2+B^2+2AB\cos\omega_B t)^{1/2}, \tag{5}$$

where $\omega_B=\omega_2-\omega_1$, the angular beat frequency. In general, for any prescribed gain function $G(E)$, the substitution of Eq. (5) into Eq. (2) yields an exact expression for the output of the AGC, which can then be analyzed in terms of suppression and distortion. It is instructive to write $G(t)$ as a Fourier series $\Sigma G_k(t)$ and to focus on the leading terms. For the power-law function of Eq. (4) this expansion is treated in the Appendix, yielding

$$\begin{aligned}
&G_0=(A^2+B^2)^{-\alpha},\\
&G_1(t)=-2G_0c_1(\alpha,R)\cos\omega_B t,\\
&G_2(t)=2G_0c_2(\alpha,R)\cos 2\omega_B t,\\
&\cdots\\
&G_k(t)=(-1)^k 2G_0c_k(\alpha,R)\cos k\omega_B t,
\end{aligned} \tag{6}$$

where $\alpha=(1-\beta)/2$, $R=AB/(A^2+B^2)$, and the positive coefficients $c_k(\alpha,R)$ are defined in Eq. (A10) and further clarified and illustrated in the Appendix. The particular choice of the gain function of Eq. (4) is reflected in the dependence on α of the above components. The dependence on the primary amplitudes enters Eq. (6) via G_0 and R. The alternation of signs $(-1)^k$ of the components is largely independent of the details of the gain function; it is merely related to its asymptotic behavior (cf. Schroeder, 1969). Note in passing that the expression for G_0 (the dc component of the gain) entails the dc component of the mutual suppression of the primaries.

It must now be taken into account that an AGC needs some time to evaluate the stimulus level: the stimulus has to be integrated over a stretch of time. As explained earlier, I incorporate this lag by a single "AGC-related delay" τ_{AGC} and do not worry at this point how it depends on other parameters. It is crucial, however, that τ_{AGC} represents the *inherent* delay of the AGC mechanism as opposed to cochlear travel times and/or "filter delays."[3]

The stage is now set for an evaluation of the distortion spectrum. Each component G_k of Eq. (6) contributes to the modulation of the primaries at ω_1 and ω_2, thereby producing sidebands at frequencies $\omega_{1,2}\pm k\omega_B$. In detail, the DPs stem from multiplying the retarded gain functions $G_k(t-\tau_{\mathrm{AGC}})$ with each of the primaries $A\cos\omega_1 t$ and $B\cos\omega_2 t$. Two spectral series emerge:

$$S_k(t)=(-1)^k AG_0c_k(\alpha,R)\cos\{\omega_1 t+k\omega_B(t-\tau_{\mathrm{AGC}})\}, \tag{7a}$$

$$T_k(t)=(-1)^k BG_0c_k(\alpha,R)\cos\{\omega_2 t+k\omega_B(t-\tau_{\mathrm{AGC}})\}, \tag{7b}$$

where $k=\ldots,-1,0,1,\ldots$ (the $k=0$ terms correctly describe the components at the primary frequencies at the output of the AGC). The spectral series of Eq. (7) are overlapping: S_k and T_{k-1} contribute to one and the same DP. Because the amplitude of the DP decreases with order k, and because k differs by 1 between the two contributions at any given frequency, a single DP is usually dominated by one of the contributions. DPs below the primary frequencies are dominated by contributions from S_k, $k<0$, while DPs above the primary frequencies are dominated by T_k, $k>0$.

We now generalize the distortion spectrum of Eq. (7) to include travel times τ_1 and τ_2 of each of the primaries toward the site of nonlinear interaction, and the travel time τ_{DP} of the DP from the interaction site toward the site where it is measured. Mathematically, this corresponds to a two-tone input $A\cos\omega_1(t-\tau_1)+B\cos\omega_2(t-\tau_2)$ and the substitution $t\rightarrow(t-\tau_{\mathrm{DP}})$ in the final expression. It is convenient to express the travel times of the primaries in an average travel time τ_{PR} and a travel-time disparity $\delta\tau$ as follows:

$$\tau_{\mathrm{PR}}=(\tau_2+\tau_1)/2, \quad \delta\tau=(\tau_2-\tau_1).$$

The physical origin of the travel-time disparity is the dispersion of wave propagation on the BM, i.e., the dependence of propagation speed on frequency. Abbreviating the amplitudes, which are unaffected by the inclusion of travel times, we obtain:

$$\begin{aligned}
S_k(t)=(-1)^k A_k\cos\{&(\omega_1+k\omega_B)(t-\tau_{\mathrm{PR}}-\tau_{\mathrm{DP}})\\
&-k\omega_B\tau_{\mathrm{AGC}}-[(k-1)\omega_1+k\omega_2]\delta\tau/2\},
\end{aligned} \tag{8a}$$

$$\begin{aligned}
T_k(t)=(-1)^k B_k\cos\{&(\omega_2+k\omega_B)(t-\tau_{\mathrm{PR}}-\tau_{\mathrm{DP}})\\
&-k\omega_B\tau_{\mathrm{AGC}}-[k\omega_1+(k+1)\omega_2]\delta\tau/2\}.
\end{aligned} \tag{8b}$$

The three terms in curly braces reflect to three type of delays:

(1) $(\omega_{1,2}+k\omega_B)(t-\tau_{PR}-\tau_{DP})$ reflects the *overall travel time* of primaries and DPs taken together. Alone, this term would represent a single tone at the DP frequency delayed by the total travel time $(\tau_{PR}+\tau_{DP})$.
(2) $k\omega_B\tau_{AGC}=k(\omega_2-\omega_1)\tau_{AGC}$ reflects the AGC-related delay τ_{AGC}.
(3) $[\cdots]$ $\delta\tau/2$ stems from the disparity $\delta\tau$ in travel time between the primaries.

The DP phases of Eq. (8) can be rewritten in a form that is more convenient for the analysis of (ω_1,ω_2) area studies (Knight and Kemp, 2000; Schneider *et al.*, 2003). With the convention that the phase of $\cos\omega t$ is zero, the phases become

$$\phi_S(k)=k\pi+\omega_1[(k-1)(\tau_{PR}+\tau_{DP}-\delta\tau/2)+k\tau_{AGC}] - \omega_2 k(\tau_{PR}+\tau_{DP}+\tau_{AGC}+\delta\tau/2), \quad (9a)$$

$$\phi_T(k)=k\pi+\omega_1 k(\tau_{PR}+\tau_{DP}+\tau_{AGC}-\delta\tau/2) - \omega_2[(k+1)(\tau_{PR}+\tau_{DP}+\delta\tau/2)+k\tau_{AGC}]. \quad (9b)$$

For situations in which the travel times τ_{PR} and τ_{DP} dominate the other time parameters τ_{AGC} and $\delta\tau$, Eq. (9) predicts that the equal-phase contours in the (ω_1,ω_2) plane are approximately parallel to the lines of constant DP frequency $(\omega_1+k\omega_B)$. Many of the DPOAE phase data of Knight and Kemp (2000) indeed show this patterning of phase contours. On the other hand, Eq. (9) also shows that both the AGC-related delay τ_{AGC} and the travel-time disparity $\delta\tau$ cause systematic deviations from this basic alignment. These deviations are further analyzed in Sec. II D.

C. Amplitude of the cubic difference tone

To illustrate the properties of the distortion spectrum of Eq. (7), we now examine the lower cubic difference tone (CDT) at $2\omega_1-\omega_2$. The two contributions to this CTD are

$$S_{-1}(t)=-AG_0c_1(\alpha,R)\cos\{(2\omega_1-\omega_2)t + (\omega_2-\omega_1)\tau_{AGC}\}, \quad (10a)$$

$$T_{-2}(t)=+BG_0c_2(\alpha,R)\cos\{(2\omega_1-\omega_2)t + 2(\omega_2-\omega_1)\tau_{AGC}\}. \quad (10b)$$

First note that the amplitude *ratio* of S_{-1} to T_{-2} is independent of the overall level of the primaries; it only depends on the amplitude ratio A/B of the primaries [see Eq. (A3)]. Next, when the overall level of the primaries is increased while fixing the relative level, then both S_{-1} and T_{-2} grow at the same compressive rate as does the *output* level of the primaries, *viz.*, at $(1-2\alpha)=\beta$ dB/dB. Moreover, this is true for the entire spectrum of Eq. (7): everything that comes out of the AGC is compressed at the same rate. This is consistent with the common observation that the amplitude ratio of DPOAEs and primary *input* decreases with increasing input level (e.g., Wilson, 1980). Taken at face value, the growth of DPOAEs provides a direct measure of the index of compression β in the cochlea.

Figure 1 shows the levels of S_{-1} and T_{-2} as a function of the input level L_2 of the second primary; a compression index $\beta=0.3$ dB/dB was used. In this graph the first primary is used as a reference in two ways: the *input* level L_2 (abscissa) is expressed *re* the fixed input level L_1, and the *output* levels of the CDT components (ordinate) are expressed *re* the output level of the first primary when presented in isolation. From the above remarks it follows that the curves of Fig. 1 do not change with overall primary level. In addition, the curves give an instant measure of the amount of distortion: 0 dB on the ordinate corresponds to the level of the undistorted (but compressed) output of the primaries. For $\beta=0.3$ dB/dB, the peak intermodulation distortion is -13 dB.

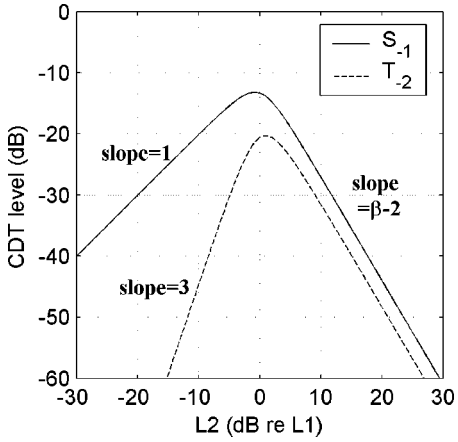


FIG. 1. Levels of the two contributions to the CDT at $2\omega_1-\omega_2$ as a function of the input level L_2 of the primary at ω_2, computed for an AGC with a compression index $\beta=0.3$ dB/dB. The two contributions S_{-1} and T_{-2} (see the text) are identified on the graph. The levels of the CDT components are expressed in dB *re* the output level of the fixed ω_1 primary when presented in isolation. Asymptotic slopes are indicated next to the curves where they apply.

Both curves in Fig. 1 show a nonmonotonic dependence on relative primary level, and both curves peak when the primary levels L_1 and L_2 are approximately equal. S_{-1} is consistently larger than T_{-2} and falls of less steeply when L_2 is decreased. The asymptotic slopes of the curves follow from Eq. (A16); they are indicated in the graph. For large L_2, the two curves become parallel and S_{-1} exceeds T_{-2} by a fixed amount $20\log(4/(3-\beta))$ dB [cf. Eq. (A17)]. In Fig. 1, where $\beta=0.3$ dB/dB, this asymptotic level difference is 3.4 dB. Figure 1 shows that the S_{-1} term will certainly dominate when, as is routinely done, a 10 dB level difference between the primaries is employed favoring L_1.

The inverted-U shape of the S_{-1} curve (Fig. 1) and its asymptotic slopes of $+1$ and $(\beta-2)$ dB/dB closely resemble data on CDT level dependence of ω_2/ω_1 ratios close to unity (e.g., Zwicker, 1981). This agreement, however, does not by itself afford an argument in favor of AGC because the patterning of CDT amplitudes is equally well described by an instantaneous type of compression (e.g., Goldstein, 1967; Duifhuis, 1976). It is in the phase behavior that the most distinctive feature of the AGC, its delayed action, is revealed most clearly.

D. Phase of the cubic difference tone

Assuming that the lower and upper CDTs are dominated by the S_{-1} and T_{+1} terms, their phases [see Eq. (8)] are

$$\phi_{\mathrm{CDTL}} = \phi_S(-1) = \pi - (2\omega_1 - \omega_2)(\tau_{\mathrm{PR}} + \tau_{\mathrm{DP}}) + (\omega_2 - \omega_1)\tau_{\mathrm{AGC}} + (\omega_1 + \omega_2/2)\,\delta\tau, \tag{11a}$$

$$\phi_{\mathrm{CDTU}} = \phi_T(+1) = \pi - (2\omega_2 - \omega_1)(\tau_{\mathrm{PR}} + \tau_{\mathrm{DP}}) - (\omega_2 - \omega_1)\tau_{\mathrm{AGC}} - (\omega_1/2 + \omega_2)\,\delta\tau. \tag{11b}$$

Note that ϕ_{CDTU} can be obtained from ϕ_{CDTL} by interchanging the primaries, i.e., $\omega_1 \leftrightarrow \omega_2$ and $\delta \rightarrow -\delta\tau$ [this reflects the more general symmetry between S_k and T_{-k} of Eq. (7)]. We therefore restrict the discussion to ϕ_{CDTL} without loss of generality.[4] Equation (11) is exact, but its interpretation is complicated by the many unknowns it contains: the travel-time parameters τ_{PR}, τ_{DP}, and $\delta\tau$, and the AGC-related delay τ_{AGC}. Moreover, it is unknown how the travel times vary with the primary frequencies. It will prove useful to linearize in $(\omega_2 - \omega_1)$. Since always $\delta\tau = 0$ for $\omega_2 = \omega_1$, we may write, to first order

$$\delta\tau = (\omega_2 - \omega_1)u, \tag{12}$$

where the parameter u quantifies the disparity of travel times[5] of the primaries at the place of CDT generation. It is important that, in writing down Eq. (12), all we use is the fact that $\delta\tau \rightarrow 0$ when $\omega_2 \rightarrow \omega_1$. This is true even in the case of different primary levels because the two primaries are always simultaneously present: when ω_1 and ω_2 coincide there exists only a single tone. For the present purposes, it is unnecessary to make further assumptions concerning the traveling wave such as ''scaling invariance'' (Shera *et al.*, 2000). The single parameter u absorbs both the change of delay at a fixed location and the shift of the CDT generation site. In particular, Eq. (12) is equally valid for ''fixed-wave,'' ''fixed-place,'' or alternative types of CDT generation (Kemp and Brown, 1983; Zweig and Shera, 1995). Using the linearization of Eq. (12), the CDT phase of Eq. (11a) becomes, to first order in $(\omega_2 - \omega_1)$,

$$\phi_{\mathrm{CDT}} = \pi - (2\omega_1 - \omega_2)(\tau_{\mathrm{PR}} + \tau_{\mathrm{DP}}) + (\omega_2 - \omega_1)(\tau_{\mathrm{AGC}} + 3/2u\omega_1). \tag{13}$$

When applying Eq. (13) to a specific experimental paradigm, the unknown character of τ_{DP} (the time it takes the DP to travel from its source to the site where it is measured) is the main stumbling block. This is particularly true for DPOAEs. First, the dependence of τ_{DP} on the primary frequencies is unknown. Second, the actual *path* taken by the DP is uncertain and may even depend on the relation between DP frequency and generation site (Knight and Kemp, 2000). Third, multiple paths of DP propagation may converge to the measurement site (Heitmann *et al.*, 1998; Kalluri and Shera, 2001).

Despite these uncertainties, Eq. (13) clearly shows that the phase effect of an AGC-related delay is entangled with the dispersive effects of primary spacing. The former is quantified by τ_{AGC} and the latter by u; their entanglement or tradeoff is represented by the factor $(\tau_{\mathrm{AGC}} + 3/2u\omega_1)$ in Eq. (13). When analyzing DP phase, therefore, ignoring the effects of τ_{AGC} will generally lead to a systematic overestimation of the parameter u, introducing biases in the estimation of traveling-wave parameters from phase data. Furthermore, due to the tradeoff between u and τ_{AGC}, any effect of overall sound level on τ_{AGC} (see the following) is easily mistaken for an effect of sound level on the traveling wave phases or DP generation.

E. Phase of cancellation tones

In the classic cancellation method (Lewis and Larsen, 1937; Zwicker, 1955), a ''cancellation tone'' at the DP frequency is presented along with the primaries. The amplitude and phase of the cancellation tone is adjusted to eliminate the percept of the DP. Cancellation data provide a test of the existence of AGC-related lags, as can be understood as follows. When the primaries are closely spaced, then ω_1, ω_2, and the frequencies of low-order DPs are all close together. Because their traveling waves largely overlap, the differences in propagation will be small and are bound to cancel when inspecting the *differential* effects measured with the cancellation method. In particular, the travel time of the DP from its generation site to its characteristic place becomes irrelevant because it is common to the DP and the cancellation tone (the DP and the cancellation tone are really one and the same tone). This situation is in stark contrast to DPOAEs, whose measured phases critically depend on their backward propagation to the ear canal.

The mathematical details of the cancellation of the CDT are as follows. The linearization that led to Eq. (12) is used to express the phase of the cancellation tone (this is justified by the close spacing of all components considered). The CDT frequency is $3/2(\omega_2 - \omega_1)$ away from the average frequency of the primaries. Using the linearization of travel times [Eq. (12)], the travel time τ_{CL} of the cancellation tone to the site of CDT generation becomes

$$\tau_{\mathrm{CL}} = \tau_{\mathrm{PR}} - 3/2(\omega_2 - \omega_1)u.$$

To first order in $(\omega_2 - \omega_1)$, then, the phase of the cancellation tone beyond the CDT generation site becomes

$$\begin{aligned}\phi_{\mathrm{CL}} &= \phi_{\mathrm{ADJ}} - (2\omega_1 - \omega_2)(\tau_{\mathrm{CL}} + \tau_{\mathrm{DP}}) \\ &= \phi_{\mathrm{ADJ}} - (2\omega_1 - \omega_2)(\tau_{\mathrm{PR}} + \tau_{\mathrm{DP}}) \\ &\quad + 3/2(\omega_2 - \omega_1)u\omega_1,\end{aligned}$$

where ϕ_{ADJ} is the (adjustable) phase of the cancellation tone in the apparatus. The requirement of cancellation with the CDT [Eq. (13)] yields

$$\phi_{\mathrm{ADJ}} = (\omega_2 - \omega_1)\tau_{\mathrm{AGC}} + \eta(\omega_2 - \omega_1)^2 + \cdots, \tag{14}$$

where the ellipsis denotes third-order and higher-order terms in $(\omega_2 - \omega_1)$. As anticipated, to leading order all travel delays have canceled in Eq. (14). This isolates the effect of the AGC-related delay, which enters the scene in an uncompromising way. If there is a nonzero AGC-related delay $\tau_{\mathrm{AGC}} > 0$, then, for small primary separations, ϕ_{ADJ} should vary linearly with $(\omega_2 - \omega_1)$, the coefficient of proportionality being τ_{AGC}. If, on the other hand, the compressive nonlinearity is instantaneous ($\tau_{\mathrm{AGC}} = 0$) then Eq. (14) tells us that ϕ_{ADJ} varies as $\eta(\omega_2 - \omega_1)^2$ for closely spaced primaries. In that case the direction of change with primary spacing depends exclusively on the sign of the coefficient η, which in turn depends on traveling wave details. Equation (14) generalizes

to higher-order DPs, the linear coefficient being $-k\tau_{\mathrm{AGC}}$ for the DP at $\omega_1+k(\omega_2-\omega_1)$.

A simple test of linear phase behavior near $\omega_2=\omega_1$ is to look for a change of sign in the cancellation phase during the transition from $\omega_2>\omega_1$ to $\omega_2<\omega_1$. In psychoacoustics this is not possible due to the inaudibility of upper CDTs. The next best method is to look for a nonvanishing slope of the phase when approaching $\omega_2=\omega_1$. The CDT, however, becomes inaudible when the primaries are too close; no data are available for primaries whose relative frequency difference is less than 5%. These two restrictions can be summarized in the single requirement that $\omega_2/\omega_1>1.05$ for psychoacoustic cancellation tests. No such restriction exists in the measurement of DPs on the basilar membrane (Robles and Ruggero, 2001), or is there a reason to expect such restrictions in the temporal analysis of auditory nerve data (Goldstein and Kiang, 1968; Buunen and Rhode, 1978). The best test of the phase behavior, therefore, can be expected from applying the cancellation method to cochlear-mechanical measurements or auditory nerve recordings.[6] Until the advent of such data, the classic psychoacoustic studies are our best source of information.

A square-law behavior of ϕ_{ADJ} was claimed by Goldstein (1967), although part of his phase data (plotted on a log–log scale) deviates from this behavior. Goldstein also gave a mathematical argument for the square-law phase behavior, based on an instantaneous nonlinearity. This theoretical argument was elaborated by Schroeder (1969). Many of Hall's (1972) phase curves look quite linear (e.g., Fig. 1, 1475 Hz curve), and there is a systematic trend for the phase curves to become less steep with increasing sound level (e.g., Fig. 1 versus Fig. 3). If the slope of the phase curves is identified with the AGC-related delay [Eq. (14)] then Hall's data suggest that τ_{AGC} decreases with sound level: the AGC becomes faster with growing intensity. The same trend is seen in many other data sets (e.g., Zwicker, 1981). The phase data of Zwicker (1981, Figs. 5, 8, and 15) were plotted using a compressed (square-root) abscissa and a linear ordinate. This choice of scaling, which is not clarified in the paper, prevents a straightforward evaluation because it makes a linear phase curve look like a parabola and a square-law relationship like a fourth power.

Wilson's (1980) cancellation data were obtained while varying the frequency ratio ω_2/ω_1 and keeping the frequency of the lower CDT constant at 800 Hz. Data from Wilson's Fig. 1 are reproduced in Fig. 2 (symbols). We plot cancellation phase ϕ_{ADJ} as a function of $(\omega_2-\omega_1)$ instead of the ratio ω_2/ω_1 used in the original graph. To estimate the contribution of a linear term as in Eq. (14), each phase curve was fitted to a polynomial $\phi_{\mathrm{ADJ}}=\gamma(\omega_2-\omega_1)+\eta(\omega_2-\omega_1)^2$. The fitted curves are shown as lines in the graph. Identifying the linear coefficient γ with the AGC-related delay, Wilson's data yield $\tau_{\mathrm{AGC}}=4.0$, 3.2, 2.5, 2.3, 1.8, 1.5, and 1.7 ms for primary levels of 20, 30, 40, 50, 60, 70 and 80 dB SL. This is a quantitative assessment of our earlier remark that τ_{AGC} seems to decrease with sound level. Note that the delays at low levels amount to several cycles of the primary frequencies.[7] Such large delays are consistent with the lag of suppression observed in the auditory nerve of the cat (van der Heijden and Joris, 2005). Delays of several cycles preclude any interpretation in terms of hysteresis (Bian *et al.*, 2004). This concludes our perusal of psychoacoustic cancellation data.

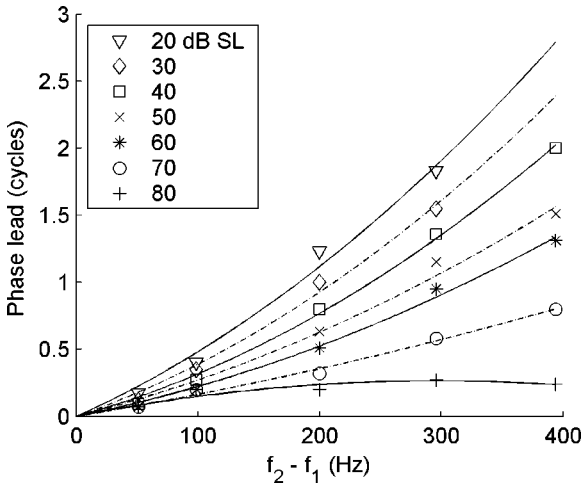


FIG. 2. Phase of the cancellation tone as a function of primary spacing. Data of Wilson (1980) (symbols) are reproduced along with the best fitting second-order polynomials with vanishing constant term (lines). Individual curves correspond to different primary levels as indicated in the graph.

How general is the analysis of cancellation phase leading to Eq. (14)? One assumption used in the derivation is that primaries, CDT, and cancellation tone all travel along the same inward path (this allows one to treat the different types of delay as cumulative, i.e., to add them). This assumption is well backed up by physiological data (for a discussion, see Buunen and Rhode, 1978) and by theoretical analysis (Shera, 2003). A second assumption is the dominance of the S_{-1} term over the T_{-2} term in Eq. (10). This is under the experimenter's control: the dominance is guaranteed by an appropriate choice of relative primary levels, viz., $L_1-L_2\geqslant 10$ dB (Fig. 1). The general case is treated in Sec. II F below.

It is important that the reasoning leading to Eq. (14) does not depend on the assumption that the nonlinear interaction takes place at a single location. Single distortion products probably arise over an extended region of the cochlea, as convincingly argued by Zwicker (1980a, 1980b). But the distributed emergence of DPs does not affect the presence or absence of the linear term in Eq. (14) because the addition of multiple second-order terms $\eta(\omega_2-\omega_1)^2$ will never create a linear term $\gamma(\omega_2-\omega_1)$. On the other hand, if the phase dispersion across the primaries is significant, multiple contributions may cancel in exceptional cases, giving rise to typical notches in DP amplitude (e.g., Goldstein, 1967). For this to happen, phase dispersion must be appreciable, and this in turn is directly linked to cochlear frequency selectivity (Schroeder, 1969; Zwicker, 1981). For an experimental assessment of the linear term in Eq. (14), it is best to stay away from the larger frequency ratios ($\geqslant$1.3) at which such anomalies usually occur.

F. Interference effects

There is another class of effects predicted by AGC but not by instantaneous nonlinearities. In much of the preced-

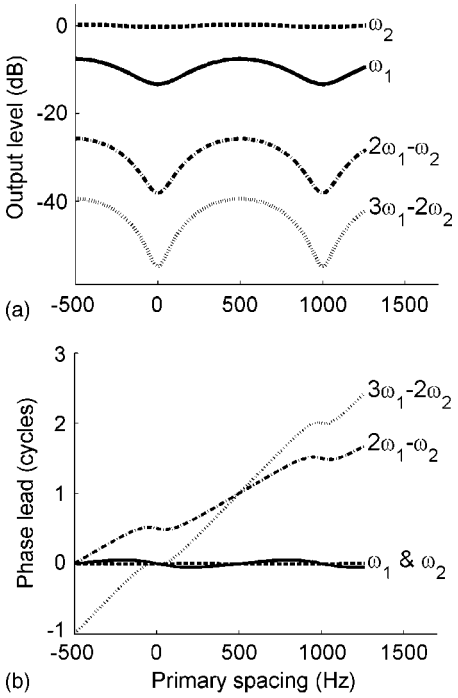


FIG. 3. AGC-related interference effects. (a) The amplitudes of four components at the output of an AGC for a two-tone input. The components shown are the primaries at ω_1 and ω_2, the CDT at $2\omega_1-\omega_2$, and the fifth-order DP at $3\omega_1-2\omega_2$, as indicated in the graph. Primary spacing at the abscissa corresponds to $\omega_2-\omega_1$; note that both $\omega_2<\omega_1$, and $\omega_2>\omega_1$ are covered. The index of compression was 0.3 dB/dB. The AGC-related delay was 1 ms. The input level L_1 of the ω_1 primary was 10 dB below L_2. At the ordinate, 0 dB corresponds to the output level of the strongest primary when presented alone. (b) The unwrapped phases of the same components.

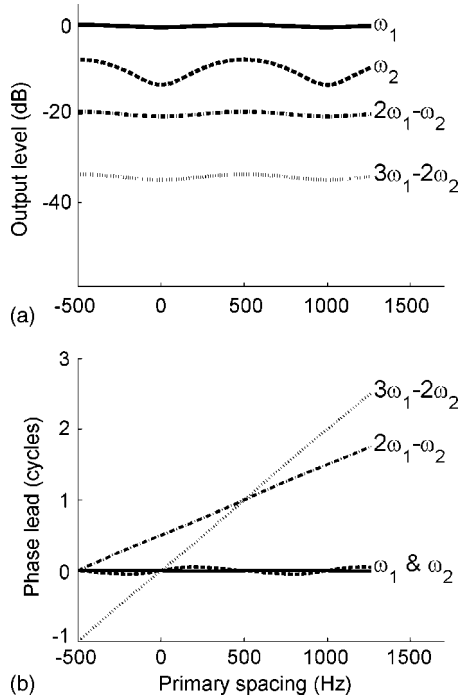


FIG. 4. Same as Fig. 3, except for the primary levels: here L_1 is 10 dB above L_2.

ing, the CDT was identified with the S_{-1} term of Eq. (10), while the T_{-2} was ignored. In reality, Fig. 1 shows that over a large range of relative primary levels the T_{-2} term is only slightly weaker than the S_{-1} term; the asymptotic S_{-1}/T_{-2} amplitude ratio amounts to 2.5–6 dB, depending on the index of compression β (the quoted values occur when $\beta=0$ and 1, respectively; see Sec. II C). Thus T_{-2} will significantly contribute to the lower CDT as long as the (effective) level L_2 of the upper primary is not much lower than the level L_1 of the lower primary. Now from Eq. (8), the *relative phase* of the S_{-1} and T_{-2} terms amounts to $(\omega_2-\omega_1)\tau_{\mathrm{AGC}}$. Thus, when the primary spacing is varied, the running relative phase of the S_{-1} and T_{-2} terms will cause fluctuations in the amplitude of the CDT. This interference effect generalizes to DPs of all orders because any DP is a mix of an S_k and a T_{k-1} term [Eq. (10)] whose phases differ by $(\omega_2-\omega_1)\tau_{\mathrm{AGC}}$. Moreover, the amplitudes of the DPs of different orders should co-vary.

Figure 3 illustrates this interference effect. The amplitudes of the different components at the output of the AGC are plotted as a function of primary spacing in (a). The curves are calculated using a constant CDT frequency of 3 kHz, primary levels $L_2=L_1+10$ dB, and an AGC-related delay τ_{AGC} of 1 ms (three cycles of the CDT). The index of compression β was 0.3 dB/dB. An important characteristic of the amplitude fluctuations is the location of a minimum at zero primary spacing for all components. Note that the graph includes conditions with $(\omega_2-\omega_1)<0$, so that ω_1 (solid line) is not necessarily the lower primary. Over the whole range of primary spacings, though, ω_1 is always the primary that is nearest to the DPs shown.

Interestingly, the primaries themselves are no exception to the interference effect. The interference of the primaries S_0 and T_0 with the T_{-1} and S_1 terms produces oscillations completely analogous to those of the CDT and higher-order DPs. This means that the suppression of one primary by the other will contain a component that oscillates with increasing primary spacing. Naturally the effect is more pronounced for the weaker than for the stronger primary [upper two curves in Fig. 3(a)].

The interference of S and T terms also affects the phase of the components [Fig. 3(b)], causing plateaus in DP phase around $(\omega_2-\omega_1)\tau_{\mathrm{AGC}}=0,\pm1,\ldots$. The plateaus spoil the otherwise linear phase behavior of the DPs. Thus, when testing the linear phase behavior (Sec. II E), it is best to choose a relative primary level that avoids these interference effects (see the following). The phase of the weaker primary at ω_1 varies along with the DP phase. This effect might underlie some of the contradictory findings regarding the effect of suppression on probe phase in BM measurements (for a discussion, see Cooper, 1996).

Figure 4 is the analogue of Fig. 3 for the opposite choice of primary levels, now favoring the primary closest to the CDT: $L_1=L_2+10$ dB. As expected, this choice of primary levels removes most of the interference effects, except for the weaker primary at ω_2. In particular, the linear phase behavior of the DPs is restored, rendering this the ideal condition for a test of Eq. (14).

It is somewhat surprising that the oscillations shown in Fig. 3, even though they result from interference, do not require a combination of two physically separated sources or

alternative travel paths of a DP. The interference effects of Fig. 3 are due to time differences, but these time differences have nothing to do with travel times or different modes of backward propagation (Heitmann *et al.*, 1998; Kalluri and Shera, 2001).

Are the effects shown in Fig. 3 a potential explanation for the notches routinely found in CDT cancellation data (e.g., Goldstein, 1967)? The interference of *S* and *T* terms is ''gentle'' in the sense that the *S* term always dominates the *T* term, preventing perfect annihilation. Specifically, the S_{-1}/T_{-2} amplitude ratio has a lower bound of $4/(3-\beta)$ [cf. Eq. (A17)], resulting in a maximum CDT notch-depth of $20\log((7-\beta)/(1+\beta))$ dB. This maximum notch-depth is largest for small β and decreases from 17 to 9.5 dB when β increases from 0 to 1. When $\beta=0.3$, the maximum notch-depth is 14 dB. The notches reported by Goldstein (1967) and by many others (e.g., Smoorenburg, 1972) are of a more violent type, with notch-depths exceeding 25 dB and sudden 0.5-cycle phase jumps. The latter anomalies point to a nearly perfect annihilation of two or more components (Zwicker, 1980a; 1980b) and cannot result from the gentle type of interference shown in Fig. 3.

The AGC-related interference described here might provide an explanation for those parts of DPOAE ''fine structure'' that are not tied to DPOAE frequency. For instance, Fig. 10b of Schneider *et al.* (2003) shows humps in the amplitudes of DPOAEs which are consistently aligned parallel to the effective ($\omega_2=\omega_1$) axis of the plot, a mode of alignment for which the authors ''cannot yet give [...] a simple interpretation.'' Consistent with Figs. 3 and 4, Schneider *et al.* observe this type of alignment when $L_1<L_2$ but not when $L_1>L_2$.

When inspecting DPOAE data for the occurrence of the interference effects described earlier, the existence of other types of DPOAE fine structure (Talmadge *et al.*, 1999) may be a confounding factor. On the other hand, the prediction that DPs of all orders should co-vary, including even the primaries themselves, is a strong prediction which should facilitate experimental testing. Evidence for these fluctuations may also be sought in measurements of BM motion or from recordings of the auditory nerve. Regardless of the method, it is best to use low overall sound levels as this will minimize the expected frequency interval between spectral maxima and minima (due to the decrease of τ_{AGC} with level).

III. SUMMARY AND CONCLUSION

Automatic gain control accounts well for the nonlinear auditory phenomena of compression, suppression, and distortion (Secs. II A, II B, and II C). The most distinctive feature of a realistic AGC is its finite integration time, which causes a delay that is an inherent property of the nonlinearity (as opposed to cochlear travel times). This AGC-related delay gives rise to a number of phase effects (Secs. II D, II F) that are alien to the instantaneous type of nonlinearity commonly considered to underlie auditory nonlinearity. Presumably the best way to explore the putative AGC-related delays is the measurement of the phase of tones that cancel the CDT produced by a pair of primary tones. Any AGC-related delay will show up as the slope of a linear relation between the phase of the cancellation tone and the frequency spacing of the primaries (Sec. II E). The absence of such a delay would result in a square-law phase behavior.

Existing psychoacoustic cancellation data, when interpreted in terms of AGC-related delays, suggest that these delays become smaller with increasing sound level (e.g., Fig. 2). The data, however, are not conclusive as to the existence of AGC-related delays. Because part of the inconclusiveness is due to inherent limitations of the psychoacoustic cancellation method, I suggest that the cancellation method be employed in future physiological measurements (Sec. II E), and indicate which stimulus conditions are optimal for that purpose (Secs. II E, II F).

If AGC-related delays τ_{AGC} exist, they have a serious impact on the interpretation of DP data in terms of traveling waves. The most important consequences are:

(1) Estimates of phase dispersion in the cochlea must be corrected for the phase effects associated with τ_{AGC} [Eqs. (9) and (13)].
(2) The effect of sound level on DP phase will need reinterpretation, as it might reflect changes in τ_{AGC} rather than changes in wave propagation or DP generation (Secs. II D, II E)
(3) Not all interference effects imply the existence of physically separate sources or alternative travel paths of DPs (Sec. II F).

Does a nonzero value of τ_{AGC} necessarily imply the parametric character of auditory compression exemplified by the leaky-integrator system (Sec II A)? It cannot be excluded off-hand that an instantaneous nonlinearity, when integrated into a dynamic system like an oscillator, would lead to the type of delay between waveform intensity and gain adjustment that is termed AGC-related in this study. The delay would then be an ''emergent property'' of the nonlinear dynamic system rather than the reflection of a physical time constant. I am not aware of an example of a nonlinear oscillator that shows this behavior and that also produces the sluggishness of suppression reported by Cooper (1996) and the delay of suppression reported by van der Heijden and Joris (2005). Gain control, on the other hand, provides a simple framework in which these different phenomena are by-products of the lowpass character of the control signal; the measured time constants should all reflect this common origin.

The existence of cochlear gain control, if confirmed, will affect the interpretation of auditory nonlinearity in terms of cochlear anatomy and physiology. Because the auditory nonlinearities in question are widely assumed to have a cochlear-mechanical origin, the anatomical structures responsible for AGC must take part in the mechanics, too. Other potential sites of AGC, such as the synapses between inner hair cells and nerve fibers (Geisler and Greenberg, 1986) are therefore unlikely to play a role in the nonlinearities at issue. Regarding the role of *outer* hair cells, an envelope-following gain [Eq. (2)] is more likely to originate from the dc component of the transduction current of outer hair cells than from the ac component.

Though most of the above deals with nonlinearities of

the cochlea, there is no a priori reason why gain-control mechanisms should be restricted to animals having a cochlea. The inner ears of other vertebrates show nonlinear behavior, too (e.g., Manley *et al.*, 1993) and the analyses offered in this study are equally applicable to these ears. In fact, the absence of traveling waves could only facilitate the measurement of any AGC-related delays.

In view of the above-mentioned ramifications, it is remarkable how little is really known about the temporal aspects of the nonlinearity of the inner ear. The present author hopes that this study will inspire colleagues from different areas to tackle this intriguing problem.

ACKNOWLEDGMENTS

I thank Philip Joris and Alberto Recio for comments on an earlier draft, Armin Kohlrausch for careful comments, and Neal Viemeister for his perceptive remark that ''it all boils down to time constants.'' This project was supported by the Fund for Scientific Research—Flanders (G.0083.02), Research Fund K.U.Leuven (OT/01/42), and a K.U.Leuven fellowship (F/00/92).

APPENDIX: SERIES EXPANSION OF THE POWER-LAW GAIN

In this Appendix the power-law gain of Eq. (4) is evaluated for a two-tone input. The Hilbert envelope of the two-tone input [Eq. (5)] yields a gain

$$G = G_0(1 + 2R\cos\omega_B t)^{-\alpha}, \tag{A1}$$

where

$$\alpha \equiv (1-\beta)/2, \tag{A2a}$$

$$G_0 \equiv (A^2 + B^2)^{-\alpha}, \tag{A2b}$$

and

$$R \equiv AB/(A^2 + B^2) \tag{A3a}$$

$$= (A/B)/(1 + (A/B)^2) \tag{A3b}$$

$$= (B/A)/(1 + (B/A)^2). \tag{A3c}$$

(The latter two expressions emphasize that R depends exclusively on the ratio of the primary amplitudes.) In order to obtain a Fourier series for G, we write $\cos\omega_B t$ as a sum of exponentials:

$$\cos\omega_B t = [\exp(i\omega_B t) + \exp(-i\omega_B t)]/2 \equiv (x + x^{-1})/2, \tag{A4}$$

and rewrite Eq. (A1) as

$$G = G_0(1 + R(x + x^{-1}))^{-\alpha}, \tag{A5}$$

leading to a power series in x,

$$G/G_0 = \sum_{k=0}^{\infty} \frac{(-\alpha)(-\alpha-1)\cdots(-\alpha-k+1)}{k!} R^k \times \sum_{m=0}^{k} \frac{k!}{(k-m)!m!} x^{k-2m}. \tag{A6}$$

This can be simplified to

$$G/G_0 = \sum_{m=0}^{\infty} \sum_{k=m}^{\infty} x^{k-2m} \frac{(\alpha)_k (-R)^k}{(k-m)!m!}, \tag{A7}$$

where we used the Pochhammer symbol $(\alpha)_k \equiv (\alpha)(\alpha+1)\cdots(\alpha+k-1)$. Equation (A7) is a Laurent series in x:

$$G/G_0 = \sum_{n=-\infty}^{\infty} (-1)^n c_n x^n \tag{A8}$$

and Eq. (A5) implies that $c_n = c_{-n}$. It therefore suffices to consider $n \geqslant 0$. The coefficients c_n are retrieved by collecting terms with $k = (n+2m)$ in Eq. (A7), yielding

$$c_n = \sum_{m=0}^{\infty} \frac{(\alpha)_{n+2m}}{(n+m)!m!} R^{n+2m}, \quad n \geqslant 0. \tag{A9}$$

A rearrangement of the factors of Eq. (A9) leads to the final expression for the coefficients,

$$c_n = R^n \frac{(\alpha)_n}{n!} F\left(\frac{\alpha+n}{2}, \frac{\alpha+n+1}{2}; n+1; (2R)^2\right), \quad n \geqslant 0, \tag{A10a}$$

$$c_{-n} = c_n, \tag{A10b}$$

where F is the hypergeometric function defined by

$$F(\alpha,\beta;\gamma;z) = \sum_{m=0}^{\infty} \frac{(\alpha)_m (\beta)_m}{(\gamma)_m m!} z^m \tag{A11}$$

(e.g., Gradstheyn and Ryzhik, 1994). Combining Eqs. (A4) and (A8), the gain is written as a Fourier series:

$$G = G_0\left[1 + 2\sum_{n=1}^{\infty} (-1)^n c_n(\alpha,R)\cos(n\omega_B t)\right]. \tag{A12}$$

This expression reveals that the coefficients c_n defined in Eq. (A10) are the relative amplitudes of the harmonics of the beat frequency in the gain function. These relative amplitudes depend exclusively on the relative amplitudes of the primaries, because R can be expressed in the ratio A/B [cf. Eq. (A3b)].

The $c_n(\alpha,R)$ of Eq. (A10) are singular for $A=B$. This divergence stems from the periodic zeros of the envelope of an equal-amplitude two-tone complex: a strict application of the power-law gain function of Eq. (4) leads to singularities for a vanishing envelope. A realistic gain function, however, has an upper limit, which is conveniently implemented by the following adaptation of the power-law gain function:

$$G(E) = (E^2 + \epsilon^2)^{(1-\beta)/2}, \tag{A13}$$

where ϵ is that value of the envelope below which the gain saturates. In practical calculations involving Eq. (A10), such a constraint amounts to the substitution

$$R \rightarrow (1-\delta)R, \tag{A14}$$

where $0 < \delta < 1$. The calculations in the present study are all based on $\delta = 0.1$. The precise value of δ is not important for the present purposes.[8]

Figure 5 shows the magnitude of the amplitude coefficients $c_1 - c_3$ as a function of relative primary level $20\log(A/B)$. An index of compression β of 0.3 dB/dB was used. The magnitude of the components c_n decreases with

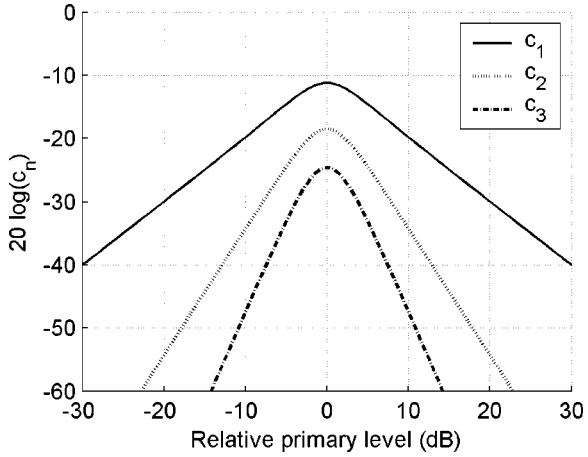


FIG. 5. Magnitude of the amplitude coefficients c_n of a power-law gain function (see the text) as a function of relative primary level $20\log(A/B)$. The curves were computed using an index of compression $\beta=0.3$ dB/dB.

the order n. The left/right symmetry of the curves is a consequence of the symmetry $R(A,B)=R(B,A)$ displayed by Eq. (A3). The asymptotic behavior of the coefficients c_n for $A\gg B$ and for $B\gg A$ is dictated by $R\to 0$ in both cases, causing the factor $F(\cdots)$ in Eq. (A10) to approach unity. This leaves

$$c_n \to R^n(\alpha)_n/n!, \quad R\to 0. \tag{A15}$$

The two limiting cases can be further evaluated by means of Eqs. (A3b), (A3c) yielding

$$c_n \to (A/B)^n(\alpha)_n/n!, \quad A\ll B, \tag{A16a}$$

$$c_n \to (B/A)^n(\alpha)_n/n!, \quad B\ll A. \tag{A16b}$$

This accounts for the asymptotic slopes of $\pm n$ dB/dB of the c_n curves in Fig. 5. An immediate consequence of Eq. (16a) is

$$c_{n+1}/c_n \to (A/B)(n+1/2-\beta/2)/(n+1), \quad A\ll B. \tag{A17}$$

This result is used in the main text to estimate the relative amplitudes of the two contributions to a single DP (Secs. II C, II F).

[1]An exception is a system in which the gain is controlled by future stimulus fluctuations. Such a system would need two parallel input channels: a fast one for determining the gain and a delayed one for the signal to be controlled.

[2]It is hard to imagine a type of *instantaneous* nonlinearity that produces suppression but no CDTs. To the extent that the cochlear-mechanical measurements are made very close to the site of nonlinear interaction, it is unconvincing to blame the absence of CDTs to poor propagation or lack of "coupling."

[3]Delays due to "filtering" are implicit in the phase profile of the basilar membrane. A distinction between traveling wave delays and so-called filter delays can be misleading in that it incorrectly suggests the operation of a separate filter outside the traveling wave.

[4]The symmetry between S_k and T_{-k} is only valid within a single idealized AGC. It does not imply a general symmetry between lower and upper CDTs, because that is generally lost upon frequency-dependent propagation in the cochlea.

[5]In view of the normal dispersion of the cochlea, one expects $u\geqslant 0$, corresponding to the *higher* primary arriving *later* at the CDT generation site than the lower primary. Anomalous dispersion at low frequencies (<1 kHz in the cat; Pfeiffer and Molnar, 1970; van der Heijden and Joris, 2003), however, will cause u to become negative. The terms "normal dispersion" and "anomalous dispersion" correspond to a decrease, respectively, increase, of propagation velocity with frequency (Elmore and Heald, 1985).

[6]In principle, one could simply measure (e.g., on the BM) the phase of the primaries and of the CDT, and then check the validity of Eq. (13). In practice, however, the absolute phase of stimulus components in the inner ear is sensitive to a host of confounding factors such as sound level and the relation between stimulus frequency and characteristic frequency of the measurement site (Robles and Ruggero, 2001). A differential method like the cancellation paradigm eliminates these confounding factors.

[7]Interestingly, it is only at higher sound levels ($\geqslant$60 dB SPL) that harmonic distortion starts to become appreciable in psychophysics (Wegel and Lane, 1924) and cochlear mechanics (Cooper, 1998). This is expected when the integration time of the AGC drops to values comparable to the signal period: for such small values of integration time, the time-varying gain $G(t)$ must contain residuals of the signal fine-structure (i.e., signal frequencies and their harmonics) in addition to the stimulus envelope. The application of such a gain function $G(t)$ to the stimulus obviously results in harmonic distortion.

[8]Only if very small values are used ($\delta<0.01$) is the patterning of amplitude coefficients significantly affected and then only in the immediate neighborhood of equal primary amplitudes. Such small values of δ would mean that the nonlinearities were dominated by the small fraction of time that the theoretical (Hilbert) envelope is almost zero. This is unrealistic because any practical scheme to extract the stimulus envelope (such as a leaky integrator) will fill up the narrow valleys of the envelope.

Bian, L., Chertoff, M. E., and Yi, X. (**2002**). "Deriving a cochlear transducer function from low-frequency modulation of distortion product otoacoustic emissions," J. Acoust. Soc. Am. **112**, 198–210.

Bian, L., Linhardt, E. E., and Chertoff, M. E. (**2004**). "Cochlear hysteresis: Observation with low-frequency modulated distortion product otoacoustic emissions," J. Acoust. Soc. Am. **115**, 2159–2172.

Buunen, T. J. F., and Rhode, W. S. (**1978**). "Responses of fibers in the cat's auditory nerve to the cubic difference tone," J. Acoust. Soc. Am. **64**, 772–781.

Cooper, N. P. (**1996**). "Two-tone suppression in cochlear mechanics," J. Acoust. Soc. Am. **99**, 3087–3098.

Cooper, N. P. (**1998**). "Harmonic distortion on the basilar membrane in the basal turn of the guinea-pig cochlea," J. Physiol. (Paris) **509**, 277–288.

Cooper, N. P. (**2004**). "Compression in the peripheral auditory system," in *Compression: From Cochlea to Cochlear Implant*, edited by S. P. Bacon, R. R. Fay, and A. N. Popper (Springer, New York), pp. 18–61.

Duifhuis, H. (**1976**). "Cochlear nonlinearity and second filter: Possible mechanism and implications," J. Acoust. Soc. Am. **59**, 408–423.

Elmore, W. C., and Heald, M. A. (**1985**). *Physics of Waves* (Dover, New York), p. 123.

Geisler, C. D., and Greenberg, S. (**1986**). "A two-stage nonlinear cochlear model possesses automatic gain control," J. Acoust. Soc. Am. **80**, 1359–1363.

Geisler, C. D., and Nutall, A. (**1997**). "Two-tone suppression of basilar membrane vibrations in the base of the guinea pig cochlea using low-side suppressors," J. Acoust. Soc. Am. **102**, 430–440.

Goldstein, J. L. (**1967**). "Auditory nonlinearity," J. Acoust. Soc. Am. **41**, 676–689.

Goldstein, J. L., and Kiang, N. Y. S. (**1968**). "Neural correlates of the aural combination tone $2f_1-f_2$," Proc. IEEE **56**, 981–992.

Gradshteyn, I. S., and Ryzhik, I. M. (**1994**). *Table of Integrals, Series, and Products*, 5th ed. (Academic, New York), Chap. 9.1.

Hall, J. L. (**1972**). "Auditory distortion products f_2-f_1 and $2f_1-f_2$," J. Acoust. Soc. Am. **51**, 1863–1871.

Heitmann, J., Waldman, B., Schnitzel, H. U., Plinkert, P. K., and Zenner, H. P. (**1998**). "Suppression of distortion product otoacoustic emissions (DPOAE) near $2f_1-f_2$ removes DP-gram fine structure—evidence for a secondary generator," J. Acoust. Soc. Am. **103**, 1527–1531.

Kalluri, R., and Shera, C. A. (**2001**). "Distortion-product source unmixing: A test of the two-mechanism model for DPOAE generation," J. Acoust. Soc. Am. **109**, 622–637.

Kemp, D. T., and Brown A. M. (**1983**). "An integrated view of cochlear mechanical nonlinearities observable from the ear canal," in *Cochlear Mechanics*, edited by E. de Boer and M. A. Viergever (Delft University Press, Holland).

Knight, R. D., and Kemp, D. T. (**2000**). "Indications of different distortion

product otoacoustic emission mechanisms from a detailed f_1, f_2 area study,'' J. Acoust. Soc. Am. **107**, 457–473.
Lewis, D., and Larsen, M. J. (**1937**). ''The cancellation, reinforcement and measurement of subjective tones,'' Proc. Natl. Acad. Sci. U.S.A. **23**, 415–421.
Lyon, R. F. (**1990**). ''Automatic gain control in cochlear mechanics,'' in *Mechanics and Biophysics of Hearing*, edited by P. Dallos, C. D. Geisler, J. W. Matthews, M. A. Ruggero, and C. R. Steele (University of Wisconsin Press, Madison, WI), pp. 395–402.
Manley, G. A., Köppl, C., and Johnstone, B. M. (**1993**). ''Distortion-product otoacoustic emissions in the bobtail lizard. I. General characteristics,'' J. Acoust. Soc. Am. **93**, 2820–2833.
Patuzzi, R. (**1996**). ''Cochlear micromechanics and macromechanics,'' in *The Cochlea*, edited by P. Dallos, A. N. Popper, and R. R. Fay (Springer, New York), pp. 186–257.
Pfeiffer, R. R., and Molnar, C. E. (**1970**). ''Cochlear nerve fiber discharge patterns: Relationship to the cochlear microphonic,'' Science **167**, 1614–1616.
Rhode, W. S. (**1971**). ''Observation of the vibration of the basilar membrane in squirrel monkeys using the Mössbauer effect,'' J. Acoust. Soc. Am. **49**, 1218–1231.
Robles, L., and Ruggero, M. A. (**2001**). ''Mechanics of the mammalian cochlea,'' Physiol. Rev. **81**, 1305–1352.
Ruggero, M. A., and Rich, N. C. (**1983**). ''Time course of suppression of spontaneous otoacoustic emissions,'' J. Acoust. Soc. Am. **73**, S60 (abstract).
Schroeder, M. R. (**1969**). ''Relation between critical bands in hearing and the phase characteristics of cubic difference tones,'' J. Acoust. Soc. Am. **46**, 1488–1492.
Schneider, S., Prijs, V. F., and Schoonhoven, R. (**2003**). ''Amplitude and phase of distortion product otoacoustic emissions in the guinea pig in an (f_1, f_2) area study,'' J. Acoust. Soc. Am. **113**, 3285–3296.
Shera, C. A., Talmadge, C. L., and Tubis, A. (**2000**). ''Interrelations among distortion-product phase-gradient delays: Their connection to scaling symmetry and its breaking,'' J. Acoust. Soc. Am. **108**, 2933–2948.
Shera, C. A. (**2003**). ''Wave interference in the generation of reflection- and distortion-source emissions,'' in *Biophysics of the Cochlea: From Molecules to Models*, edited by A. W. Gummer (World Scientific, Singapore), pp. 439–453.
Smoorenburg, G. F. (**1972**). ''Combination tones and their origin,'' J. Acoust. Soc. Am. **52**, 615–632.
Talmadge, C. L., Long, G. R., Tubis, A., and Dhar, S. (**1999**). ''Experimental confirmation of the two-source interference model for the fine structure of distortion product otoacoustic emissions,'' J. Acoust. Soc. Am. **105**, 275–292.
van der Heijden, M., and Joris, P. X. (**2003**). ''Cochlear phase and amplitude retrieved from the auditory nerve at arbitrary frequencies,'' J. Neurosci. **23**, 9194–9198.
van der Heijden, M., and Joris, P. X. (**2005**). ''The speed of auditory low-side suppression,'' J. Neurophysiol. **93**, 201–209
Wegel, R. L., and Lane, C. E. (**1924**). ''The auditory masking of one pure tone by another and its probable relation to the dynamics of the inner ear,'' Phys. Rev. **23**, 266–285.
Weiss, T. F., and Leong, R. (**1985**). ''A model for signal transmission in an ear having cells with free-standing stereocilia. IV. Mechanoelectric transduction stage,'' Hear. Res. **20**, 175–195.
Wilson, J. P. (**1980**). ''The combination tone, $2f_1-f_2$, in psychophysics and ear-canal recording,'' in *Psychophysical, Physiological and Behavioral Studies in Hearing*, edited by G. van den Brink and F. A. Bilsen (Delft University Press, Delft, The Netherlands), pp. 43–52.
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. (**1955**). ''Der ungewöhnliche Amplitudengang der nichtlinearen Verzerrungen des Ohres'' (The remarkable amplitude behavior of auditory nonlinear distortions), Acustica **5**, 67–74.
Zwicker, E. (**1980a**). ''Cubic difference tone level and phase dependence on frequency difference and level of primaries,'' in *Psychophysical, Physiological and Behavioral Studies in Hearing*, edited by G. van den Brink and F. A. Bilsen (Delft University Press, Delft, The Netherlands), pp. 286–273.
Zwicker, E. (**1980b**). ''Nonmonotonic behavior of $(2f_1-f_2)$ explained by a saturation-feedback model,'' Hear. Res. **2**, 513–518.
Zwicker, E. (**1981**). ''Dependence of level and phase of the $(2f_1-f_2)$-cancellation tone on frequency range, frequency difference, level of primaries, and subject,'' J. Acoust. Soc. Am. **70**, 1277–1288.
Zwislocki, J. J., Szymko, Y. M., and Hertig, L. Y. (**1996**). ''The cochlea is an automatic gain control system after all,'' in *Diversity in Auditory Mechanics*, edited by E. Lewis, G. R., Long, R. F. Lyon, P. M. Narins, C. R. Steele, and E. Hecht-Poinar (World Scientific, Singapore), pp. 621–627.

On the large-scale spectral structure of otoacoustic emissions

Renata Sisto[a)]
Dipartimento Igiene del Lavoro, ISPESL, Via Fontana Candida, 1, 00040 Monte Porzio Catone (Roma), Italy

Arturo Moleti[b)]
Dipartimento di Fisica, Università di Roma "Tor Vergata," Via della Ricerca Scientifica, 1, 00133 Roma, Italy

(Received 5 October 2004; revised 3 December 2004; accepted 3 December 2004)

Transient evoked and distortion product otoacoustic emission data, showing a characteristic slowly oscillating spectral shape, are presented. Such peculiar behavior had also been observed in earlier studies, and deserves some theoretical explanation. A simple model of the cochlear reflectivity, based on the analogy between the cochlear transmission line equations and the Schrödinger wave equation for the motion of an elementary particle above a one-dimensional potential well, is presented. Wave mechanics predicts indeed reflection from a negative potential well, which is quasiperiodically dependent on the width and depth of the well, i.e., on the quality factor of the cochlear resonance. The model, whose quantitative predictions are dependent on the rather uncertain level and slope of the cochlear tuning curve, proves capable of explaining, at least qualitatively, the observed experimental behavior. © *2005 Acoustical Society of America.*
[DOI: 10.1121/1.1853208]

PACS numbers: 43.64.Jb [BLM] Pages: 1234–1240

I. INTRODUCTION

Otoacoustic emissions (OAEs) are a by-product of the active mechanisms of the inner ear. Their study is important to get information on the physiology and pathology of the cochlea (Probst *et al.*, 1991). A spectral fine structure has been observed both in spontaneous OAEs (SOAEs) and in distortion product OAEs (DPOAEs) (see, e.g., Talmadge *et al.*, 1998; Mauermann *et al.*, 1999a, b). This structure is associated with peculiar aspects of cochlear mechanics, and its analysis has been effectively used to test cochlear models. DPOAE's fine structure is generally assumed to be due to interference between two different cochlear sources (Talmadge *et al.*, 1999). A possible explanation of SOAE quasiperiodicity was provided by Zweig and Shera (1995) as being due to a coherent reflection mechanism, which would favor those frequencies matching a periodic standing wave condition (Talmadge *et al.*, 1998). The frequency spacing of multiple SOAEs has also been associated with the critical band (CB) and an interesting structure has been observed (Braun, 1997) within CB and 2CB, favoring low-order frequency ratios (5:4, 6:5).

In addition to these ''small-scale'' spectral features (fine structure), TEOAE and DPOAE spectral responses also reveal some ''large-scale'' structure, with a characteristic dip around 2–3 kHz (e.g., Smurzynski and Kim, 1992; Gorga *et al.*, 1993; Kummer *et al.*, 1998). Possible explanations were proposed, in the case of DPOAEs, involving calibration of the stimuli in the ear canal, the effect of standing waves in the sealed ear canal (Siegel and Hirohata, 1994), or individual variations of middle-ear and inner-ear resonances (Lonsbury-Martin *et al.*, 1990). In this work, even if we cannot rule out other possible explanations, we suggest instead that this peculiar OAE large-scale spectral structure could find a simple explanation within very general cochlear transmission and reflection models. We present new TEOAE and DPOAE data of normal-hearing and hearing-impaired ears confirming this spectral behavior, and provide a possible explanation for it, starting from the equation describing the cochlear transmission as that of an equivalent electrical transmission line. We note that the differential equation for the cochlear transmission line is formally equivalent to the Schrödinger wave equation describing the motion of a particle in a one-dimensional potential. The transverse impedance minimum at the resonant place is equivalent to a potential well, and the initial energy of the particle, which is only kinetic, is equivalent to the square of the initial wave vector amplitude. Reflection from the well can be computed as a function of the particle energy and of the well depth, which is related to the quality factor of the cochlear resonance. As it is well known from elementary quantum mechanics (Landau and Lifshitz, 1977), reflection from a potential well is quasiperiodically dependent upon the kinetic energy of the incoming particle and the well depth and width. As a consequence, in our equivalent cochlear problem, the combination of the dependence on frequency of the initial wave number and of the quality factor produces a characteristic oscillating dependence on frequency of the reflected fraction of the incoming acoustic wave. Although absorption at the resonance place should also be taken into account in a more complete model, the frequency of the reflection maxima and minima can be approximately determined even neglecting absorption. In the Results section, the model is applied to the analysis of experimental TEOAE and DPOAE spectral responses, showing that the observed spectral shape could be effectively analyzed within this framework.

[a)]Electronic mail: r.sisto@dil.ispesl.it
[b)]Electronic mail: arturo.moleti@roma2.infn.it

II. METHODS

Audiometric recordings and TEOAEs and DPOAEs of a population of 83 adult male subjects exposed to noise, including normal-hearing and hearing-impaired subjects, have been analyzed. These data were collected within the framework of the Finalized Project on Otoacoustic Emissions of the Italian Ministry of Health, aiming at proving the effectiveness of OAE techniques to detect and monitor mild cochlear damage in workers exposed to noise. Pure-tone audiograms were recorded, for all subjects, in an acoustically shielded room. The audiometric test frequencies were 0.25, 0.5, 1, 2, 3, 4, 6, and 8 kHz. The population was divided in three different audiometric classes, according to the measured hearing threshold. The ear was defined as ''normal'' if no absolute threshold level higher than 10 dB was measured over the whole frequency range (96 ears). If a threshold higher than 20 dB was observed in any audiometric range, the ear was defined as ''impaired'' (26 ears). An intermediate set of normal-hearing subjects was identified with maximum threshold level between 10 and 20 dB, defined in this work as ''mild hearing loss'' (44 ears).

TEOAE and DPOAE recordings were obtained with the ILO-96 system (Otodynamics, Ltd.). Probe calibration was repeated at the beginning of every measurement session, using the standard ILO 2cc cavity.

TEOAEs were recorded in the standard ''derived nonlinear'' mode, using a stimulus amplitude of (80 ± 3) dB and the standard ILO trapezoidal window.

DPOAEs were recorded with a stimulus frequency ratio $f_2/f_1=1.22$, and levels $L_1=65$ dB and $L_2=55$ dB, in the range 500–8000 Hz, with 1/3-octave resolution. Actually, only in 102 of the total of 166 ears, DPOAEs were recorded using exactly the 65–55-dB stimulus paradigm. In the other ears examined, slightly different stimulus levels were used. Although the results are very similar, to avoid confusion only these 102 DPOAE recordings were used in the following analysis.

TEOAE and DPOAE absolute response amplitudes were averaged over the ears of the three audiometric classes. The TEOAE waveforms of the normal-hearing ears have also been analyzed to measure the OAE spectral latency, to get an estimate of the average cochlear tuning curve, according to the procedures described in Sisto and Moleti (2002) and Moleti and Sisto (2003).

III. MODEL

We refer to a simple model of the cochlear transmission described as a transmission line, which is locally resonant at different frequencies, according to a tonotopic map (Greenwood, 1990) relating frequency and longitudinal position x along the cochlea. Within the framework of the transmission line formalism, the differential equation describing the motion of the traveling wave has the form

$$\frac{\partial^2 P_d(x,\omega)}{\partial x^2}=\frac{-k_0^2\omega^2}{\Delta(x,\omega)}P_d(x,\omega), \tag{1}$$

where P_d is the differential pressure applied to the cochlear membrane, k_0 is a constant dependent on density and geometry of the basilar membrane and cochlear cavity, and $1/\Delta(x,\omega)$ is a tonotopically resonant function (see, e.g., Talmadge *et al.*, 1998, Sisto and Moleti, 2002).

This wave equation is formally equivalent to the Schrödinger equation for the motion of a particle in a one-dimensional potential

$$-\frac{\hbar^2}{2m}\frac{\partial^2\psi}{\partial x^2}+U(x)\psi=E\psi. \tag{2}$$

The increase of the absolute value of the (negative) second member of Eq. (1) approaching the resonant place is equivalent to a potential well $U(x)$, and the initial energy of the particle, which is only kinetic, is equivalent to the square of the wave vector initial amplitude.

As the particle kinetic energy increases approaching the potential well, similarly, in the cochlea, for a wave of given frequency, the wave vector increases (the wavelength decreases) approaching the resonant place for that frequency. This analogy may be useful because the problem of one-dimensional motion of a quantum particle has been extensively studied, and analytical solutions of many different problems are already available in the literature. Of course, in the transmission line equation there is also the imaginary part of the transverse impedance, associated with absorption near the resonance place. This dissipative part of the equations is not taken into account in the simplest version of this model, because we wish to explain the main relative features of the OAE spectral structure, but it should be considered if a quantitative estimate of the reflection coefficient is needed to predict also the absolute OAE level. It can be shown that inclusion of dissipative effects leads to prediction of the typical traveling wave behavior, with a strong damping approaching the resonant place, where the impedance becomes imaginary. Neglecting dissipation, it is still possible to predict the spectral shape of the reflectivity function.

A. Semiclassical solution for the transmission line models

In full cochlear model based on resonant transmission lines, the basic equations of motion are

$$\frac{\partial^2 P_d(x,t)}{\partial x^2}=k_0^2\sigma_{\rm bm}\frac{\partial^2\xi(x,t)}{\partial t^2}, \tag{3}$$

$$\frac{\partial^2\xi(x,t)}{\partial t^2}+\Gamma(x,\xi)\frac{\partial\xi(x,t)}{\partial t}+\omega^2(x)\xi(x,t)=\frac{P_d(x,t)}{\sigma_{\rm bm}}, \tag{4}$$

where P_d is the differential pressure applied to the basilar membrane, ξ is the transverse displacement along the membrane, Γ is a (generally nonlinearly dependent on ξ) damping function, and $\sigma_{\rm bm}$ the basilar-membrane surface density. The line is locally resonant at frequency $\omega(x)$ at each site x along the basilar membrane. The relation between position x and the resonance frequency is given by the Greenwood map (Greenwood, 1990)

$$\omega(x)=\omega_1+\omega_{\max}e^{-k_\omega x}, \tag{5}$$

where $\omega_{\max}=2\pi\cdot20655$ rad/s, $\omega_1=2\pi\cdot145$ rad/s, and $k_\omega=1.382\ \text{cm}^{-1}$.

Equation (3) describes the propagation of the differential pressure applied to the basilar membrane. Equation (4) is that of a resonant oscillator forced by the differential pressure P_d. In general, the oscillator described by Eq. (4) belongs to a class of nonlinear limit-cycle oscillators. The most studied of these oscillators is the Van der Pol oscillator, in which the nonlinearity is in the term proportional to $\dot{\xi}$ and gives both the saturation of the response at high stimulus levels and the existence of equilibrium oscillations associated with spontaneous emissions.

Neglecting all nonlinear terms, the basic linearized equations in the frequency domain (Talmadge *et al.*, 1998) become

$$\xi(x,\omega)=\frac{1}{\sigma_{\mathrm{bm}}\Delta(x,\omega)}P_d(x,\omega), \tag{6}$$

$$\Delta(x,\omega)=\omega^2(x)-\omega^2+i\omega\Gamma, \tag{7}$$

$$\frac{\partial^2 P_d(x,\omega)}{\partial x^2}=-k_0^2\sigma_{\mathrm{bm}}\omega^2\xi(x,\omega)=\frac{-k_0^2\omega^2}{\Delta(x,\omega)}P_d(x,\omega)$$
$$=-k^2(x,\omega)P_d(x,\omega). \tag{8}$$

Note that ω is the frequency of the considered Fourier component, while ω(x) is the local resonance frequency at the place x along the basilar membrane. The first equation represents the relation between the potential and the current for each frequency ω as a function of x. The denominator is a resonant function of the line transverse impedance for each frequency at each site x. For the frequency ω, the real part of the impedance is zero at the resonant place $x(\omega)=\hat{x}$.

Let us consider Eq. (8). This equation is formally identical to the Schrödinger equation of a particle of mass m moving in the potential

$$U(x,\omega)=\frac{-k_0^2\omega^2}{\Delta(x,\omega)}=\frac{-k_0^2\omega^2}{\omega^2(x)-\omega^2+i\omega\Gamma}, \tag{9}$$

$$-\frac{\hbar^2}{2m}\frac{\partial^2 P_d}{\partial x^2}+U(x,\omega)P_d=EP_d. \tag{10}$$

The total energy of the particle is $E=0$. In our units $\hbar=1$, $m=1/2$.

The real part of the potential accounts for the oscillating behavior of the solution while the imaginary part represents the dissipative behavior, characterized by exponentially decaying amplitude.

If we look for semiclassical solutions (which are valid in the limit in which the spatial derivative of the wave vector can be neglected with respect to the phase variation of the solution) of the type

$$P_d(x,\omega)=e^{-i\int dx' k(\omega,x')}P_d(\omega), \tag{11}$$

a mathematical relation between the wave vector k, frequency ω, and position x along the basilar membrane is obtained, showing the characteristic tonotopically resonant behavior

$$k(\omega,x)=\frac{\omega k_0}{\sqrt{\omega^2(x)-\omega^2+i\omega\Gamma(x)}}, \tag{12}$$

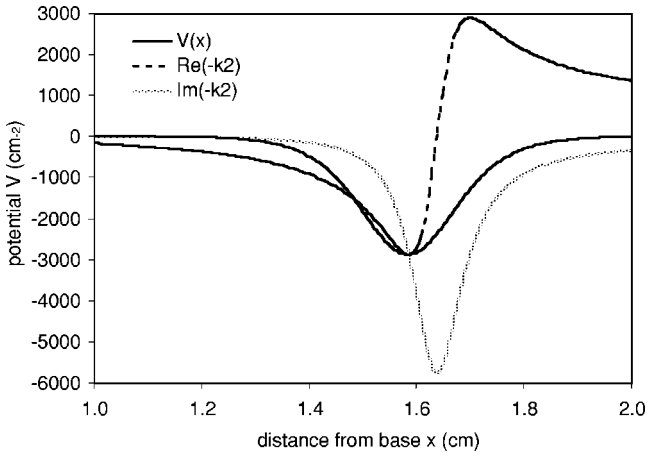


FIG. 1. Potential $U(x)=-\mathrm{Re}(k^2)$, and its symmetric approximation $V(x)$, plotted against the cochlear position x, for a frequency $f=\omega/2\pi=2$ kHz. The imaginary part of k^2 is also shown.

$$P_d(x,\omega)=e^{-\int dx' \mathrm{Im}\, k(x',\omega)}e^{-i\int dx' \mathrm{Re}\, k(\omega,x')}P_d(\omega). \tag{13}$$

In the semiclassical limit

$$k^2(x,\omega)=E-U=\frac{k_0^2\omega^2}{\Delta(x,\omega)}, \tag{14}$$

$$k^2=|k^2|e^{i\Phi}, \tag{15}$$

$$|k^2|=\frac{k_0^2\omega^2}{\sqrt{(\omega^2(x)-\omega^2)^2+\omega^2\Gamma^2}}, \tag{16}$$

$$\Phi=\arctan\left\{\frac{\mathrm{Im}(k^2)}{\mathrm{Re}(k^2)}\right\}=\arctan\left\{\frac{-\omega\Gamma}{\omega^2(x)-\omega^2}\right\}, \tag{17}$$

$$\mathrm{Re}(k)=\sqrt{|k^2|}\cos\left(\frac{\Phi}{2}\right), \tag{18}$$

$$\mathrm{Im}(k)=\sqrt{|k^2|}\sin\left(\frac{\Phi}{2}\right). \tag{19}$$

In Fig. 1 we show the real and imaginary parts of k^2, plotted as a function of the cochlear position x for a frequency $f=\omega/2\pi=2$ kHz.

B. Reflection from the potential well

Let us now consider the Schrödinger Eq. (10) with

$$U(x,\omega)=-\mathrm{Re}(k^2). \tag{20}$$

The total energy E is zero and, up to the resonant place, the potential is negative, so the particle is moving in a potential well with an energy that is higher than the potential level.

Here, we neglect for simplicity the dissipative part of the potential

$$E>U(x,\omega). \tag{21}$$

The limits of the potential are

$$U(x,\omega)\to-k_0^2\frac{\omega^2}{\omega_{\max}^2}\quad\text{for}\quad x\to 0, \tag{22}$$

$$U(x,\omega)=U_{\min}(\omega)=-\frac{k_0^2 Q}{2}=-\frac{\hat{k}^2}{2}\quad \text{for } x$$

$$=x(\omega+\Gamma/2)=\hat{x}-\frac{1}{2k_\omega Q}, \tag{23}$$

$$U(x,\omega)=\frac{U_{\min}(\omega)}{2}=-\frac{k_0^2 Q}{4}\quad \text{for } x$$

$$=x(\omega+2\Gamma)=\hat{x}-\frac{2}{k_\omega Q}, \tag{24}$$

$$U(x,\omega)=0\quad \text{for } x=\hat{x}, \tag{25}$$

$$U(x,\omega)\to k_0^2\quad \text{for } x\gg\hat{x}, \tag{26}$$

where $\hat{k}$ is the modulus of the wave vector at resonance and Q is the quality factor at frequency ω

$$Q=\frac{\omega}{\Gamma};\quad \hat{k}=k_0\sqrt{Q};\quad \Delta\hat{x}=\frac{1}{k_\omega Q}. \tag{27}$$

In the region between $\hat{x}-2\Delta\hat{x}$, where the potential is half its minimum value, and $\hat{x}$, where the potential crosses zero, a well is present [see Eqs. (23–25)]. In our quantum-mechanical analogy, $\Delta\hat{x}$ is the well width. From Eq. (22) we see that, for $x\to 0$ (base), the potential approaches from below a negative asymptotic value, which is close to zero if $\omega\ll\omega_{\max}$. Neglecting to compute accurately the limit of the potential near $x=\hat{x}$ (in the region close to $\hat{x}$ the modulus of the imaginary part of k^2 abruptly increases, so the amplitude of resonant component of the traveling wave is rapidly damped), we can approximate the potential U with a symmetric function of the type

$$\tilde{U}(x,\omega)=-\frac{\frac{\hat{k}^2}{2}-\left(k_0\frac{\omega}{\omega_{\max}}\right)^2}{\cosh^2\alpha(x-(\hat{x}-\Delta\hat{x}))}-\left(k_0\frac{\omega}{\omega_{\max}}\right)^2, \tag{28}$$

with

$$\alpha=\frac{1}{\Delta\hat{x}}=k_\omega Q. \tag{29}$$

By doing so, we neglect the asymmetry of the potential well, which probably could be better described by a Morse-type potential. This is beyond the scope of the present work, but could be done in a more accurate computation also including dissipative effects.

We prefer to put the potential offset in the second member of Eq. (11), redefining the potential and the total energy, which is initially kinetic only

$$V(x,\omega)=-\frac{\frac{\hat{k}^2}{2}-\left(k_0\frac{\omega}{\omega_{\max}}\right)^2}{\cosh^2\alpha(x-(\hat{x}-\Delta\hat{x}))}, \tag{30}$$

$$E=k_0^2\frac{\omega^2}{\omega_{\max}^2}. \tag{31}$$

In Fig. 1 we show the symmetric potential $V(x)$, which we use in the following, compared to the potential $-\mathrm{Re}(k^2)$.

If we consider an incoming particle with wave number

$$k=\sqrt{E}=k_0\frac{\omega}{\omega_{\max}}, \tag{32}$$

the transmission coefficient from the potential well of Eq. (30) is analytically computed (Landau and Lifshitz, 1977)

$$D=\frac{\sinh^2(\pi k/\alpha)}{\sinh^2(\pi k/\alpha)+\cos^2\left(\frac{\pi}{2}\sqrt{1+\frac{4V_{\min}}{\alpha^2}}\right)}=\frac{\sinh^2\left(\frac{\pi k_0\omega}{k_\omega\omega_{\max}Q}\right)}{\sinh^2\left(\frac{\pi k_0\omega}{k_\omega\omega_{\max}Q}\right)+\cos^2\left(\frac{\pi}{2}\sqrt{1+\frac{2k_0^2}{k_\omega^2 Q}\left(1-\left(\frac{2\omega^2}{\omega_{\max}^2 Q}\right)\right)}\right)}, \tag{33}$$

where

$$V_{\min}=-\frac{\hat{k}^2}{2}\left(1-\frac{2\omega^2}{\omega_{\max}^2 Q}\right) \tag{34}$$

is the depth of the potential well, whose minimum occurs at the cochlear position $\hat{x}-(\Delta\hat{x}/2)$.
The reflection coefficient is

$$R=1-D=\frac{\cos^2\left(\frac{\pi}{2}\sqrt{1+\frac{2k_0^2}{k_\omega^2 Q}\left(1-\left(\frac{2\omega^2}{\omega_{\max}^2 Q}\right)\right)}\right)}{\sinh^2\left(\frac{\pi k_0\omega}{k_\omega\omega_{\max}Q}\right)+\cos^2\left(\frac{\pi}{2}\sqrt{1+\frac{2k_0^2}{k_\omega^2 Q}\left(1-\left(\frac{2\omega^2}{\omega_{\max}^2 Q}\right)\right)}\right)}. \tag{35}$$

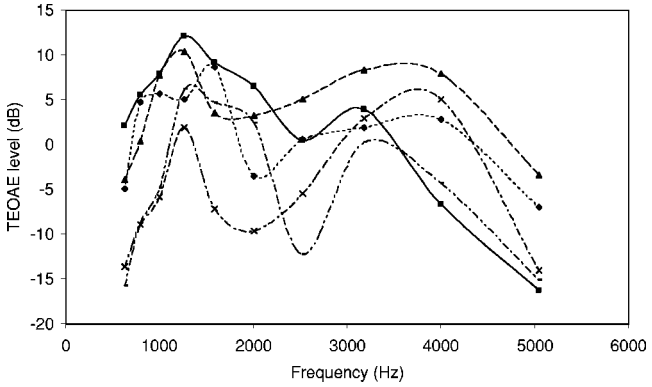


FIG. 2. TEOAE response of a set of five normal ears, with third-octave spectral resolution. A characteristic dip around $f_2=2$ kHz is visible in most of the examined ears.

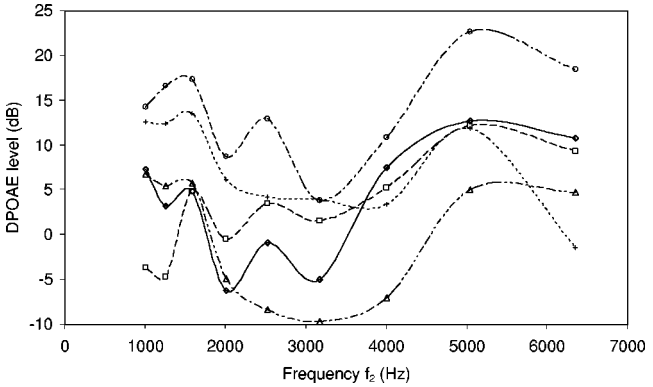


FIG. 3. DPOAE response of a set of five normal ears, with third-octave spectral resolution. The response at frequency $f_{DP}=2\,f_1-f_2$ is plotted as a function of f_2, as usually done in the literature. A characteristic dip around $f_2=3$ kHz is visible in most of the examined ears.

Neglecting the slow dependence on frequency of the $\sinh^2$ term in Eq. (35), local maxima of the reflectivity occur where the following condition is matched:

$$\cos^2\left(\frac{\pi}{2}\sqrt{1+\frac{2k_0^2}{k_\omega^2 Q}\left(1-\left(\frac{2\omega^2}{\omega_{\max}^2 Q}\right)\right)}\right)=1. \tag{36}$$

For $\omega \ll \omega_{\max}\sqrt{Q}$, this implies a simple condition on Q

$$Q \approx \frac{k_0^2}{2k_\omega^2 n^2\left(1-\frac{1}{4n^2}\right)}, \tag{37}$$

where n is any integer. Using the value of $k_0=31\ \mathrm{cm}^{-1}$ (Talmadge *et al.*, 1998), Eq. (37) can be rewritten as

$$Q \approx \frac{250}{n^2\left(1-\frac{1}{4n^2}\right)}. \tag{38}$$

Hence, we expect reflectivity maxima for $Q=(\ldots,64,41,28,21,15.5,12.5,10,8,7,6,5.1,4.5,\ldots)$. As Q is a function of frequency, varying roughly in the range between 3 and 20 in the OAE frequency range 1–5 kHz (Moleti and Sisto, 2003), a few maxima of the reflectivity should be present in this range. The exact position and spacing of the maxima are slightly different, if all terms are correctly taken into account, and they are dependent on the height and slope of the tuning curve, which cannot be accurately estimated. What is important is that a slowly oscillating structure is, however, predicted for the reflectivity spectral shape.

We remark that this model does not include randomly distributed mechanical irregularities, which are considered as ''place-fixed'' physical scattering centers for coherent reflection in some cochlear models (Zweig and Shera, 1995; Talmadge *et al.*, 1998). The present model in this simple form implies ''wave-fixed'' reflection as a consequence of the resonant nature of the transmission line (for a discussion on wave-fixed and place-fixed models, see, e.g., Moulin and Kemp, 1996). Micromechanical irregularities could also be included in a composite ''intermediate'' model, but here we are interested in the prediction of quasiperiodic spectral shape given by the present model in its simplest form.

IV. RESULTS AND DISCUSSION

In this section, experimental TEOAE and DPOAE data are shown and compared to the predictions of the model. Individual TEOAE and DPOAE responses show irregular patterns, with some common large-scale features, namely a dip in the midfrequency range (2–3 kHz). This is shown for TEOAEs and DPOAEs in Fig. 2 and Fig. 3, respectively, for small subsets of normal-hearing ears. For TEOAEs, the cutoff around 5 kHz is due to the limited extension of the spectrum of the click stimulus used. Averaged responses show much more clearly the large-scale structure, because the small-scale irregularities associated with noise and with the individual OAE response tend to disappear. The typical behavior of the average TEOAE response is shown in Fig. 4, where the average response of 166 ears is shown with third-octave resolution. The three audiometric classes of ears show different levels of the response, as discussed in many previous studies (e.g., Lucertini *et al.*, 2002). For the purposes of this work, the interesting feature is that all of them show a similar spectral pattern, with a dip in the middle-frequency range, around 2–3 kHz. A different, but similar, pattern is shown in Fig. 5, where the average DPOAE response of 102

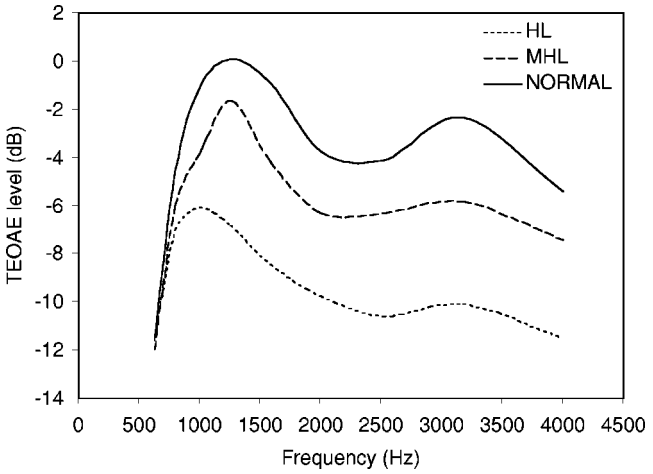


FIG. 4. TEOAE average response of the three classes of ears, with third-octave spectral resolution. The same spectral structure is present in all three classes, while the response level is significantly different.

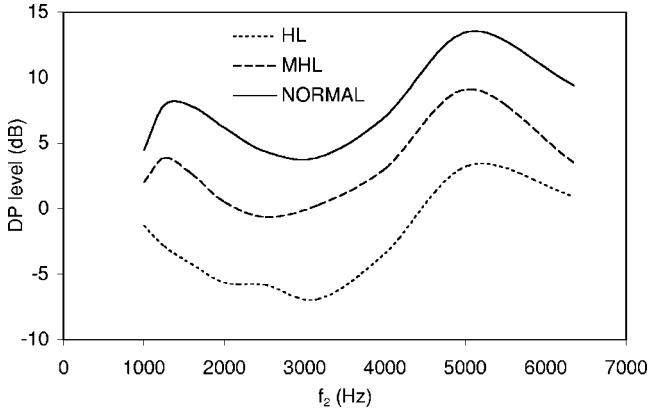


FIG. 5. DPOAE average response of the three classes of ears, with third-octave spectral resolution. The response at frequency $f_{DP}=2\,f_1-f_2$ is plotted as a function of f_2. A slowly oscillating spectral pattern is present, at different levels of the response, for all three classes, as in Fig. 4.

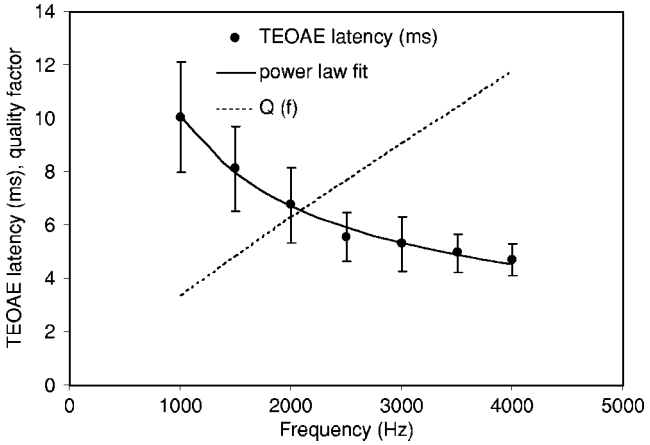


FIG. 7. TEOAE spectral latency, averaged over the population of normal ears, and cochlear tuning curve, estimated from the latency data. Error bars represent 1 standard deviation.

of the 166 ears is shown, with the same frequency resolution. The response at frequency $2\,f_1-f_2$ is plotted as a function of f_2, as usual.

In the case of DPOAEs, it should be stressed that the DPOAE response is made up of two contributions: nonlinear generation of the combination tone at the f_2 place and reflection at the f_{DP} place. The present model should predict the frequency dependence of the power reflected from the f_{DP} place, so the usual plot of the DPOAE response as a function of f_2 should be replaced, for this purpose, with a plot representing the response amplitude as a function of f_{DP}. The comparison with the model predictions is shown in Fig. 6. An experimental tuning curve was used, obtained from TEOAE latency estimates of the same subjects. We note that the reflection coefficient of Eq. (35) depends on the quality factor, which is dependent on the stimulus amplitude. Therefore, the comparison with the model should be done using appropriate values for the tuning curve. The latency estimates were based on time-frequency wavelet analysis of the TEOAE data, according to the procedures described in Sisto and Moleti (2002), while the tuning estimates were obtained an analysis technique based on a model of the cochlear transmission (Moleti and Sisto, 2003). The latency data and the tuning curve are shown in Fig. 7, where the standard deviation of the measured latencies is also reported. These estimates are rather indirect and affected by large uncertainties; thus, the comparison shown in Fig. 6 should be considered as a qualitative result. Indeed, the agreement with the data is dependent on the chosen tuning curve function. We did not attempt to obtain a better fit to the data by modifying the tuning curve, which would obviously be a quite useless task. The interesting result, both for TEOAEs and DPOAEs, is that the model predicts a spectral structure characterized by slow oscillations that is very similar to the structure of the experimental OAE spectra. This is not yet a proof of the validity of the model, but encourages us to do further work in this direction to improve it, including dissipative effects, in order to permit quantitative comparisons with experimental data. This result also suggests trying to improve the accuracy of the estimates of cochlear tuning, working on both the data acquisition and data analysis techniques. The data presented in this work come from standard DPOAE and TEOAE clinical tests, using rather high levels of stimulation, with the advantages of a large available database and of a good signal-to-noise ratio. In further experiments, it also would be useful to use lower levels of stimulation for both DPOAEs and TEOAEs to test the dependence of the results on the level of stimulation. Indeed, the quality factor of the cochlear resonances is dependent on the stimulus level, and low-level DPOAEs would be relatively more sensitive to the

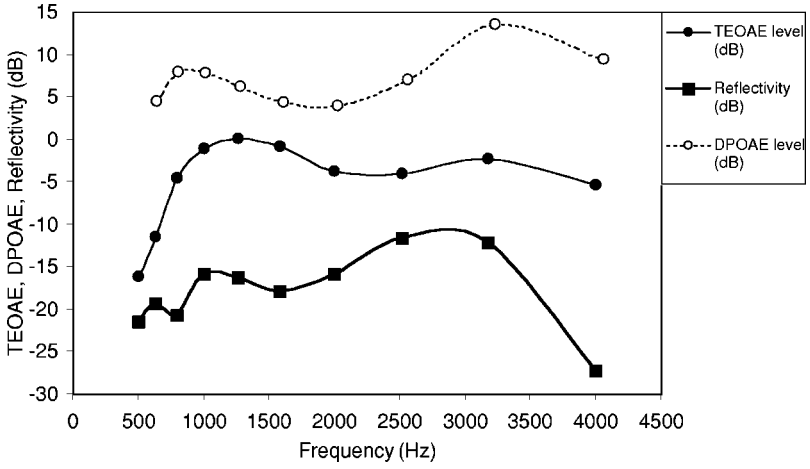


FIG. 6. Comparison between the experimental average DPOAE and TEOAE spectral responses of normal ears and the reflection coefficient (expressed in dB) predicted by the model of the present work. In this figure, the response at frequency $f_{DP}=2\,f_1-f_2$ is plotted as a function of f_{DP}, instead of f_2.

reflection source, which is assumed here to be responsible for the observed spectral oscillations.

V. CONCLUSIONS

A simple model for OAE emission by reflection of the cochlear traveling wave has been used to explain the slowly oscillating spectral structure observed in the TEOAE and DPOAE spectra, when averaged over a large number of ears. The model exploits a formal analogy between the transmission line equation and the Schrödinger equation for a particle moving above a one-dimensional potential well. The discovery of this analogy permits us to use a large set of analytical solutions that is easily available in quantum mechanics literature.

The comparison with experimental DPOAE and TEOAE spectra shows that the model predictions qualitatively agree with the observed spectra. The effect of absorption near the resonance place, which has been neglected here for simplicity, should strongly reduce the predicted reflection amplitude, leaving almost unchanged the frequencies of maxima and minima of the predicted spectral oscillations.

As alternative explanations of the observed spectral structure cannot be ruled out by this experiment, we suggest repeating our observations with a better control on probe and speaker calibration. Measurements at different stimulus levels would also give important information, because the cochlear tuning curve is dependent on the level of stimulation.

ACKNOWLEDGMENTS

This work was supported by ISPESL, in the framework of the Finalized Project on Otoacoustic Emissions of the Italian Ministry of Health. We wish to thank Dr. S. Pellegrini and Mrs. S. Chelotti for having provided the experimental data.

Braun, M. (**1997**). ''Frequency spacing of multiple spontaneous otoacoustic emissions shows relation to critical bands: A large-scale cumulative study,'' Hear. Res. **114**, 197–203.

Gorga, M. P., Neely, S. T., Bergman, B., Beauchaine, K. L., Kaminski, J. R., Peters, J., and Jesteadt, W. (**1993**). ''Otoacoustic emissions from normal-hearing and hearing-impaired subjects: Distortion product responses,'' J. Acoust. Soc. Am. **93**, 2050–2060.

Greenwood, D. D. (**1990**). ''A cochlear frequency position function for several species—29 years later,'' J. Acoust. Soc. Am. **87**, 2592–2605.

Kummer, P., Janssen, T., and Arnold, W. (**1998**). ''The level and growth behavior of the 2 f_1-f_2 distortion-product otoacoustic emission and its relationship to auditory sensitivity in normal hearing and cochlear hearing loss,'' J. Acoust. Soc. Am. **103**, 3431–3444.

Landau, L. D., and Lifshitz, E. M. (**1977**). ''*Quantum Mechanics (Non-Relativistic Theory)*,'' 3rd ed. (Pergamon, Oxford, England).

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.

Lucertini, M., Moleti, A., and Sisto, R. (**2002**). ''On the detection of early cochlear damage by otoacoustic emission analysis,'' J. Acoust. Soc. Am. **111**, 972–978.

Mauermann, M., Uppenkamp, S., van Hengel, P. W., and Kollmeier, B. (**1999a**). ''Evidence for the distortion product frequency place as a source of distortion product otoacoustic emission (DPOAE) fine structure in humans. I. Fine structure and higher-order DPOAE as a function of the frequency ratio f_2/f_1,'' J. Acoust. Soc. Am. **106**, 3473–3483.

Mauermann, M., Uppenkamp, S., van Hengel, P. W., and Kollmeier, B. (**1999b**). ''Evidence for the distortion product frequency place as a source of distortion product otoacoustic emission (DPOAE) fine structure in humans. II. Fine structure for different shapes of cochlear hearing loss,'' J. Acoust. Soc. Am. **106**, 3484–3491.

Moleti, A., and Sisto, R. (**2003**). ''Objective estimates of cochlear tuning by otoacoustic emission analysis,'' J. Acoust. Soc. Am. **113**, 423–429.

Moulin, A., and Kemp, D. T. (**1996**). ''Multicomponent acoustic distortion product otoacoustic emission phase in humans. II. Implications for distortion product otoacoustic emissions generation,'' J. Acoust. Soc. Am. **100**, 1640–1662.

Probst, R., Lonsbury-Martin, B. L., and Martin, G. K. (**1991**). ''A review of otoacoustic emissions,'' J. Acoust. Soc. Am. **89**, 2027–2067.

Siegel, J. H., and Hirohata, E. T. (**1994**). ''Sound calibration and distortion product otoacoustic emissions at high frequencies,'' Hear. Res. **80**, 146–152.

Sisto, R., and Moleti, A. (**2002**). ''On the frequency dependence of the otoacoustic emission latency in hypoacoustic and normal ears,'' J. Acoust. Soc. Am. **111**, 297–308.

Smurzynski, J., and Kim, D. O. (**1992**). ''Distortion-product and click-evoked otoacoustic emissions of normally hearing adults,'' Hear. Res. **58**, 227–240.

Talmadge, C. L., Tubis, A., Long, G. R., and Piskorski, P. (**1998**). ''Modeling otoacoustic emission and hearing threshold fine structures,'' J. Acoust. Soc. Am. **104**, 1517–1543.

Talmadge, C. L., Long, G., Tubis, A., and Dhar, S. (**1999**). ''Experimental confirmation of the two-source interference model for the fine structure of distortion product otoacoustic emissions,'' J. Acoust. Soc. Am. **105**, 275–292.

Zweig, G., and Shera, C. A. (**1995**). ''The origin of periodicity in the spectrum of otoacoustic emissions,'' J. Acoust. Soc. Am. **98**, 2018–2047.

Investigation of potential effects of cellular phones on human auditory function by means of distortion product otoacoustic emissions

Thomas Janssen,[a] Paul Boege, and Jutta von Mikusch-Buchberg
ENT-Department, Laboratory for Experimental Audiology, Technical University Munich, Germany

Johannes Raczek
T-Systems, Technology Centre, Darmstadt, Germany

(Received 16 September 2004; revised 10 December 2004; accepted 10 December 2004)

Outer hair cells (OHC) are thought to act like piezoelectric transducers that amplify low sounds and hence enable the ear's exquisite sensitivity. Distortion product otoacoustic emissions (DPOAE) reflect OHC function. The present study investigated potential effects of electromagnetic fields (EMF) of GSM (Global System for Mobile Communication) cellular phones on OHCs by means of DPOAEs. DPOAE measurements were performed during exposure, i.e., between consecutive GSM signal pulses, and during sham exposure (no EMF) in 28 normally hearing subjects at tone frequencies around 4 kHz. For a reliable DPOAE measurement, a 900-MHz GSM-like signal was used where transmission pause was increased from 4.034 ms (GSM standard) to 24.204 ms. Peak transmitter power was set to 20 W, corresponding to a specific absorption rate (SAR) of 0.1 W/kg. No significant change in the DPOAE level in response to the EMF exposure was found. However, when undesired side effects on DPOAEs were compensated, in some subjects an extremely small EMF-exposure-correlated change in the DPOAE level (<1 dB) was observed. In view of the very large dynamic range of hearing in humans (120 dB), it is suggested that this observation is physiologically irrelevant. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1854331]

PACS numbers: 43.64.Jb [BLM] Pages: 1241–1247

I. INTRODUCTION

To date, there is an increasing public and scientific debate on whether and what biological systems may be influenced by radio-frequency electromagnetic fields (EMF) transmitted by cellular phones. For instance, headache, sensation of warmth around the ear, alteration of the blood–brain barrier, disturbance of sleep, and alteration of cognitive functions have been reported as effects resulting from cellular phone use (Frey *et al.*, 1998; Oftedal *et al.*, 2000; Fritze *et al.*, 1997; Borbely *et al.*, 1999; Preece *et al.*, 1999). A further omnipresent question in general scientific and public discussion is whether EMFs do influence the auditory function. The fluid-filled inner ear possesses microhomeostatic mechanisms that are essential for the functioning of the mechanoelectrical transduction of the auditory hair cells in the cochlea (Wangemann and Schacht, 1996; Kros, 1996). Thus, disturbances of the mechanoelectrical transduction and thermal effects may appear during cellular phone use, especially because there is a close proximity of the cellular phone to the cochlea resulting *a priori* in a fairly high absorption rate of the incident EMF (Burkhardt and Kuster, 2000).

The auditory sense organ in mammals has developed special sensory cells—the outer hair cells (OHC)—to enhance sensitivity and frequency selectivity by amplifying low-level sound signals mechanically (Dallos, 1992). *In vitro*, OHCs are capable of fast contractions and elongations of their cell body in response to an electric field (Brownell *et al.*, 1985). This electromotility is suggested to be a result from a protein in the OHC basolateral membrane that undergoes structural rearrangements in response to changes in the transmembrane voltage, and is assumed to produce the amplification of vibrations in the cochlea during acoustic stimulation (Zheng *et al.*, 2000).

The purpose of the present study was to find out whether EMFs transmitted by cellular phones of the GSM (Global System for Mobile Communication) type may interfere with OHC electromotility. To test this, a noninvasive physiological measure was used which is known to directly reflect OHC function, i.e., distortion product otoacoustic emissions (DPOAEs) (e.g., Mills and Rubel, 1996).

Due to their nonlinear transmission characteristics and corresponding intermodulation distortion, OHCs evoke intermodulation vibrations in cochlear micromechanics and fluid when stimulated by two tones f_1 and f_2 ($f_2 > f_1$) of neighboring frequencies ($f_2/f_1 = 1.2$) which can be measured by a microphone positioned in the outer ear canal (Kemp, 1978; Brownell, 1990). In humans, the $2f_1 - f_2$ DPOAE has the highest level of all occurring intermodulation products and is, therefore, primarily used for diagnosing cochlear dysfunction (Gorga *et al.*, 2000). Even minute changes in the functioning of OHCs—e.g., caused by low-level noise exposure, increased body temperature due to fever, administration of salicylate, or alteration of body posture—are known to considerably affect DPOAE amplitude (Skellet *et al.*, 1996; O'Brien *et al.*, 1994; McFadden and Plattsmier, 1984; Janssen *et al.*, 2000; Frank *et al.*, 2000). Even if DPOAEs do only reflect peripheral sound processing, they are able to

[a]Electronic mail: t.janssen@lrz.tum.de

detect alterations in hearing function at the most sensitive and most vulnerable stage in the auditory pathway, i.e., the cochlea.

The 900-MHz GSM-signal form was chosen as GSM-cellular phones are currently widely used in everyday life. For instance, according to a declaration of the German Federal Office for Radiation Protection, in Germany about $\frac{3}{4}$ of the inhabitants uses GSM cellular phones. In the Scandinavian countries and Italy the percentages are still higher.

When using the common DPOAE measuring technique, DPOAEs cannot be measured during EMF exposure since the electro-acoustic transducers of the DPOAE sound probe used are disturbed when the cellular phone is transmitting the EMF. To overcome this problem DPOAEs were measured within transmission pauses, i.e., between EMF-signal pulses. Thus, an undistorted measurement of DPOAEs was possible during EMF exposure. To our knowledge, there is no study in which otoacoustic emissions have been measured with such a technique for investigating possible effects of cellular phone use on human hearing.

II. METHODS

A. Subjects

Measurements were performed on a group of 28 normally hearing subjects consisting of 14 males and 14 females aged between 16 and 30 years (mean age 26.9 years). All subjects had a negative history of hearing disorders. Their hearing thresholds were smaller than 15 dB HL at octave tone frequencies from 0.125 to 8 kHz and interoctave frequencies from 0.75 to 6 kHz. Middle-ear disorders were excluded by tympanometry.

The experimental protocol was in accordance with the revised Helsinki Declaration (1983) and was approved by the local ethics committee of the medical faculty of the Technical University of Munich.

B. DPOAE measurement

DPOAEs were recorded with an ear-canal sound probe consisting of two loudspeakers and one microphone (ER-10C, Etymotic Research) and a digital signal processor (PCMCIA-DSP-card, DP2000, Starkey). Primary tones for eliciting DPOAEs were separately calibrated in the subject's ear canal using a broadband signal. Calibration was performed before each measurement in order to minimize undesired side effects on DPOAEs due to an altered position of the sound probe in the outer ear canal or changes in atmospheric pressure within the tympanic cavity which can result in altered primary tone and hence DPOAE signal levels. The DPOAE level L_{dp} at the frequency $2f_1-f_2$ and the noise floor L_{noise} (being the average of three spectral lines below and above $2f_1-f_2$) were obtained from the Fourier transform of the microphone signal.

Despite the shielding of the sound probe and cables by means of sheet copper and ferrite rings, distortion occurred in the electro-acoustic transducers. The distortion's die-out time amounted to 6 ms. For a reliable DPOAE measurement between consecutive GSM pulses, transmission pauses had to be increased from 4.034 ms (GSM standard) to 24.204 ms.

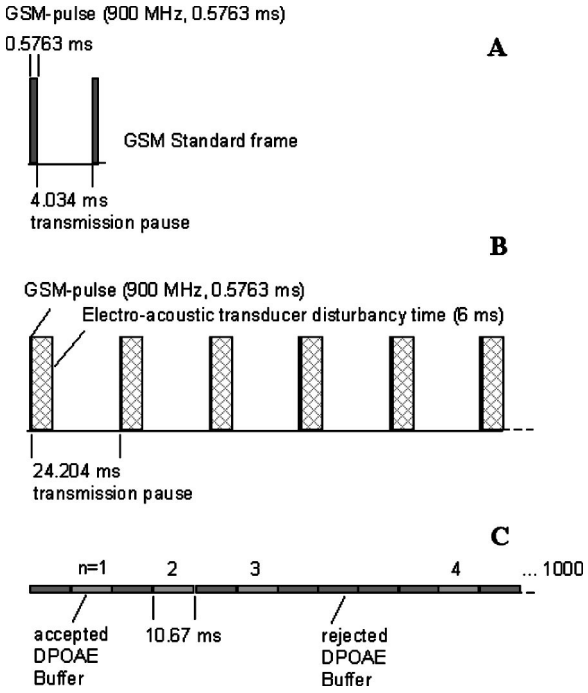


FIG. 1. Time lapse of GSM standard frame with a pulse width of 0.5763 ms, transmission pause of 4.61 ms, and carrier frequency of 900 MHz (A), altered GSM frame of increased transmission pause (24.78 ms) used in the present study comprising periods (6 ms) of electro-acoustic transducer disturbance after an EMF pulse (B), and scheme for sampling undistorted DPOAE buffers (C). EMF pulses and DPOAE buffers were recorded simultaneously. During off-line analysis only those buffers were accepted and averaged in which no EMF pulse appeared and no disturbance of electro-acoustic transducers was present.

For off-line analysis only those buffers were accepted and averaged ($n=1000$) in which no EMF pulse appeared. The time lapse of the modified GSM frame, the electro-acoustic transducer disturbance time and the rejected/accepted DPOAE buffers are shown in Fig. 1.

DPOAEs were measured at three primary tone frequencies, i.e., $f_2=4125$ Hz, $f_2=3937.5$ Hz, and $f_2=3750$ Hz, corresponding to the most sensitive frequency range in human hearing. Another reason for choosing frequencies around 4 kHz was the fact that the noise floor which was produced by the breathing of the subject and by the microphone was minimum at these frequencies. Due to the fact that DPOAEs are more sensitive to cochlear impairment when elicited at low primary tone levels L_1 and L_2, medium to close-to-threshold levels were applied, i.e., $L_2=60$, $L_2=50$, $L_2=40$, and $L_2=30$ dB SPL ($L_1=0.4L_2+39$; $f_1=f_2/1.2$) (Janssen *et al.*, 1998, Kummer *et al.*, 1998). Custom-made software (using MATLAB 5.3) was developed for DPOAE measurements and data analysis.

For each subject, each DPOAE measurement was repeated 8 times, applying EMF and sham exposure in alternating order. In 14 of 28 subjects tests started with EMF exposure, in the other 14 subjects with sham exposure. DPOAE measurements began with the highest primary tone frequency ($f_2=4125$ Hz), varying the primary tone level from high to low ($L_2=60$, 50, 40, 30 dB SPL). The same procedure was carried out for $f_2=3937.5$ and $f_2=3750$ Hz. Thus, in total, each individual subject underwent

12 tests. The duration of each test amounted to about 24 min. Testing was broken up into multiple sessions.

C. EMF exposure

Subjects were exposed to the GSM-like signal with a carrier frequency of 900 MHz, a pulse duration of 0.5763 ms, and a pause of 24.204 ms between two consecutive pulses. The peak pulse transmitter power was set to 20 W, i.e., the average transmitter power amounted to 0.465 W, resulting in a specific absorption of SA = 2.42 mWs/kg, which is comparable to that when using a GSM mobile phone under typical conditions. The corresponding specific absorption rate (SAR) was 0.1 W/kg. A custom-made amplifier was used to achieve the required power level of the EMF. The EMF signals were emitted by a monopole antenna which was mounted onto a mobile phone dummy. The distance between the antenna and the subject's ear was kept at 5 cm. The measurements were carried out in a sound-proof, electromagnetically shielded room ($6.5\times2.80\times3.0$ m). Its walls, floor, and ceiling were covered with radio-frequency-absorbing materials to suppress unwanted signal reflections. With the exception of the subject, the mobile phone dummy, and a plastic camp bed, no other things were in the room. The subjects were instructed to lie quietly on the camp bed during the measurements.

The measuring procedure was the same for EMF and sham exposure. The only difference was that EMF was switched off during sham exposure. Thus, for the subjects there was no noticeable difference because acoustic stimulation was the same in both conditions.

D. Criteria for DPOAE stability

DPOAEs from 11 normally hearing ears obtained in subjects who did not participate in this study were used in order to establish criteria for evaluating the stability of the DPOAE measurements. For that purpose, repetitive DPOAE measurements ($n=10$) were carried out for each subject with unchanged sound probe position in the outer ear canal. DPOAE measurements were performed at four different primary tone levels ($L_2=60$, 40, 30, 20 dB SPL, $L_1=0.4L_2+39$) and at six different tone frequencies around 4 kHz. With increasing signal-to-noise ratio ($\mathrm{SNR}=L_{dp}-L_{noise}$), the corresponding standard deviation of the DPOAE level σ_{Ldp} decreased. For example, at a SNR of 10 dB the σ_{Ldp} amounted to 1.8 dB, at 20 dB to 0.7 dB. At a very high SNR of 40 dB a very low standard deviation of 0.1 dB was present (Fig. 2). That means: the higher the SNR, the higher the reliability of the DPOAE measurement concerned. This finding is important with respect to the evaluation of low DPOAE changes and proved DPOAEs to be an accurate, reliable, and extremely sensitive measure for detecting minute disturbances of cochlear sound processing.

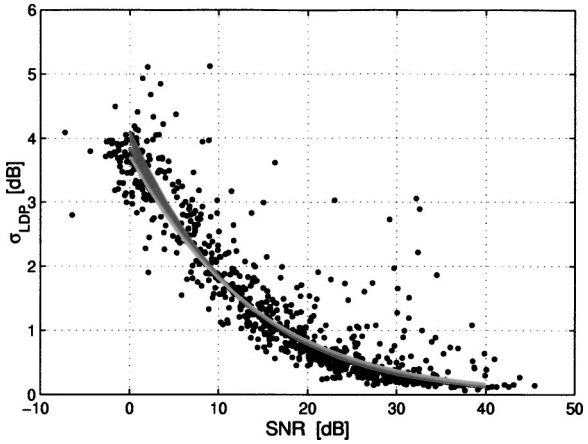


FIG. 2. Standard deviation of the DPOAE level σ_{Ldp} for repetitive measurements ($n=10$, unchanged sound probe) in 11 normally hearing ears across signal-to-noise ratio $\mathrm{SNR}=L_{dp}-L_{noise}$. DPOAE measurements were performed at different primary tone frequencies and levels.

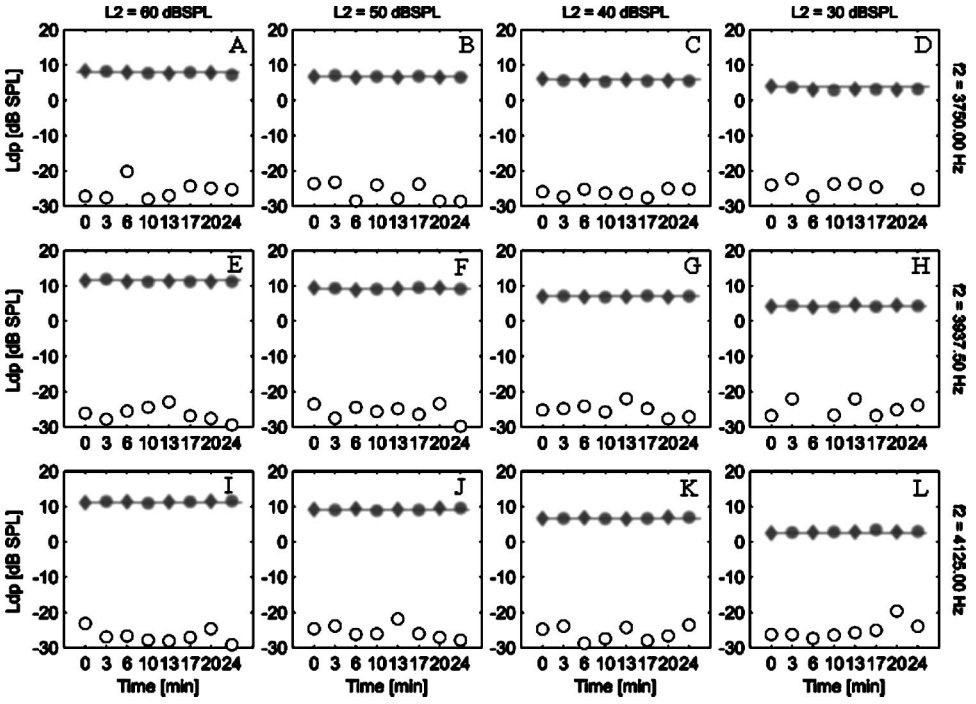


FIG. 3. DPOAE level L_{dp} measured for EMF-exposure, i.e., within transmission pauses (rhombuses), and for sham exposure (filled circles), for one individual subject. Open circles indicate the corresponding noise-floor level L_{noise}. There were 12 different tests where frequency f_2 and level L_2 of the primary tones were varied. f_2 and L_2 are indicated at the respective panels (A)–(L). Chosen case is representative of the 24 subjects in whom the EMF-correlated change in DPOAE level was smaller than 0.25 dB.

E. Data analysis

Since stability, and thus reliability, of the DPOAE level L_{dp} depends on the SNR, the DPOAE level was weighted prior to the statistical analysis, whereby the reciprocal of the corresponding standard deviation σ_{Ldp} of the DPOAE level L_{dp} served as weighting factor (see Sec. II D). The statistical evaluation of the data was performed using the average value of the weighted DPOAE levels of each test for the two exposure conditions (EMF and sham exposure, respectively). Criteria for accepting one test to be valid were (i) only those DPOAE levels with an SNR higher than 10 dB SPL were accepted; (ii) for computing the average value, at least three of the four measurements had to be valid. The two-sided t-test was applied on the difference of the average DPOAE levels for EMF and sham exposure for the different tests (three different tone frequencies and four primary tone levels). The significance level was set to 5%. Investigations were performed in the form of a single-blind study: the examiner knew whether the EMF was on or off; the subjects, however, were not instructed.

III. RESULTS

DPOAEs of the 12 tests (A–L) are presented in Fig. 3 for an individual subject. Rhombuses indicate DPOAEs when measured during EMF exposure; circles indicate DPOAEs when measured during sham exposure. For a better visualization of the variation of the DPOAE level, a horizontal line which starts from the first DPOAE is inserted in each panel. No systematic change in DPOAE level was visible when comparing neighboring DPOAEs representing the alternating conditions (EMF and sham exposure). However, independent of EMF or sham exposure, there was a small change in DPOAE level over time (compare the first and last DPOAE measurements in panels in Fig. 3) which was, on average, lower than 1 dB for the 12 tests. This variation in DPOAE level over time is suggested to be attributed to undesired side effects due to altered position of the sound probe in the outer ear canal and to changes in the middle-ear transfer function caused by minute changes in the atmospheric pressure in the tympanic cavity that can occur during swallowing.

The difference of the averages of the weighted DPOAE levels for EMF and sham exposure was computed for all 12 tests for each of the 28 individual subjects. The average difference for the 28 subjects amounted to -0.0049 dB, and the standard deviation to 1.0353 dB. The highest negative difference was -2.5614 dB, the highest positive difference $+2.0767$ dB. Note that a positive difference means that the average DPOAE level for EMF exposure was lower than the corresponding average DPOAE level for sham exposure. Differences, when classified using 2.5-dB scaling and plotted in the form of a histogram, revealed a Gaussian-shaped distribution (Fig. 4). For evaluating a possible influence of tone level and tone frequency, data were analyzed separately for each of the four tone levels, and for each of the three tone frequencies. Distribution of differences for these tests were also Gaussian-shaped (not shown). The average differences were around 0 dB SPL for all tests (Table I). The standard

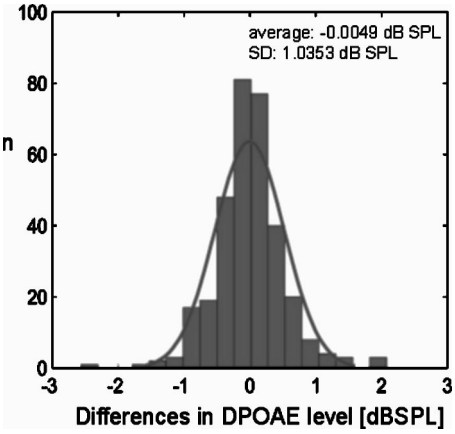


FIG. 4. Distribution of the difference of average DPOAE levels (EMF exposure versus sham exposure) obtained in the 12 tests (see Fig. 2) for all 28 subjects. Moreover, the Gaussian distribution calculated using the measured average and standard deviation is shown.

deviation varied from 0.8377 to 1.4171 dB. The higher standard deviation observed at low primary tone levels can be attributed to the fact that the SNR, and consequently the reliability of the measurement, decreased with decreasing primary tone level.

In order to test whether there is a significant EMF-exposure-induced DPOAE change or not, a two-sided t-test was performed for the 28 subjects for each of the 12 tests. The t-test did not show any significant difference in DPOAE level obtained for EMF and sham exposure in the 12 tests (i.e., $p<0.05$). However, when data were separated by gender, the 14 female subjects exhibited a significant difference of the DPOAE level for EMF and sham exposure for one of the 12 tests ($L_2=60$ dB SPL, $f_2=3750$ Hz). For the 14 male subjects the t-test showed no significant difference for all 12 tests.

To compensate for undesired side effects on DPOAE level and hence to improve the detection of small DPOAE level changes possibly induced by the EMF, DPOAE variation over time was eliminated by fitting DPOAE data of each test by a third-order polynomial. When plotting the deviation of DPOAE level from the polynomial, neighboring DPOAEs exhibited a clear up and down, revealing a systematic EMF-exposure-correlated change in DPOAE level (Fig. 5). However, these changes were very small and amounted to only some tenths of 1 dB. Please note the high y-axis resolution. It should be emphasized that the DPOAE level was higher after EMF exposure compared to that after sham exposure. The first DPOAE measurement exhibited the highest difference between sham exposure and EMF exposure. The maximum difference was 0.55 dB (at $L_2=30$ dB SPL and $f_2=4125$ Hz) [see Fig. 5 (L)].

SNR varied between 33 and 24 dB. That means that, especially at high primary tone levels, a high measuring accuracy was achieved. At an SNR of 33 dB, the corresponding standard deviation σ_{Ldp} of the DPOAE level amounted to 0.25 dB (see the criteria for DPOAE stability in Fig. 2). That means the observed changes in DPOAE level at high tone levels, being higher than 0.25 dB, are presumably a result of the EMF exposure and cannot be attributed to an insufficient

TABLE I. Mean and standard deviation (SD) of the difference of the average DPOAE levels of the two conditions (EMF exposure and sham exposure). Results were obtained by combining the measurement results for all primary tone frequencies f_2 at each of the primary tone levels L_2 separately (see columns 2–5) and by combining the results for all L_2 at each of the three f_2 separately (see columns 6–8), respectively (28 subjects).

	$L_2=60$ dB SPL all f_2	$L_2=50$ dB SPL all f_2	$L_2=40$ dB SPL all f_2	$L_2=30$ dB SPL all f_2	$f_2=4125$ Hz all L_2	$f_2=3937.5$ Hz all L_2	$f_2=3735$ Hz all L_2
Mean [dB]	−0.0130	−0.0401	−0.0169	0.0554	−0.0086	0.0718	−0.0813
SD [dB]	0.9286	0.8377	0.8988	1.4171	0.9118	1.1572	1.0075

measurement accuracy. Also, at the lowest primary tone level ($L_2=30$ dB) with a SNR of 24 dB, the standard deviation of the DPOAE level is relatively small and amounts to 0.6 dB.

It should be emphasized that the tests on the other 27 subjects showed a considerably smaller or no EMF-exposure-correlated change in DPOAEs. Four subjects exhibited an EMF-exposure-correlated change in DPOAE level which was higher than 0.25 dB at least in one of the 12 tests. In these subjects, an EMF effect cannot definitely be excluded. In 24 of the 28 subjects the change in DPOAE level was smaller than 0.25 dB. This change in DPOAE level cannot be attributed to the EMF because the change was within the variability of the measurement accuracy. Due to the small number of subjects who showed an EMF-correlated change in DPOAE level and the different natures of the responses, no statistical analysis of the data was performed.

For identifying possible technical artifacts, tests were performed in an artificial ear (B&K 4157) using the same sound probe, the same measuring procedure, and the same EMF exposure and sham exposure, respectively. The DPOAE signal was simulated by delivering a sinusoidal signal with frequency $2f_1-f_2$ via the sound probe's loudspeaker. The $2f_1-f_2$ sinusoidal signal was adjusted such that the sound-pressure level in the artifical ear was 5 dB SPL, which is a typical DPOAE level that can be found in a normally hearing human subject. Changes in the level of the simulated DPOAE in the dummy ear when exposed to EMF were very small (typically 0.05 dB), compared to that observed in the human ear (maximum 0.55 dB; see Fig. 5). Since the change in level of the simulated DPOAE in the dummy ear was considerably smaller than the change in DPOAE level observed in the human ears, technical artifacts can be excluded.

IV. DISCUSSION

The hearing organ and particularly the auditory sensory cells are known to be highly sensitive to exogenous and endogenous agents (Frey *et al.*, 1998; Oftedal *et al.*, 2000; Fritze *et al.*, 1997; Borbely *et al.*, 1999; Preece *et al.*, 1999) and are therefore a suited object for investigating potential impact of cellular phones on biological systems. Since hearing capability is not subjectively altered during cellular phone use, only minute changes in hearing function are expected, if any.

For investigating minute changes in auditory function, high measurement accuracy is needed. Since measurement accuracy depends directly on signal-to-noise-ratio (SNR), detection of EMF-exposure-induced changes in DPOAE level requires high SNRs (Fig. 2). Most subjects exhibited SNRs up to 40 dB that corresponded to a DPOAE level

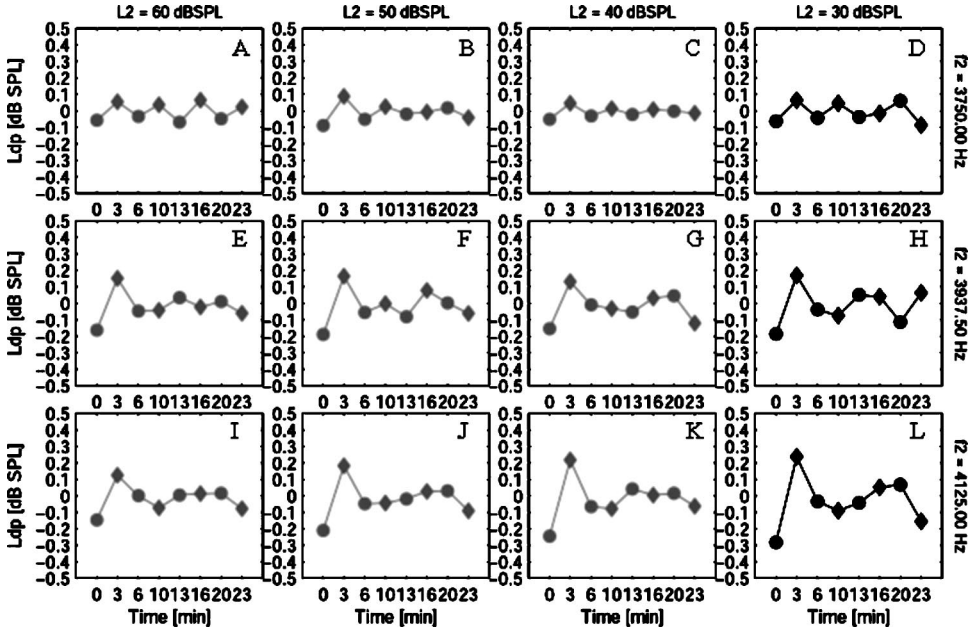


FIG. 5. DPOAE level L_{dp} measured for EMF exposure, i.e., within transmission pauses (rhombuses), and for sham exposure (filled circles), for one individual subject, plotted relatively to the fitting polynomial. Primary tone level L_2 und primary tone frequency f_2 used in the 12 tests are indicated at the respective panels. Note the high y-axis resolution.

stability of about 0.1 dB. Thus, DPOAEs provide a highly sensitive method for assessing minute changes in the functioning of auditory sensory cells.

The t-test did not reveal any significant EMF-exposure-induced change in the DPOAE level, not even at low primary tone levels, at which the DPOAEs are reported to be most sensitive to cochlear impairment (Janssen *et al.*, 1998; Kummer *et al.*, 1998), and at primary tone frequencies at 4 kHz, where the human hearing organ is most sensitive and thus most vulnerable. When separating females from males, there were also no significant changes in the DPOAE level, with one exception for females at the highest primary tone level ($L_2 = 60$ dB SPL) and the lowest primary tone frequency ($f_2 = 3750$ Hz) used. It should be noted that one result within the significance limit (5%) for male and female ($2 \times 12 = 24$ tests in total) can occur by pure randomness, even if no effect exists in reality. It is noteworthy in this context that the one significance occurred at the most intense tone level ($L_2 = 60$ dB SPL) so that a random effect is more likely.

The stability of the DPOAE level in the individual ear and hence the standard deviation of the DPOAE level in the subject sample is believed to be influenced also by the variation of the DPOAE level over time due to altered sound probe position (which lead to changes in stimulus level) and atmospheric pressure difference between outer ear canal and tympanic cavity (which alters middle-ear sound conduction). The observation that width of the distribution of the DPOAE differences is largest at the lowest primary tone level (1.4171 dB; see standard deviation in Table I) can be attributed to the fact that with decreasing primary tone level the SNR is decreasing and thus the measurement reliability is smaller.

To compensate for undesired side effects on the DPOAE level caused by altered sound probe position and changes in middle-ear sound conduction, the DPOAE levels of a measuring sequence were fitted by a polynomial. By referring the measured data to the polynomial, the variation of the DPOAE level due to non-EMF exposure effects could be eliminated to a great extent. In doing this, a clear EMF-exposure-correlated change in the DPOAE level became evident (see Fig. 5). However, the change in the DPOAE level was extremely small (on the order of some tenths of one dB). Measurements of such high accuracy could be performed only thanks to an extremely high SNR. The fact that changes in the DPOAE level were not found in the dummy ear suggested the observed DPOAE variation to be actually caused by the EMF exposure. It should be emphasized that this finding was present in a few subjects only and did not exhibit a unique characteristic form. A statistical analysis of the data was therefore not expedient. In view of the very large dynamic range of hearing in humans, which amounted to more than 120 dB, changes in DPOAE level of some tenths of 1 dB are suggested to be irrelevant with respect to a potential health risk.

Surprisingly, the DPOAE level increased during EMF exposure (see Fig. 5). OHC dysfunction is known to reduce the DPOAE level. In fact, patients suffering from cochlear hearing loss show decreasing DPOAE levels with increasing hearing loss (Kummer *et al.*, 1998). However, patients whose hearing loss is associated with tinnitus are reported to exhibit abnormal high OAE levels despite the hearing loss (Janssen *et al.*, 2000; 1998; Attias *et al.*, 1996). In these ears, a peculiar impairment of OHCs appears to generate reinforced distortion within cochlear micromechanics that gives rise to the tinnitus and leads to higher DPOAE levels than expected. However, compared to tinnitus patients who show DPOAE level elevation on the order of several dBs, the DPOAE level change during EMF exposure is considerably smaller.

The electromotility of OHCs is suggested to result from a membrane-based force generator mechanism associated with conformational changes and rearrangement of a voltage-sensitive integral membrane protein (Zheng *et al.*, 2000). It is believed to generate mechanical force upon stimulation up to stimulus frequencies of at least 24 kHz (Dallos *et al.*, 1995). This ''piezoelectrical'' function of OHCs might be influenced by external EMFs. Otoacoustic emissions are a by-product of OHC motility and thus might be able to reflect these functional alterations.

If the observed change in the DPOAE level was actually caused by the EMF, then one can speculate that EMF pulses may have changed the piezoelectric transduction of OHCs. The higher DPOAE level during EMF exposure would then suggest an induced electrical stimulation of OHCs resulting in an increase of the membrane-based force generator mechanism and hence in an increase of the DPOAE amplitude. However, whether the observed EMF-exposure-correlated change in DPOAE level is really due to the EMF or to unknown side effects other than reported here cannot be definitely answered and has to be further investigated, preferably in animal studies where higher EMF intensities can be applied. In those studies thermal effects also could be investigated since measurements of the cochlear fluids' temperature are possible.

In the literature (PubMed database search), only one single report on potential effects on OAEs after cellular phone use can be found (Ozturan *et al.*, 2002). However, the authors investigated long-term effects since they compared OAEs measured before and after EMF exposure (10 min). This is in contrast to our study where DPOAEs were measured during EMF exposure. Ozturan *et al.* did not find a statistically significant difference in their measures and thus excluded long-term thermal effects which may influence auditory sensory transduction. Beside the above-cited OAE paper, only three other papers investigated potential impact of cellular phones on neural function (Bak *et al.*, 2003; Kizilay *et al.*, 2003; Arai *et al.*, 2003). However, in these studies auditory brainstem responses served as a probe for assessing potential effects of EMF on neural auditory sound processing. The authors did not find any significant change when comparing their measures before and after EMF exposure.

For the further investigation of possible effects of cellular phones on auditory function by means of OAEs, additional studies are necessary. Since the time constant of sensory transduction is very small (tenths of milliseconds) it is suggested that the time constant of the recovery of distorted OHCs is also small. Thus, if radio-frequency EMFs really affect the electromotility of OHCs, possible effects can be identified best when measuring the DPOAEs directly after

the EMF pulses. Recall that in the present study DPOAEs could be measured only 6 ms after the EMF pulse (see Fig. 1). Therefore, it is necessary to improve the electromagnetic shielding of the electro-acoustic transducers and cable of the sound probe, allowing DPOAEs to be measured without disturbance within tenths of milliseconds after the end of the EMF pulses. If sufficient electromagnetic shielding would be available or, alternatively, when using optical microphones and delivering primary tones via tubes driven by external sound sources, undistorted DPOAE measurements may be possible even directly after or even during EMF pulses. With that technique, the original GSM frame length could be applied. Using such a technique, the investigation of possible effects of the nonpulsed EMF of the third generation mobile telecommunication systems (UMTS) on auditory function would also be possible.

Furthermore, in order to verify the observed EMF-exposure-correlated variation in the DPOAE level, subjects have to be exposed to an EMF of higher intensity than that used in the present study. To enable dose-response evaluation, subjects have to be exposed to different power flux densities close to the safety limits. Moreover, investigations should be performed in animals, in which power flux densities above the safety limit for humans can be applied. Since there is evidence that auditory sensory cells work like piezoelectric transducers, disturbances of sensory transduction are suggested to occur. The question is, however, at which power-flux density is that effect verifiable. From the point of view of searching for potential health risks of humans by using cellular phones, further investigations are necessary on this topic.

ACKNOWLEDGMENT

The study was funded by T-Systems, Germany, who provided the measuring equipment and financed the experiments.

Arai, N., Enomoto, H., Okabe, S., Yuasa, K., Kamimura, Y., and Ugawa, Y. (**2003**). ''Thirty minutes mobile phone use has no short-term adverse effects on central auditory pathway,'' Clini. Neurophys. **114**, 1390–1394.

Attias, J., Bresloff, I., and Furman, V. (**1996**). ''The influence of the efferent auditory system on otoacoustic emissions in noise induced tinnitus: Clinical relevance,'' Acta Oto-Laryngol. **116**, 534–539.

Bak, M., Sliwinska-Kowalska, M., Zymslony, M., and Dudarewicz, A. (**2003**). ''No effect of acute exposure to the electromagnetic field emitted by mobile phones on brainstem auditory potentials in young volunteers,'' Int. J. Occup. Environ. Health **16**, 201–209.

Borbely, A.A., Huber, R., Graf, T., Fuchs, B., Gallmann, E., and Achermann, P. (**1999**). ''Pulsed high-frequency electromagnetic field affects human sleep and sleep electroencephalogram,'' Neurosci. Lett. **275**, 207–210.

Brownell, W.E. (**1990**). ''Outer hair cell electromotility and otoacoustic emissions,'' Ear Hear. **11**, 82–92.

Brownell, W.E., Bader, C.R., Bertrand, D., and de Ribaupierre, Y. (**1985**). ''Evoked mechanical responses in isolated cochlear outer hair cells,'' Science **227**, 194–196.

Burkhardt, M., and Kuster, N. (**2000**). ''Appropriate modeling of the ear for compliance testing handheld MTE with SAR safety limits at 900/1800 MHz,'' IEEE Trans. Microwave Theory Tech. **48**, 1927–1934.

Dallos, P. (**1992**). ''The active cochlea,'' J. Neurosci. **12**, 4575–4585.

Dallos, P., and Evans, B.N. (**1995**). ''High frequency motility of outer hair cells and the cochlear amplifier,'' Science **267**, 2006–2009.

Frank, A.M., Alexiou, C., Hulin, P., Janssen, T., Arnold, W., and Trappe, A.E. (**2000**). ''Non-invasive measurement of intracranial pressure changes by otoacoustic emissions (OAEs)—a report of preliminary data,'' Zentralbl Neurochir. **61**, 177–180.

Frey, A.H. (**1998**). ''Headaches from cellular telephones: Are they real and what are the implications?,'' Environ. Health Perspect. **106**, 101–103.

Fritze, K., Sommer, B., and Schmitz, B. (**1997**). ''Effects of global system for mobile communication (GSM) microwave exposure on blood-brain barrier permeability in rat,'' Acta Neuropathol. (Berl) **94**, 465–470.

Gorga, M.P., Nelson, K., Davis, T., and Dorn, P.A. (**2000**). ''Distortion product otoacoustic emission test performance when both $2f_1-f_2$ and $2f_2-f_1$ are used to predict auditory status,'' J. Acoust. Soc. Am. **107**, 2128–2135.

Janssen, T., Boege, P., Oestreicher, E., and Arnold, W. (**2000**). ''Tinnitus and $2f_1-f_2$ distortion product otoacoustic emissions following salicylate overdose,'' J. Acoust. Soc. Am. **107**, 1790–1792.

Janssen, T., Kummer, P., and Arnold, W. (**1998**). ''Growth behavior of the $2f_1-f_2$ distortion product otoacoustic emission in tinnitus,'' J. Acoust. Soc. Am. **103**, 3418–3430.

Kemp, D.T. (**1978**). ''Stimulated acoustic emissions from within the human auditory system,'' J. Acoust. Soc. Am. **64**, 1386–1391.

Kizilay, A., Ozturan, O., Erdem, T., Kalcioglu, M.T., and Miman, M.C. (**2003**). ''Effects of chronic exposure of electromagnetic fields from mobile phones on hearing in rats,'' Auris Nasus Larynx **20**, 239–245.

Kros, C.J. (**1996**). ''Physiology of mammalian cochlear hair cells,'' in *The Cochlea*, edited by P. Dallos, A.N. Popper, and Fay R.R. (Springer, New York), Chap. 9, pp. 318–385.

Kummer, P., Janssen, T., and Arnold, W. (**1998**). ''The level and growth behavior of the $2f_1-f_2$ distortion product otoacoustic emission and its relationship to auditory sensitivity in normal hearing and cochlear hearing loss,'' J. Acoust. Soc. Am. **103**, 3431–3444.

McFadden, D., and Plattsmier, H.S. (**1984**). ''Aspirin abolishes spontaneous otoacoustic emissions,'' J. Acoust. Soc. Am. **76**, 443–448.

Mills, D.M., and Rubel, E.D. (**1996**). ''Development of the cochlear amplifier,'' J. Acoust. Soc. Am. **100**, 428–411.

O'Brien, A. (**1994**). ''Temperature dependency of the frequency and level of a spontaneous otoacoustic emission during fever,'' Audiology **28**, 281–290.

Oftedal, G., Wilen, J., Sandstorm, M., and Mild, K.H. (**2000**). ''Symptoms experienced in connection with mobile phone use,'' Rev. Asoc Prof. Hosp. Nac. Odontol. **50**, 237–245.

Ozturan, O., Erdem, T., Miman, M.C., Kalcioglu, M.T., and Oncel, S. (**2002**). ''Effects of electromagnetic field of mobile telephones on hearing,'' Acta Oto-Laryngol. **122**, 289–293.

Preece, A.W., Iwi, G., and Davies-Smith, A. (**1999**). ''Effect of a 915 MHz simulated mobile phone signal on cognitive function in man,'' Int. J. Radiat. Biol. **75**, 447–56.

Skellett, R.A., Crist, J.R., Fallon, M., and Bobbin, R.P. (**1996**). ''Chronic low-level noise exposure alters distortion product otoacoustic emissions,'' Hear. Res. **98**, 68–76.

Wangemann, P., and Schacht, J. (**1996**). ''Homeostatic mechanisms in the cochlea,'' in *The Cochlea*, edited by P. Dallos, A.N. Popper, and Fay R.R. (Springer, New York), Chap. 3, pp. 130–385.

Zheng, J., Shen, W., He, D.Z., Long, K.B., Madison, L.D., and Dallos, P. (**2000**). ''Prestin is the motor protein of cochlear outer hair cells,'' Nature (London) **405**, 149–155.

Distortion-product otoacoustic emission measured with continuously varying stimulus level

Stephen T. Neely, Tiffany A. Johnson, and Michael P. Gorga
Boys Town National Research Hospital, Omaha, Nebraska 68131

(Received 2 August 2004; revised 6 December 2004; accepted 6 December 2004)

Distortion-product otoacoustic emissions (DPOAE) are measured by stimulating the ear with two simultaneous tones. A novel method for measuring DPOAEs has been developed in which the tone levels vary continuously instead of in discrete steps. Varying the tone levels continuously may offer advantages for characterizing DPOAE level as a function of stimulus level. For equivalent primary levels, DPOAE levels measured with the continuous-level method were the same as levels obtained with the discrete-level method, thus validating the new method. Continuous-level measurements were used to determine the optimal L_1 for each L_2 in individual subjects ($N=20$) at $f_2=1$, 2, 4, and 8 kHz by using a Lissajous path that covered a wide range of stimulus levels. The optimal L_1 (defined as the L_1 that resulted in the largest DPOAE for each L_2) varied across subjects and across frequency. The optimal difference between L_1 and L_2 decreased with increasing L_2 at all frequencies, and increased with frequency when L_2 was low. When the optimal L_1 was determined individually for each ear, the DPOAE levels were larger and less variable than those obtained using the equation for L_1 suggested by Kummer *et al.* [J. Acoust. Soc. Am. **103**, 3431–3444 (1998)].

[DOI: 10.1121/1.1853253]

PACS numbers: 43.64.Jb, 43.64.Yp [BLM] Pages: 1248–1259

I. INTRODUCTION

Distortion-product otoacoustic emission (DPOAE) is sound generated within the cochlea due to nonlinear interaction between two stimulus tones of slightly different frequency. DPOAEs can be recorded in the ear canal, and have been used as an indicator of hearing status (e.g., Gorga *et al.*, 1997) because the nonlinearity that generates DPOAEs is characteristic of normal cochlear function.

DPOAE level varies with the level and frequency of each of the two tones in the stimulus. Measurement of DPOAE level over the complete four-dimensional stimulus space is seldom attempted because it requires so much time. Mills (2002) described DPOAE measurements in gerbil over the entire four-dimensional space, including more than 200 possible stimulus level combinations at each of 42 frequency combinations. Such comprehensive coverage of stimulus space is more difficult in human subjects because DPOAE levels tend to be lower than they are in experimental animals and, thus, longer measurement time is required in order to achieve similar signal-to-noise ratios. Typically, human DPOAE studies cover only a small region of the four-dimensional stimulus space and focus on a restricted region of the two-dimensional frequency or level space (e.g., Harris *et al.*, 1989; Brown and Gaskill, 1990; Gaskill and Brown, 1990; Hauser and Probst, 1991; Whitehead *et al.*, 1995a, b).

In the present study, we focus on representing the two-dimensional L_1,L_2 space at a few frequencies. Whitehead *et al.* (1995b) describe DPOAE measurements from several human subjects over L_1,L_2 space for fixed primary frequencies (near 3 kHz). They suggested that there would be clinical advantage in selecting an optimum L_1 for each L_2, where optimum is defined as the L_1 that results in the largest DPOAE for each L_2. Kummer *et al.* (1998), based on data previously reported by Gaskill and Brown (1990), suggested that the equation $L_1=39+0.4\cdot L_2$ produces, on average, the largest DPOAE level at each L_2. Kummer *et al.* (2000), based on their own measurements at seven frequencies in 22 subjects, found essentially the same equation for L_1 and suggested that it was independent of stimulus frequency.

Traditional DPOAE measurement methods vary the primary-tone levels in discrete steps. When there is interest in measuring DPOAE levels for many closely spaced stimulus levels, it may be more efficient to vary the primary-tone levels continuously. In this paper, we describe the use of continuously varying primary-tone levels in two ways. First, we describe an approach that allowed us to obtain DPOAE levels over a two-dimensional region of L_1,L_2 space by following a Lissajous path, where L_1 and L_2 are the primary-tone levels. Second, we use these measurements to determine an optimal, linear path through the L_1,L_2 space and then obtain DPOAE levels along this linear path by setting $L_1=a+b\cdot L_2$ in a subsequent measurement. Lissajous-path parameters were selected to provide good coverage of L_1,L_2 space in a short time. These parameters are described in detail in the Methods section. Having the ability to rapidly explore a region of L_1,L_2 space in individual ears allowed us to determine the optimal L_1 for each ear. Using continuous-level primaries, we compare the DPOAE levels along this individualized optimal path to levels obtained along alternate linear paths.

There were several objectives in the present study: (1) To determine whether DPOAE levels can be measured reliably with continuous-level stimuli. (2) To determine whether the DPOAE levels produced with this procedure are the same as those produced by traditional measurement paradigms in which primary levels are varied in discrete steps. (3) To compare the optimal-level path derived from Lissajous-path measurements with the paths suggested in other publications. (4)

To examine the frequency dependence of the optimal linear path. (5) To determine whether the DPOAE input/output (I/O) functions based on the optimal linear path are less variable across subjects than those based on the path suggested by Kummer *et al.* (1998). The first two objectives relate to the efficacy and validity of the continuous-level procedure.

II. METHODS

Our measurement system includes a high-quality (24-bit) soundcard (CardDeluxe, Digital Audio Labs) and a DPOAE probe-microphone system (ER-10C, Etymotic Research). We used locally developed software for DPOAE measurements. EMAV (Neely and Liu, 1993) was used when stimuli were presented at discrete levels, while SYSRES (Neely and Stevenson, 2002) was used to present stimuli with continuously varying level.

Twenty normal-hearing young adults served as subjects. All subjects had audiometric thresholds $\leqslant 10$ dB HL (ANSI, 1996) and normal 226-Hz tympanograms (ASHA, 1990). Subjects were tested in a sound-treated room while sitting in a comfortable recliner. They were asked to remain quiet throughout the test session. Each test session lasted 2 h and three test sessions typically were needed to collect all of the data from each subject.

A. Heterodyne analysis

To measure DPOAE levels when stimulus levels are continuously varying, we used a frequency-domain, heterodyne technique to extract the time-varying level of specific frequency components from the measured response. This technique was described by Kim *et al.* (2001) for tracking the time course of DPOAE levels. In the present study, the size of stimulus and response buffers was 2^{21} (or about 2 million) samples. Because our sampling rate was 32 000 samples/second, the duration of the stimulus was about 66 s. To perform the heterodyne analysis, we took discrete Fourier transforms of the entire (2^{21} sample) response buffer. The stimulus waveform was always specified such that the pattern in the first half of the stimulus buffer was repeated in the second half. Our calculation of the $2f_1-f_2$ DPOAE level (L_d) was based on the sum of the first and second halves of the response buffer. Our estimate of the noise level (L_n) at the same frequency was based on the difference between the first and second halves. Although the software used for continuous-level measurements (SYSRES) does not incorporate artifact rejection, the data described below indicate that DPOAEs could still be measured by this method over a wide range of levels.

B. Lissajous-path stimuli

Lissajous figures are constructed by following a path such that the x and y coordinates vary sinusoidally with distance along the path. If the rate of sinusoidal variation is equal for the x and y coordinates, then the Lissajous figure will appear to be a diagonal line, circle, or ellipse, depending on the relative phase of the sinusoidal variation. If the relative rate of sinusoidal variation is a ratio of two relatively prime integers, then the Lissajous figure becomes more complex, with the pattern coursing through more places in the two-dimensional space.

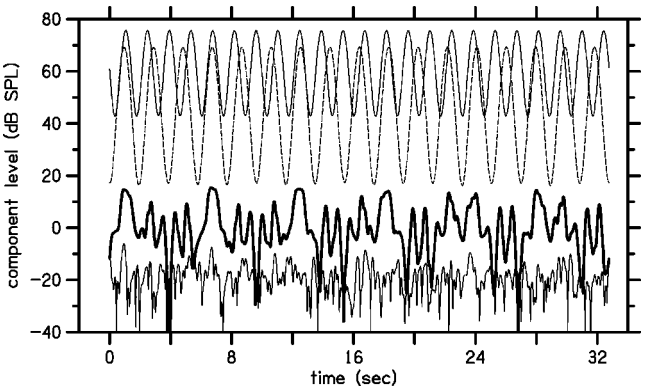


FIG. 1. Levels of measured components as a function of time in response to a Lissajous-path stimulus. The two thin lines near the top of this figure show the levels of the two tones for a Lissajous-path stimulus. The measured L_1 (solid line) varies from 43 to 76 dB SPL with 23 modulation cycles, while L_2 (dashed line) varies from 17 to 69 dB SPL with 17 modulation cycles. The heavy line near 0 dB SPL shows the $2f_1-f_2$ distortion level L_d and the light line near the bottom of the figure shows the noise level L_n at $2f_1-f_2$.

To cover a two-dimensional region of L_1, L_2 space with continuously varying stimulus level, we used a stimulus that follows a Lissajous path. In this stimulus, the L_1 and L_2 levels were each amplitude modulated such that the levels varied sinusoidally in dB. Figure 1 provides an example of the stimulus and response levels measured in the ear canal of one subject for a typical DPOAE measurement using a Lissajous-path stimulus. In this case, $f_2=4$ kHz, $f_1=f_2/1.22$, L_2 varied from 17 to 69 dB SPL, and L_1 varied from 43 to 76 dB SPL. The upper two lines show the time course of the primary-tone levels. The solid line represents L_1 and the dashed line represents L_2 over the course of 32 s, which comprised the first half of the stimulus buffer. In the bottom portion of the figure, the thick line varying around 0 dB SPL represents the time course of DPOAE level at $2f_1-f_2$ and the thin line represents the corresponding noise level at this frequency. The variation in DPOAE level is expected because of the wide range over which the primary levels are varying. At its maximum, which occurred when stimulus levels were large, the DPOAE level is about 15 dB SPL. The noise was computed as the instantaneous difference between the first and second halves of the response buffer. The noise level varies from about -10 to less than -30 dB SPL. There were only a few moments in time when the DPOAE level was at or below the noise level. This situation usually occurred when L_2 was near its lowest level and was often associated with transient acoustic artifacts, presumably due to swallowing or cable rub. These data demonstrate that the present procedure can produce DPOAE levels that exceed the noise floor for a wide range of stimulus-level conditions.

In this study, all Lissajous-path stimuli had 23 cycles of L_1 modulation and 17 cycles of L_2 modulation in each half of the stimulus buffer. Only the first half of the response buffer is shown in Fig. 1. The primary-tone levels from Fig. 1 are replotted as thin lines in Fig. 2 to illustrate the Lissa-

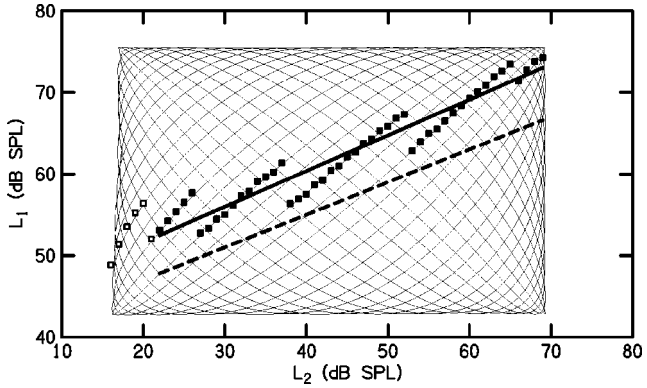


FIG. 2. Lissajous-path measurements in L_1, L_2 space. The thin line shows the same L_1 and L_2 levels from Fig. 1. The squares indicate the L_1 at which L_d was largest at each L_2 in 1-dB steps. Open squares did not meet the 9-dB SNR criterion (see the text). The thick, solid line is a fit to the filled squares, where the SNR was $\geqslant 9$ dB, and so represents an optimal linear path through L_1, L_2 space. The thick, dashed line shows the scissors path recommended by Kummer *et al.* (1998).

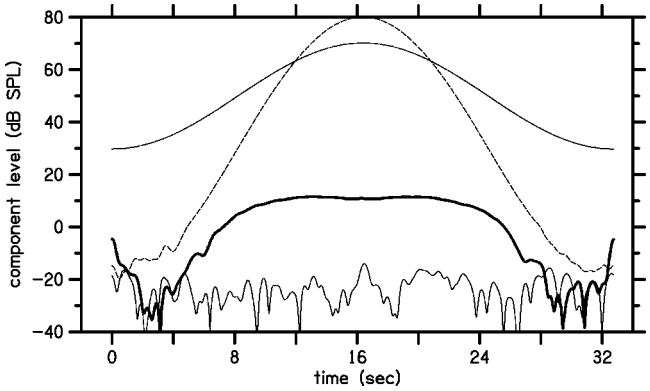


FIG. 3. Levels of measured components as a function of time in response to a linear-path stimulus. The two thin lines near the top of this figure show the levels of the two tones for a scissors-path stimulus. L_1 (solid line) varies from 31 to 71 dB SPL, while L_2 (dashed line) varies from −20 to 75 dB SPL, each with 1 modulation cycle. The heavy line near 0 dB SPL shows the $2f_1-f_2$ distortion level L_d and the light line near the bottom of the figure shows the noise level L_n at $2f_1-f_2$.

jous path covered by this stimulus in L_1, L_2 space. Every point along the Lissajous path represents an L_1, L_2 combination for which a DPOAE level was measured.

A roughly 2-Hz bandwidth was used for the heterodyne analysis of the Lissajous-path response. This was accomplished by applying a 12-Hz-wide Blackman window to the response spectrum that was centered on the frequency component of interest ($2f_1-f_2$). This 12-Hz portion of the spectrum was shifted in frequency to the origin of the frequency axis, then inverse-Fourier transformed to complete the heterodyne analysis. The resulting time-domain signal is complex and represents the time course of the amplitude and phase of the selected frequency component.

C. Optimal path

Because one of our objectives was to examine the variability in DPOAE I/O functions across normal-hearing ears, we wanted to maximize signal-to-noise ratio (SNR) in our I/O functions by selecting L_1 to produce the largest L_d at each L_2. We used the response to a Lissajous-path stimulus to determine a linear path through L_1, L_2 space that would produce the largest L_d at each L_2, which we defined as the "optimal path." This was accomplished by first examining the set of all L_d values at each L_2 in 1-dB steps. We then identified the L_1 associated with the largest L_d at each L_2. In determining the largest L_d, we applied an SNR criterion, defined as the difference (in dB) between the DPOAE level and noise level. As before, the noise level was computed as the instantaneous difference between the two halves of the response buffer, which made it sensitive to the presence of transient acoustic artifacts. When determining the largest L_d at each L_2, we only used points with at least a 9-dB SNR; however, if no response point met this criterion, then L_d was selected without imposing any SNR criterion. The L_1 values that produced the largest L_d are indicated by squares in Fig. 2. Open squares denote the L_1 values where the SNR criterion was not met. This latter result was observed mostly when L_2 was at or below 20 dB SPL

To define an optimal linear path through L_1, L_2 space, we used a linear regression on the subset of the optimal L_1 values for which the 9-dB SNR criterion was met. The thick, solid line in Fig. 2 is an example of an optimal path obtained by this method. For comparison, the thick, dashed line in Fig. 2 represents the linear path $L_1 = 39 + 0.4 \cdot L_2$ recommended by Kummer *et al.* (1998). This linear path has been called "pegelshere" (Janssen, 1995, in German) or "level scissors" (Kummer *et al.*, 2000). We will call it simply the *scissors* path to distinguish it from *optimal* paths that we derived from our *Lissajous*-path measurements. Note that in this example, the optimal path derived from the Lissajous-path measurements differs from the scissors path, with the optimal path occurring at L_1 values that were about 5 dB higher than the scissors path. Stated another way, the scissors path did not result in the largest DPOAEs for this subject at this f_2.

D. Linear-path stimuli

In addition to their use in Lissajous-path stimuli, continuously varying levels also can be used to measure DPOAE levels along any linear path in L_1, L_2 space. To illustrate this method, Fig. 3 shows the response to a stimulus in which the primary levels followed the scissors path with $f_2 = 4$ kHz. The range of L_2 was from −20 to 80 dB SPL, with only one cycle of modulation over the same time span as used for the Lissajous-path stimuli. Because the stimulus levels are modulated at a slower rate in the linear-path stimulus, the analysis bandwidth can be reduced, thereby increasing the SNR without causing excessive smoothing of the time course of the DPOAE level. In this example, the upper thin and dashed lines represent L_1 and L_2 over the 32-s buffer. The thick and thin solid lines towards the bottom of the figure represent the DPOAE and noise levels, respectively. Because the stimulus levels ramp up (0 to 16 s) and down (16 to 32 s), two independent DPOAE I/O functions are measured, one for ascending levels and one for descending levels. This inherent replication of the I/O functions helps to validate the

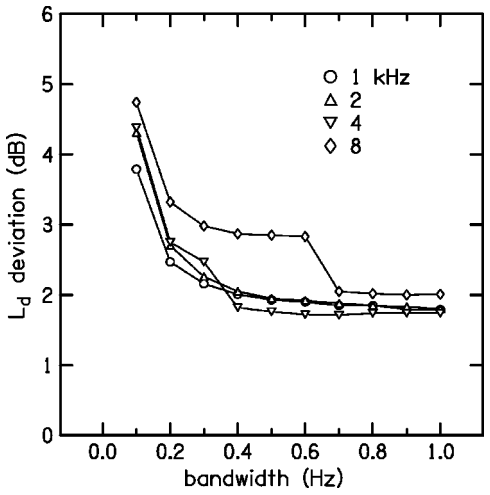


FIG. 4. Deviation of continuous-level measurements from discrete-level measurements as a function of heterodyne analysis bandwidth. The parameter in this figure is f_2. The average, absolute L_d difference increases at small bandwidths due to excessive smoothing of the continuous-level measurement.

DPOAE levels measured by this method. Note that the noise was at or below -20 dB SPL in this example, despite the lack of artifact rejection.

E. Analysis bandwidth

Comparisons of DPOAE levels using continuous-level stimuli (as shown in Fig. 3) with those obtained using discrete-level stimuli led to the observation that the agreement in DPOAE levels between these methods was sensitive to the bandwidth used for the heterodyne analysis. If the bandwidth was too small, then excessive smoothing of the DPOAE response caused the continuous-level I/O function to lie below the discrete-level I/O function. If the bandwidth was too large, then excessive noise decreased the SNR, causing the continuous-level I/O function to deviate randomly from the discrete-level I/O function. To quantify the deviation between these two types of I/O functions, as a means to determine the bandwidth that resulted in the best compromise between smoothing and SNR, we used the absolute difference (in dB) between the continuous-level and discrete-level measurements averaged across all subjects and across all L_2 levels where the SNR was greater than 10 dB. The influence of analysis bandwidth on I/O function deviation is shown in Fig. 4 for each f_2. Although Fig. 4 does not illustrate the adverse effect of increasing noise with increasing bandwidth, it was decided on the basis of this figure that 0.7 Hz represented a reasonable trade-off between excessive smoothing and excessive noise. This bandwidth was used for all subsequent extraction of DPOAE levels obtained with linear-path, continuous-level stimuli.

F. Procedures

DPOAEs were measured in one ear of each subject at four f_2 frequencies (1, 2, 4, and 8 kHz) with $f_1 = f_2/1.22$. The following DPOAE measurements were made at each frequency. First, a Lissajous-path measurement, similar to the one shown in Fig. 1, was made to determine the optimal

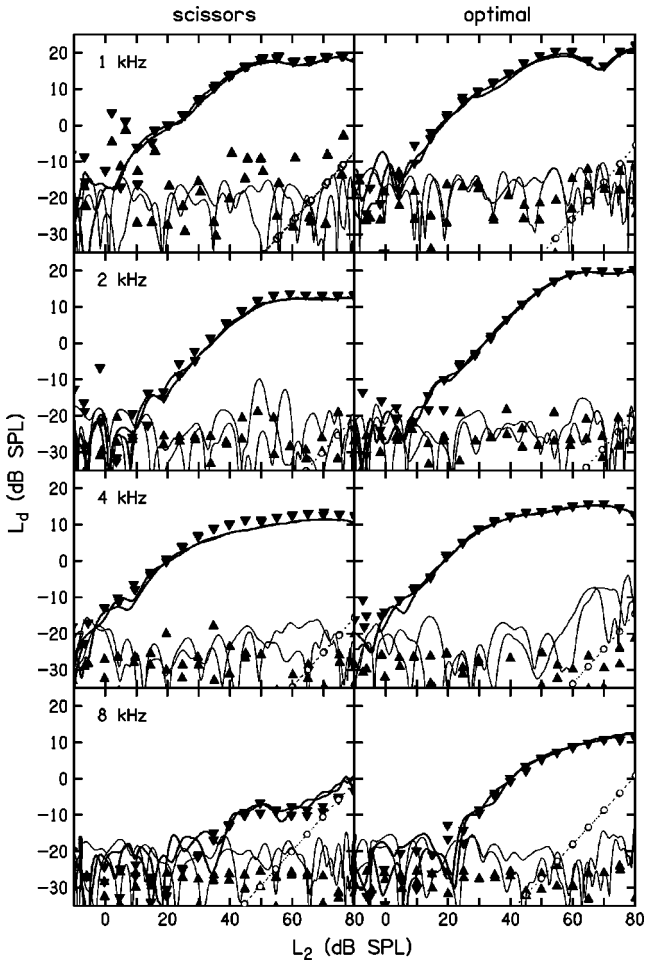


FIG. 5. DPOAE level comparison between continuous and discrete I/O functions. Results are shown for both scissors (left column) and optimal (right column) paths. Each row shows a different f_2 frequency. DPOAE levels (inverted triangles) and corresponding noise levels (upright triangles) are shown for the discrete-level stimulus at 5-dB intervals. The estimated system-distortion levels (circles) for these measurements are also shown. The DPOAE levels (thick lines) and corresponding noise levels (thin lines) for the continuous-level stimulus are superimposed.

linear path for that ear. Then, the DPOAE I/O function was measured in four different ways, using both discrete-level and continuous-level stimuli with both scissors and optimal paths. All discrete-level measurements were repeated because the continuous-level stimuli produced both ascending and descending versions of the I/O functions (see Fig. 3).

III. RESULTS

First, we describe some observations of typical DPOAE I/O function measurements from a single subject. Then, we summarize the results of our I/O function measurements from all 20 subjects.

A. Typical I/O functions

DPOAE I/O functions obtained with linear-path, continuous-level primaries are compared in Fig. 5 with I/O functions for the same linear paths using discrete levels. The I/O functions on the left are for scissors-path stimuli, while those on the right are for optimal-path stimuli. Each row shows results for a different f_2. The symbols represent mea-

surements using discrete stimulus levels (inverted triangles for DPOAE and upright triangles for noise), while the lines represent measurements using continuous-level stimuli (thick for DPOAE and thin for noise). The panel on the left in Fig. 5 labeled ''4 kHz'' shows the same continuous-level data shown in Fig. 3; however, L_{d} and L_n are now plotted as functions of L_2.

Although difficult to distinguish in Fig. 5, each I/O function appears twice for different reasons. For the discrete-level stimulus, each stimulus condition was repeated. For the continuous-level stimuli, one I/O function represents the ascending level portion of the stimulus, while the other represents the descending portion (see Fig. 3). For either procedure, the superimposed DPOAE levels were within 1–2 dB of each other for L_2 conditions that were at least 10 dB above DPOAE threshold (the lowest L_2 at which a response was measured above the noise floor). Thus, the responses were highly reproducible within an ear. Perhaps more importantly, there is good agreement between the DPOAE levels observed for continuous-level and discrete-level measurements in this example. These data from a single subject provide support for the validity of the continuous-level procedure that was used to obtain the data. Below we will provide further validation, based upon data from all 20 subjects. Using either the previously described scissors path or the newly obtained optimal path, our continuous-level I/O procedure resulted in very similar DPOAE levels compared to measurements using discrete stimulus levels.

In general, DPOAE levels for the optimal path (right column) in Fig. 5 are larger than those for the scissors path (left column). Below, we extend this observation to all subjects. At 8 kHz, the DPOAE level for the scissors path (lower-left panel) was not much higher than our estimate of system distortion (circles), whereas the corresponding DPOAE levels for the optimal path (lower-right panel) were at least 10 dB higher than our system distortion. In this subject, the optimal path permits reliable measurements of DPOAE levels at 8 kHz that are not possible with the scissors path.

Figure 6 shows the DPOAE phases associated with the DPOAE levels in Fig. 5, and follows the same convention for lines and symbols. As demonstrated for the DPOAE level measurements in Fig. 5, phase measurements were highly reproducible within a subject once the SNR was positive. There is good agreement between the continuous-level and discrete-level DPOAE phase measurements, just as there was for the DPOAE level measurements in Fig. 5. The variability on the left side of each panel is due to the poor SNR for those stimulus conditions, which made phase measurements less reliable. Also of interest in Fig. 6 is the decreasing trend of phase with increasing stimulus level. The phase decrease can be interpreted as a latency decrease, such as would be expected if the DPOAE generation site moved closer to the base of the cochlea with increasing stimulus level.

B. Optimal paths

The optimal paths for all subjects are superimposed as the thin, solid lines in Fig. 7 to illustrate the distribution of these paths across subjects. Each row shows functions de-

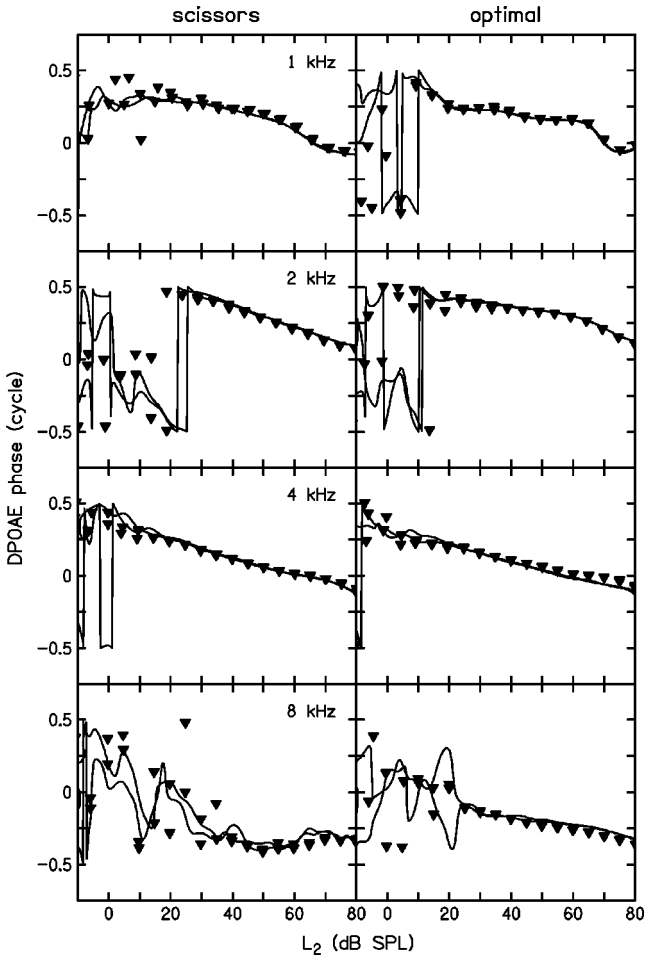


FIG. 6. DPOAE phase comparison between continuous and discrete I/O functions. Results are shown for both scissors (left column) and optimal (right column) paths. Each row shows a different f_2 frequency. DPOAE phases (inverted triangles) are shown for the discrete-level stimulus at 5-dB intervals. The DPOAE phases (thick lines) for the continuous-level stimulus are superimposed. No attempt was made to unwrap the phase in this figure. Evidence of phase wrapping can be seen in several panels, but only where the stimulus levels were low and phase values less reliable.

rived from data for a different f_2. The dashed lines show the scissors path for comparison. Recall that the scissors path was found to be independent of frequency (Kummer *et al.*, 2000); thus, the dashed line is the same in each panel. Note that the optimal paths in each panel tend to lie above the scissors path. Thus, for these 20 subjects, the largest DPOAE for most L_2 levels was observed when L_1 was set higher than recommended by the scissors equation. This effect became progressively more pronounced as frequency increased. At 4 kHz and especially at 8 kHz, the optimal L_1 exceeded the scissors values in practically every subject at all L_2 levels. The slopes of the optimal paths tends to be steeper than the slope of the scissors path when $f_2 = 1$ kHz and less steep when $f_2 = 8$ kHz.

These trends are summarized in Table I, which lists the average optimal path at each frequency, as well as an overall average for the present set of data. Because the optimal L_1 is greater than L_2 at low levels and its slope is less than 1, there will be some value of L_2 for which $L_1 = L_2$ will be optimal. That level is also listed in Table I under the heading *equal*. Finally, comparable values suggested by Kummer *et al.*

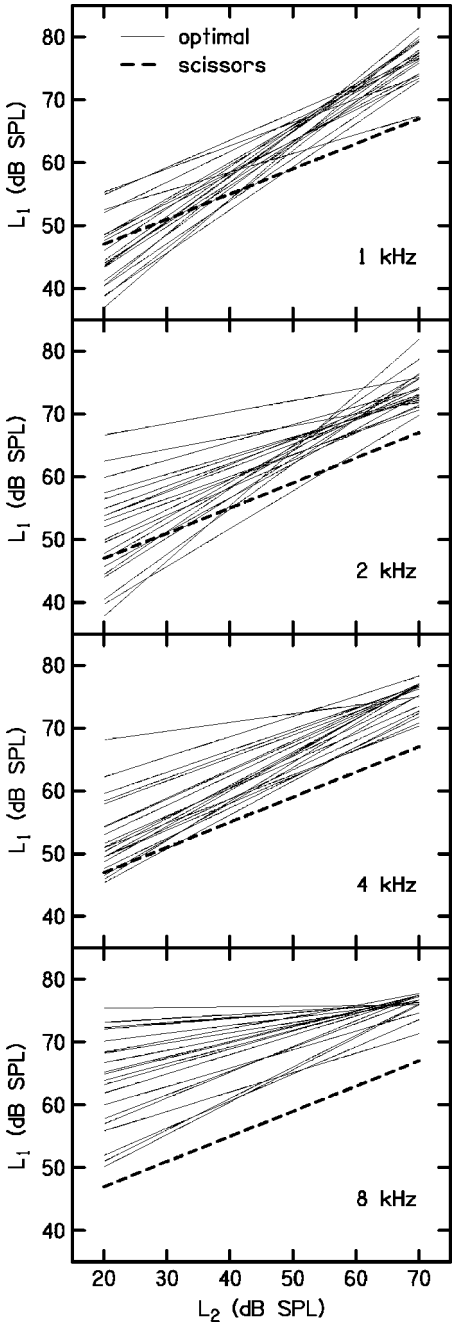


FIG. 7. Optimal paths for all 20 subjects (thin, solid lines) are compared with the scissors path (thick, dashed line). Each panel shows data for a different f_2. This figure illustrates the distribution of the optimal paths across subjects at each frequency, as well as their deviation from the scissors path.

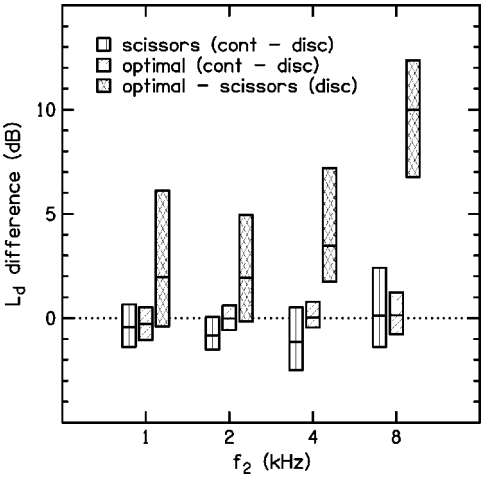


FIG. 8. Distribution (median and interquartile range) of DPOAE level differences at each frequency. The scissors (vertical hatch) and optimal (diagonal hatch) distributions compare the continuous-level measurement with the discrete-level measurement. These distributions are relatively narrow and centered near zero, indicating good agreement between the continuous-level and discrete-level measurements, especially for the optimal path. The distributions of differences between the optimal path and the scissors path for discrete-level stimuli (cross hatch) are also shown at each frequency. These distributions tend to be centered above zero, indicating higher DPOAE levels were measured for the optimal path, especially when $f_2 = 8$ kHz.

(1998), Kummer *et al.* (2000), and Whitehead *et al.* (1995b) are also provided. The overall average in the present study is closer to the value reported by Whitehead *et al.* than it is to the values reported by Kummer *et al.* However, the present data suggest that there is a systematic effect of frequency, which is obscured in the average across frequency. Only the optimal linear path at 2 kHz from the present study approximates the path recommended by Kummer *et al.*

C. I/O function comparisons

To assess the similarity between the I/O functions obtained with discrete-level and continuous-level stimuli, we examined the differences between I/O functions obtained by these two methods for both the scissors and optimal paths. Figure 8 shows the distribution of these differences in the first two boxes at each of four test frequencies. These data included all subjects and all values of L_2 for conditions in which the SNR was greater than 10 dB. Separate distributions of these differences are shown for the scissors path

TABLE I. Coefficients of the optimal linear path $L_1 = a + b \cdot L_2$ averaged across all subjects at each frequency (kHz). The *equal* column indicates the stimulus level at which $L_1 = L_2$. The linear fit suggested by Kummer *et al.* (1998) was based on the data of Gaskill and Brown (1990).

Study	Subjects	Frequency (kHz)	*a* (dB SPL)	*b*	Equal (dB SPL)
Present	20	1	33	0.62	87
		2	42	0.46	77
		4	44	0.45	79
		8	58	0.26	78
		Average	44	0.45	78
Whitehead *et al.* (1995a)	8	3	45	0.5	90
Kummer *et al.* (1998)	5	Average	39	0.4	65
Kummer *et al.* (2000)	22	Average	41	0.41	70

(vertical hatch) and optimal path (diagonal hatch). Thus, the first two boxes at each frequency compare the DPOAE levels for continuous and discrete presentation of the stimuli, holding constant the paradigm that was used to set stimulus level (scissors or optimal path). The boxes in Fig. 8 show the interquartile range (25th to 75th percentile) of the DPOAE level differences. The horizontal line through the middle of each box shows the median (50th percentile) of the distribution. The fact that the scissors (vertical-hatch boxes) and optimal (diagonal-hatch boxes) are relatively short with medians near zero indicates that the I/O functions for continuous levels are similar to the I/O functions for discrete stimulus levels regardless of whether the levels are selected according to the scissors or optimal equation. Since the discrete-level method represents the standard approach for measuring DPOAE I/O functions, the similarity between DPOAE levels measured with these two methods validates the continuous-level method.

The cross-hatch boxes in Fig. 8 represent the distribution of DPOAE level differences when the optimal path is compared to the scissors path for measurements with discrete stimulus levels. In contrast to the first two boxes at each f_2, these boxes are relatively tall with medians always above zero, indicating that the optimal path produces larger DPOAE levels than the scissors path. The advantage of the optimal path increases as frequency increases and is greatest at 8 kHz, where the median difference in DPOAE level is about 10 dB. It is not surprising that the optimal path results in responses that are at least as large as those observed with the scissors path, since the stimulus levels used for the optimal-path measurements were, by definition, the levels that produced the largest DPOAE levels. However, the size and direction of the differences and the fact that these differences were observed in about 70% of the stimulus conditions at 1 and 2 kHz and 90%–100% of the cases at 4 and 8 kHz suggests that the scissors path does not, on average, produce the largest DPOAEs in subjects with normal hearing. The DPOAE level differences when the optimal path is compared to the scissors path for measurements with continuous-level stimuli are not shown in Fig. 8, but are similar to the differences obtained with discrete-level stimuli.

D. Measurement time

One of the advantages that the continuous-level stimulus may offer over discrete-level stimuli is the time required to measure a DPOAE I/O function. The continuous-level paradigm includes a fixed amount of time, independent of the conditions of the measurement. Specifically, 262, 131, 66, and 66 s of data-collection time were used at 1, 2, 4, and 8 kHz, regardless of noise levels. Furthermore, this paradigm does not incorporate artifact rejection, so transient spikes potentially could be included in the response. In contrast, the time required to measure an I/O function using discrete-level stimuli may vary from condition to condition in the present study because measurement-based stopping rules controlled when each measurement was terminated. These rules allowed averaging to continue until either noise level was less than -25 dB SPL or the length of artifact-free averaging time was 32 s, whichever occurred first. Thus, the use of measurement-based stopping rules and the inclusion of artifact rejection potentially could increase the total discrete-level measurement time. These factors should be kept in mind when comparing the data-collection times for the continuous-level and discrete-level conditions.

TABLE II. Average time (s) required for measurement of two I/O functions for each level variation method (discrete or continuous) at each frequency (kHz). For the discrete-level stimuli, the two I/O functions represent repeated conditions. For the continuous-level stimuli, the two I/O functions represent ascending and descending levels.

Frequency (kHz)	Discrete (s)	Continuous (s)
1	999	262
2	601	131
4	340	66
8	272	66
Average	553	131

The average time (across subjects) that was required to measure two scissors or two optimal path I/O functions using discrete-level stimuli is listed in Table II for each f_2 frequency. The discrete-level I/O functions were repeated for comparison with continuous-level stimuli, because the latter always produces two I/O functions, one for ascending levels and one for descending levels. On average, the time required to measure I/O functions using the discrete-level method was quadruple the time required for the continuous-level method.

The data-collection times listed in Table II should be interpreted in relation to the noise levels that were achieved by the continuous-level and discrete-level paradigms. That is, the advantage in averaging time for the continuous-level paradigm would not be impressive if it were achieved at the expense of a concomitant increase in noise levels. Table III compares the noise levels at each frequency for the two different paradigms. As can be seen in this table, the additional averaging time allowed the discrete-level measurements to achieve lower noise levels. The discrete-level noise levels were lower than those achieved by the continuous-level paradigm, ranging from 2.4 dB lower (at 1 kHz) to 6.9 dB lower (at 8 kHz), with an overall average that was lower by 4.2 dB. These results suggest that the data-collection time advantage of the continuous-level paradigm was essentially counterbalanced by the increased noise levels at 8 kHz (i.e., a fourfold decrease in averaging time would be expected to result in a 6-dB increase in noise level). However, at all other frequencies (and in the average across frequency), the reduced aver-

TABLE III. Average noise level (dB SPL) for each level variation method (discrete or continuous) at each frequency (kHz). The averages include both scissors and optimal paths because the noise level was independent of stimulus level.

Frequency (kHz)	Discrete (dB SPL)	Continuous (dB SPL)
1	-20.9	-18.5
2	-27.5	-24.3
4	-29.1	-24.5
8	-29.1	-22.2
Average	-26.6	-22.4

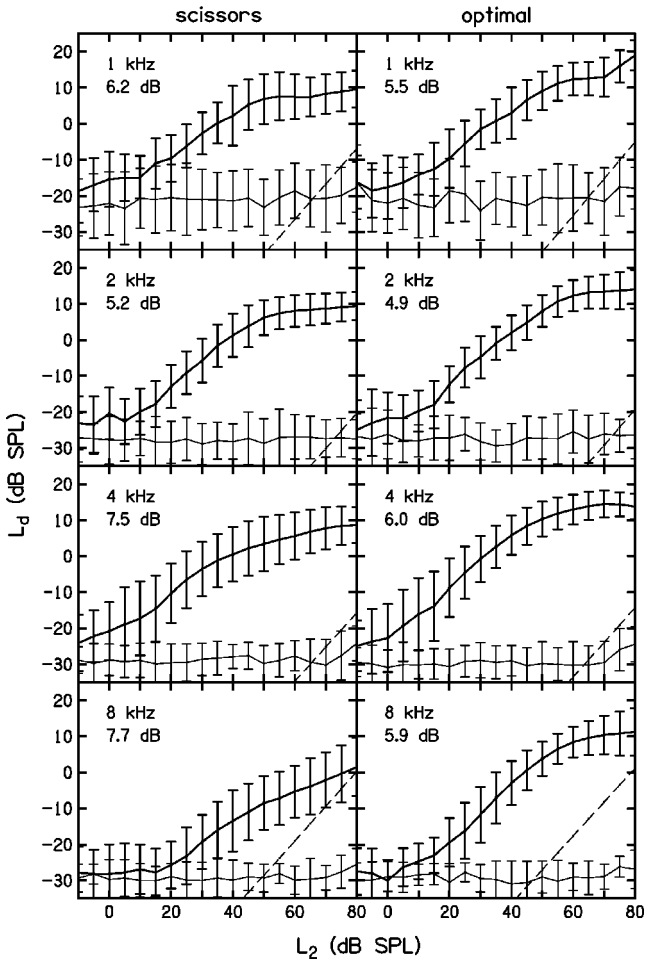


FIG. 9. Average DPOAE I/O functions for discrete-level measurements. The mean DPOAE level (thick line) and corresponding noise level (thin line) are shown for each f_2 frequency for both scissors (left column) and optimal (right column) paths. The error bars indicate 1 standard deviation from the mean. Dashed lines represent estimated system distortion based on cavity measurements. The dB values in each panel are the average of all standard deviations with SNR $\geqslant$10 dB.

aging time for the continuous-level paradigm did not result in an equivalent increase in noise level.

E. Optimal versus scissors for discrete-level I/O functions

The discrete-level DPOAE I/O functions, averaged across subjects, are shown in Fig. 9. The thick line represents the mean DPOAE level at $2f_1-f_2$ and the thin line represents the noise at this frequency. The format of Fig. 9 is similar to the one used in Fig. 5, with data from the scissors path on the left, optimal path on the right, and a different f_2 in each row. The error bars in Fig. 9 represent plus or minus 1 standard deviation from the mean for each L_2 from -5 to 80 dB SPL. Thus, the error bars in this figure represent variability for all of the data, including those conditions for which the SNR is low ($L_2 \leqslant 20$ dB SPL). The dB value in each panel is the average standard deviation across the L_2 for which the SNR was greater than 10 dB. Thus, it represents an estimate of the intersubject variability in DPOAE level, restricted to conditions where the response was reliably measured. The average level produced by the optimal-level path

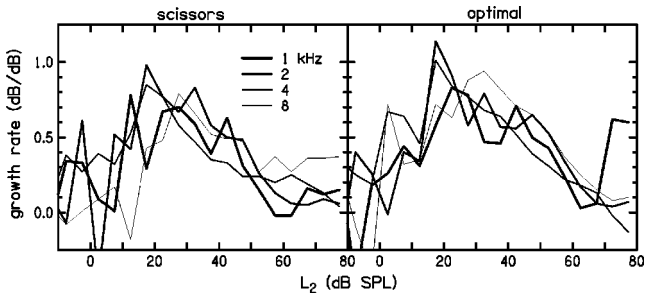


FIG. 10. DPOAE growth rate as a function of stimulus level for scissors (left) and optimal (right) discrete-level measurements. These growth rates are the slopes of the I/O functions shown in Fig. 9. The growth rate for the optimal path shows less variability across frequency compared to the scissors path.

exceeds that observed for the scissors path, especially at 8 kHz, consistent with the distribution of DPOAE level differences shown in Fig. 8. Visual inspection of the functions and their associated error bars suggests that there is little difference in the variability that is observed across subjects. However, comparing the optimal path to the scissors path, we see that the average standard deviation always decreased, especially at 8 kHz. In other words, the variability averaged across L_2 tends to be less for the optimal path than for the scissors path. Stated another way, there is less intersubject variability in DPOAE levels when individually determined optimal level paths are used, an effect that is most pronounced at the two f_2 frequencies for which the DPOAE-level advantage is greatest (4 and 8 kHz). Still, the differences are not large, suggesting that there are other sources contributing to intersubject variability.

One additional feature that appears to distinguish the optimal path from the scissors path is the similarity in the shape of the I/O functions across frequency. This feature can be observed directly in Fig. 9 and indirectly by noting the similarity in the slopes of the I/O functions shown in Fig. 10. The growth rate (dB/dB) in Fig. 10 was computed from the slopes of the I/O functions in Fig. 9. Greater emphasis should be placed on the slopes for L_2 levels greater than about 20 dB SPL, because the SNR was more favorable at these levels for both paradigms and, therefore, the results are considered more reliable. At levels below about 20 dB SPL, there was less separation between signal and noise, resulting in less confidence in our calculations of slope and, therefore, a more cautious assessment of slope estimates at these levels. The similarity in the optimal-path growth rates across frequency in Fig. 10 appears to be greater than that observed for the scissors path for L_2 levels ranging from 20 to about 65 dB SPL. Stated another way, the optimal-level path resulted in DPOAE I/O functions that were less dependent on frequency.

The average phase of the I/O function across subjects was obtained by averaging complex pressure values at $2f_1-f_2$, and is shown in the upper half of Fig. 11. Only measurements which had an individual SNR greater than 3 dB were included in the average. The lower half of Fig. 11 shows the vector strength (Goldberg and Brown, 1969) associated with each phase average. The line weights in Fig. 11 follow the convention used in Fig. 10. Just as we observed in

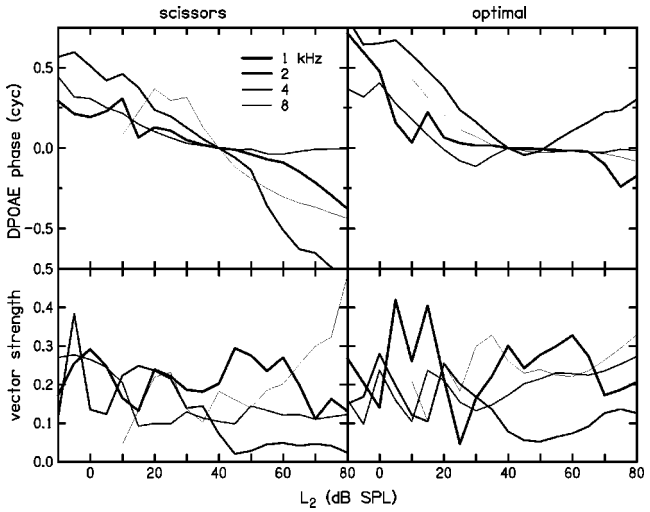


FIG. 11. Average DPOAE phase (upper) and vector strength (lower) as a function of stimulus level for scissors (left) and optimal (right) discrete-level measurements. Different frequencies are represented by varying linewidths. The complex pressure at 2 f_1-f_2 was averaged across all subjects to obtain these phase values. To facilitate comparison, each phase curve has been unwrapped and shifted vertically so that its value would be zero at 40 dB SPL.

the individual data shown in Fig. 6, the average DPOAE phase in Fig. 11 has a tendency to decrease with increasing stimulus level at all frequencies and for both scissors and optimal paths. The total decrease in phase tends to be about 0.5 to 1 cycles between 0 and 80 dB SPL, independent of frequency. This trend is observed despite low values of vector strength, which indicate broad distributions of individual phase values across subjects. The phase decrease for the scissors path (left panel) tends to be uniform across all levels, whereas, for the optimal path, the phase decreases more below 40 dB SPL than it does at higher levels. For the optimal path, there was little phase change between 40 and 60 dB SPL, except at 2 kHz, where the phase increased above 45 dB SPL. While the general trend was a decrease in phase with increasing stimulus level, regardless of the particular stimulus path, this result should be interpreted cautiously, given the variability across subjects suggested by the low vector-strength values.

IV. DISCUSSION

The major findings from this study are (1) the development of a stimulus paradigm that allows one to explore a large L_1,L_2 space in a relatively short period of time; (2) validation of this technique by demonstrating the similarity in DPOAE levels obtained with it to the DPOAE levels obtained by more traditional measurement techniques at comparable stimulus levels; (3) demonstration that the optimal-level path depends on frequency; (4) demonstration that the optimal-level path results in larger responses than another commonly used paradigm for selecting stimulus level; and (5) that the DPOAE I/O functions are more similar across frequency and less variable across subjects when the optimal-level path is used, compared to the scissors path.

A. Analysis bandwidth

A 2-Hz bandwidth was selected for heterodyne analysis of the Lissajous-path responses because a smaller bandwidth would distort the shape of the Lissajous pattern formed by the primary levels and a larger bandwidth would decrease the SNR. This bandwidth was too small to allow accurate representation of DPOAE levels because these levels were changing so rapidly. In other words, the smoothing of the DPOAE level by the heterodyne analysis caused these levels to differ from those measured with discrete-level stimuli. However, because smoothing did not alter the locations of the most prominent DPOAE-level maxima in L_1,L_2 space, the Lissajous-path measurements retained sufficient information to determine an optimal path, despite the smoothing.

Possible evidence of excessive smoothing can be observed in the large deviation of some squares from the solid line in Fig. 2. The main reason for these large deviations was because DPOAE level tended to be larger during portions of the stimulus when L_1 and L_2 were both increasing. The dependence of DPOAE levels on the direction of stimulus level changes indicates that the stimulus levels were changing too rapidly to approximate steady-state conditions; however, this dependence was apparently exaggerated by the narrowness of the analysis bandwidth. Local maxima in DPOAE level were spread over a wider range as a result of smoothing. We observed that increasing the analysis bandwidth reduced the largest deviations of the squares from the solid line in Fig. 2; however, it was decided that the present amount of smoothing was acceptable and was selected because it provided better SNR than was provided by the larger bandwidth. Another way to reduce the influence of smoothing would have been to reduce the rate of stimulus-level changes, either by increasing the duration of stimulus or by reducing the number of cycles of level modulation. Both of these modifications to the stimulus have corresponding disadvantages and were not explored in the present study. Considering the alternatives, the 2-Hz analysis bandwidth appears to represent a reasonable compromise between competing needs. Still, the selection of analysis bandwidth for the Lissajous-path measurement warrants further consideration in future studies.

B. Frequency dependence

The data displayed in Fig. 7 suggest that the optimal linear path for maximum DPOAE level depends on frequency, contrary to the suggestion of Kummer *et al.* (2000). The observation of frequency dependence, however, is perhaps not surprising. One might expect the optimal path to have a steeper slope at low frequencies because the basilar-membrane (BM) excitation patterns for these stimuli overlap more than at high frequencies, resulting in more similar BM growth rates for the two tones at the site of DPOAE generation. The excitation patterns for higher frequencies with the same f_2/f_1 will be separated to a larger degree, which might result in more different growth rates at the DPOAE generation site. If the growth of BM response to $f_1=6.7$ kHz at the $f_2=8$-kHz place is more rapid than the $f_1=0.83$ kHz response growth at the $f_2=1$-kHz place, then the rate at which L_1 will need to increase to maintain the maximum interac-

tion at the f_2 place will depend on frequency. Under these conditions, maintaining similar BM displacements at the generation site for the two tones requires L_1 to change level much more slowly than L_2 when $f_2=8$ kHz than when $f_2=1$ kHz. Presumably, this is why the optimal path has a shallower slope at high stimulus frequencies, compared to lower frequencies.

According to this view, one would predict an interaction between primary-level ratio and primary-frequency ratio. Thus, the frequency dependence of the optimal-path slope might be reduced by selecting a smaller f_2/f_1 ratio for higher frequencies. In this way, the relative overlap between the BM excitation patterns for f_1 and f_2 could be made to be more consistent across frequency. Stated another way, either decreasing f_2/f_1 or increasing L_1 can compensate for differences in frequency selectivity as a function of place along the basilar membrane. By decreasing f_2/f_1, the relative growth rates between the BM responses at the place of DPOAE generation would be more consistent. This consistency should be reflected in the slope of the optimal path. In support of this interpretation, Mills (2002) observed in gerbils that the f_2/f_1 producing the largest DPOAE levels decreases as f_2 increases.

C. Measurement time

The advantage of reduced measurement time for the continuous-level I/O function compared to the discrete-level I/O function, as shown in Table II, should also take into account the higher noise level for the continuous-level I/O functions listed in Table III. At 8 kHz, the increased noise levels essentially negated the benefit of the decreased data-collection time. While the continuous-level paradigm resulted in higher noise levels than the discrete-level paradigm at other frequencies, the increased noise level was less than would be predicted by the decreased data-collection time. If noise equivalency is a goal, we should be able to achieve similar noise levels with continuous-level measurements by doubling the measurement time at 1 and 2 kHz and quadrupling the measurement time at 4 and 8 kHz; however, this increase in measurement time would reduce its advantage over discrete-level measurements to less than a factor of 2. An alternative way to achieve a similar noise floor with the continuous-level measurement would be to decrease the analysis bandwidth; however, this approach would lead to about 1- or 2-dB greater deviations from DPOAE levels measured with discrete-level stimuli. Considering the inescapable trade-offs among (1) measurement time; (2) noise level; and (3) DPOAE-level accuracy, the apparent advantage of the continuous-level measurement appears smaller than suggested by the times listed in Table II. However, continuous-level measurements may still be preferable to discrete-level measurements in some situations, especially when accuracy of DPOAE levels is less important, such as in the determination of the optimal linear path. Another situation in which the continuous levels could offer a significant advantage over discrete levels is when the SNR is sufficiently high that the transient responses due to switching stimulus levels occupy a significant percentage of the averaging time required at each level. This may be the case in typical laboratory studies because the SNR is usually much greater in experimental animals than in humans, so avoiding transient-response time has a greater influence on total measurement time.

D. I/O function variability

The similarity across frequency of the DPOAE I/O function growth rates shown in Fig. 10 for the optimal path suggests that the optimal path reduces individual variability in I/O functions by compensating for individual-ear difference in the ear-canal geometry and/or middle-ear transmission. The compressive shape of the DPOAE I/O function is due to compressive growth of BM vibrations. The similarity of DPOAE growth across frequency suggests a corresponding similarity in BM growth across frequency. Individual differences in ear-canal and/or middle-ear characteristics could affect the relative levels of primary-tone responses within the cochlea and, as a consequence, cause the DPOAE growth rates to differ, despite consistency of BM growth rates for single tones. The use of optimal-path stimuli can compensate for these differences in individual ears. Variability across ears, however, cannot be completely attributed to level differences reaching the site of generation, because variability remained even after individualized optimal-level paths were used.

Similarity of the I/O functions across frequency provides evidence that the cochlear mechanisms that cause BM growth to be nonlinear have relatively invariant characteristics along the length of the cochlea, at least for the cochlear regions over which frequencies from 1 to 8 kHz are represented. The assumption of such invariance in BM growth has been used by Keefe (2002) to compute estimates of middle-ear forward- and reverse-transfer functions based on measurements of DPOAE I/O functions. Miller and Shera (2002) observed, however, that the I/O functions used by Keefe had fixed values of L_1 for each L_2 despite differing values of middle-ear forward transfer at f_1 and f_2. Consequently, the relative level of these two frequency components within the cochlea was not held constant across frequency. The fact that the optimal path I/O functions in the present study were less variable across frequency and across subjects suggests that relative levels within the cochlea were more consistent. Therefore, estimates of middle-ear transfer functions derived from individualized optimal-path I/O functions could be more accurate than those derived from I/O functions measured with any stimuli that have fixed level differences for all subjects. At the very least, the frequency effect described in the present study suggests that the DPOAE I/O functions obtained with a fixed primary-level ratio regardless of frequency may not provide the best data from which to derive estimates of forward and reverse energy transmission.

E. Optimal path

Our results indicate an optimal path through L_1,L_2 space, averaged over subjects and over stimulus frequency, is $L_1=44+0.45\cdot L_2$. This result is consistent with previous

recommendations to make $L_1 > L_2$ at low stimulus levels (Gaskill and Brown, 1990; Whitehead *et al.*, 1995b; Kummer *et al.*, 1998, 2000).

Whitehead *et al.* (1995b), based on measurements in eight human subjects, suggested the use of $L_1 = 45 + 0.5 \cdot L_2$ at 3 kHz, which is close to our optimal path when averaged across frequency. They also observed some frequency dependence to their optimal path by noting that the constant term was less (in the 32- to 42-dB range) at other frequencies (both above and below 3 kHz) when the slope (0.5) was kept constant.

Our optimal path lies above the Kummer *et al.* (1998) scissors path $L_1 = 39 + 0.4 \cdot L_2$, which was based on the data of Gaskill and Brown (1990). According to our results, L_1 needs to be higher than the scissors path to obtain the largest DPOAE levels at each L_2. Kummer *et al.* (2000), based on measurements in 22 subjects, essentially confirmed their earlier recommendation for the optimum path, although they suggest a slightly higher path $L_1 = 41 + 0.41 \cdot L_2$. Contrary to our results, they found little frequency dependence in their scissors path, although their methods were similar to ours. One possible reason for the discrepancy between their results and ours could be the more limited dynamic range of their stimuli. Their stimulus levels went no higher than $L_1 = 70$ and $L_2 = 65$ dB SPL. Their maximum measurement time of 4 s per stimulus condition also limited the minimum stimulus levels at which reliable measurements could be made (as evidenced by the SNR). However, the frequency effects described in Figs. 7 and 8 of the present paper are large. Therefore, differences in the range of stimulus levels may not be adequate to explain the differences between the present results and those described in Kummer *et al.* (2000).

The continuous-level method, which allows us to explore a large region of L_1, L_2 space in a relatively short amount of time, estimates the optimal path in less time than traditional methods. Although we have taken advantage of this method to provide average optimal paths at each frequency, as well as averaged across frequency, the preferred approach would be to use individual estimates of the optimal path for each ear. This is because current stimulus calibration methods, whether in-the-ear or in an acoustic cavity, are unable to guarantee that the desired level difference between the two tones will be achieved within the cochlea. Determining the individual optimal path in each ear provides the best way to maximize the DPOAE level as a function of stimulus level. While this approach might be preferable in normal-hearing subjects under laboratory conditions, the problem becomes more complex when measuring responses in possibly hearing-impaired subjects under clinical conditions. Here, the approach that is taken to select stimulus level probably will depend on the question that is being asked. If the aim is to distinguish ears with normal hearing from ears with hearing loss, then it might be more appropriate to select the average normal optimal-level path and apply those values with patients in the clinic. Although this path might not be optimal for individual hearing-impaired subjects (or for individual normal-hearing subjects), it might result in greater separation between the response distributions from normal and impaired ears, which is the goal in dichotomous clinical decisions. On the other hand, individual optimal-level paths might be more useful in clinical applications where the DPOAE I/O function will be used to predict behavioral threshold (Boege and Janssen, 2002; Gorga *et al.*, 2003) or loudness growth (Müller and Janssen, 2004). Under these circumstances, selecting a stimulus path that takes into account any individual differences in calibration and/or energy transfer through the middle ear might provide a better estimate of BM response growth, which is the variable of interest when evaluating the relationship between DPOAE I/O functions and behavioral thresholds or loudness growth.

F. Level-dependent phase shift

It is of some interest that the same trend of decreasing phase with increasing level that was observed for a single subject in Fig. 6 is preserved in the average data from 20 subjects shown in Fig. 11. As stated above, we interpret the decrease in phase to indicate a shift in the place of DPOAE generation of about 1 cycle. The suggestion of a shift towards the base as level increases is not surprising, given the direction in which excitation spreads in the cochlea (e.g., Rhode and Recio, 2000). This result is also consistent with the upward spread of DPOAE suppression (Gorga *et al.*, 2002), which similarly suggests that the place of DPOAE generation moves toward the base of the cochlea as stimulus levels increase.

Shera and Guinan (2003) have estimated that the wavelength of the BM response in a human cochlea is about 1.3 mm at 1 kHz, but decreases with increasing frequency at a rate of about 25% per octave. The level-dependent phase shift in the present study shows no such consistent trend with frequency over a three-octave range. This suggests that the level-dependent shift in the place of DPOAE generation is more likely to be consistent in terms of cycles than distance along the BM.

V. CONCLUSIONS

The Lissajous-path measurement using continuous-level stimuli provides an efficient method to determine an optimal linear path that will usually produce larger DPOAE levels than more traditional approaches to setting primary levels, such as the Kummer *et al.* (1998) scissors path. In addition to producing larger DPOAE levels, the individualized optimal path produces I/O functions that are slightly less variable across subjects and slightly more consistent in shape across frequency. Consequently, DPOAE I/O functions measured with individualized optimal paths may yield more accurate predictions of hearing thresholds or loudness growth.

The optimal path differs across subjects and across frequency. The slope of the optimal path is steeper at low frequencies and less steep at high frequencies. This observation is consistent with direct observations of cochlear mechanics that show nonlinear BM growth more spatially distributed at low frequencies and more localized at high frequencies (Cooper and Rhode, 1995; Robles and Ruggero, 2001). Despite this clear variation in nonlinear growth characteristics across frequency, DPOAE I/O functions maintain a similar shape across frequency. This observation appears to be con-

sistent with psychophysical data that show similar loudness growth across frequency (e.g., Fletcher and Munson, 1933; Müller and Janssen, 2004).

ACKNOWLEDGMENTS

This work was supported by grants R01-DC02251, T32-DC00013, and P30-DC04662 from the National Institutes of Health, NIDCD. We thank Darcia Dierking and Brenda Hoover for their assistance with data collection. Preliminary results of this study were presented at a meeting of the Association for Research in Otolaryngology (ARO) in February, 2004.

American National Standards Institute (**1996**). ANSI S3.6-1996, ''Specifications for audiometers'' (American National Standards Institute, New York).

American Speech-Language-Hearing Association (ASHA) (**1990**). ''Guidelines for screening for hearing impairment and middle-ear disorders,'' Working Group on Acoustic Immittance Measurements and the Committee on Audiologic Evaluation, ASHA Suppl. (2), 17–24.

Boege, P., and Janssen, T. (**2002**). ''Pure-tone threshold estimation from extrapolated distortion product otoacoustic emission I/O functions in normal and cochlear hearing loss ears,'' J. Acoust. Soc. Am. **111**, 1810–1818.

Brown, A. M., and Gaskill, S. A. (**1990**). ''Measurement of acoustic distortion reveals underlying similarities between human and rodent mechanical responses,'' J. Acoust. Soc. Am. **88**, 840–849.

Cooper, N. P., and Rhode, W. S. (**1995**). ''Nonlinear mechanics at the apex of the guinea-pig cochlea,'' Hear. Res. **82**, 225–243.

Fletcher, H., and Munson, W. A. (**1933**). ''Loudness, its definition, measurement, and calculation,'' J. Acoust. Soc. Am. **5**, 82–108.

Gaskill, S. A., and Brown, A. M. (**1990**). ''The behavior of the acoustic distortion product, $2f_1-f_2$, from the human ear and its relation to auditory sensitivity,'' J. Acoust. Soc. Am. **88**, 821–839.

Goldberg, J. M., and Brown, P. B. (**1969**). ''Response of binaural neurons of the dog superior olivary complex to dichotic tonal stimuli: Some physiological mechanisms of sound localization,'' J. Neurophysiol. **32**, 613–636.

Gorga, M. P., Neely, S. T., Ohlrich, B., Hoover, B., Redner, J., and Peters, J. (**1997**). ''From laboratory to clinic: A large scale study of distortion product otoacoustic emissions in ears with normal hearing and ears with hearing loss,'' Ear Hear. **18**, 440–455.

Gorga, M. P., Neely, S. T., Dorn, P. A., Dierking, D., and Cyr, E. (**2002**). ''Evidence of upward spread of suppression in DPOAE measurements,'' J. Acoust. Soc. Am. **112**, 2910–2920.

Gorga, M. P., Neely, S. T., Dorn, P. A., and Hoover, B. M. (**2003**). ''Further efforts to predict pure-tone thresholds from distortion product otoacoustic emission input/output functions,'' J. Acoust. Soc. Am. **113**, 3275–3284.

Harris, F. P., Lonsbury-Martin, B. L., Stagner, B. B., Coats, A. C., and Martin, G. K. (**1989**). ''Acoustic distortion products in humans: Systematic changes in amplitude as a function of f_2/f_1 ratio,'' J. Acoust. Soc. Am. **85**, 220–229.

Hauser, R., and Probst, R. (**1991**). ''The influence of systematic primary-tone level variation $L2-L1$ on the acoustic distortion product emission $2f_1-f_2$ in normal human ears,'' J. Acoust. Soc. Am. **89**, 280–286.

Janssen, T., Kummer, P., and Arnold, W. (**1995**). ''Wachstumsverhalten der Distorsionsproduktemissionen bei kochleären Hörstörungen,'' or ''Growth behavior of distortion-product emission in cochlear hearing-impairment,'' Oto-Rhino-Laryngol. Nova **5**, 34–46 (in German).

Keefe, D. H. (**2002**). ''Spectral shapes of forward and reverse transfer functions between ear canal and cochlea estimated using DPOAE input/output functions,'' J. Acoust. Soc. Am. **111**, 249–260.

Kim, D. O., Dorn, P. A., Neely, S. T., and Gorga, M. P. (**2001**). ''Adaptation of distortion-product otoacoustic emission in humans,'' J. Assoc. Res. Otolaryngol. **2**, 31–40.

Kummer, P., Janssen, T., and Arnold, W. (**1998**). ''Level and growth behavior of the $2f_1-f_2$ distortion-product otoacoustic emission and its relation to auditory sensitivity,'' J. Acoust. Soc. Am. **103**, 3431–3444.

Kummer, P., Janssen, T., Hulin, P., and Arnold, W. (**2000**). ''Optimal $L(1)-L(2)$ primary-tone level separation remains independent of test frequency in humans.'' Hear. Res. **146**, 47–56.

Miller, A. J., and Shera, C. A. (**2002**). ''Using DPOAEs to measure forward and reverse middle-ear transmission noninvasively,'' Assoc. Res. Otolaryngol. Abs. **769**.

Mills, D. M. (**2002**). ''Interpretation of standard distortion product otoacoustic emission measurements in light of the complete parametric response,'' J. Acoust. Soc. Am. **112**, 1545–1560.

Müller, J., and Janssen, T. (**2004**). ''Similarity in loudness and distortion product otoacoustic emission input/output functions: Implications for an objective hearing aid adjustment,'' J. Acoust. Soc. Am. **115**, 3081–3091.

Neely, S. T., and Liu, Z. (**1993**). ''EMAV: Otoacoustic emission averager,'' Tech. Memo. **17**, Boys Town National Research Hospital, Omaha, NE.

Neely, S. T., and Stevenson, R. (**2002**). ''SYSRES,'' Tech. Memo. **19**, Boys Town National Research Hospital, Omaha, NE.

Rhode, W. S., and Recio, A. (**2000**). ''Study of mechanical motions in the basal region of the chinchilla cochlea,'' J. Acoust. Soc. Am. **107**, 3317–3332.

Robles, L., and Ruggero, M. A. (**2001**). ''Mechanics of the mammalian cochlea,'' Physiol. Rev. **81**, 1305–1352.

Shera, C. A., and Guinan, J. J. (**2003**). ''Stimulus-frequency-emission group delay: A test of coherent reflection filtering and a window on cochlear tuning,'' J. Acoust. Soc. Am. **113**, 2762–2772.

Whitehead, M. L., McCoy, M. J., Lonsbury-Martin, B. L., and Martin, G. K. (**1995a**). ''Dependence of distortion-product otoacoustic emissions on primary levels in normal and impaired ears. I. Effects of decreasing $L2$ below $L1$,'' J. Acoust. Soc. Am. **97**, 2346–2358.

Whitehead, M. L., Stagner, B. B., McCoy, M. J., Lonsbury-Martin, B. L., and Martin, G. K. (**1995b**). ''Dependence of distortion-product otoacoustic emissions on primary levels in normal and impaired ears. II. Asymmetry in $L1,L2$ space,'' J. Acoust. Soc. Am. **97**, 2359–2377.

The Allen-Fahey experiment extended

Egbert de Boer[a)]
Room D2-226, Academic Medical Center, University of Amsterdam, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands

Alfred L. Nuttall
Oregon Hearing Research Center, NRC04, Oregon Health & Science University, 3181 SW Sam Jackson Park Road, Portland, Oregon 97239-3098 and Kresge Hearing Research Institute, University of Michigan, 1301 E. Ann Street, Ann Arbor, Michigan 48109-0506

Ning Hu, Yuan Zou, and Jiefu Zheng
Oregon Hearing Research Center, NRC04, Oregon Health & Science University, 3181 SW Sam Jackson Park Road, Portland, Oregon 97239-3098

(Received 9 August 2004; revised 15 December 2004; accepted 15 December 2004)

An ingenious experiment has been performed by Allen and Fahey [J. Acoust. Soc. Am. **92**, 178–188 (1992)], in which they attempted to estimate the gain of the cochlear amplifier by comparing responses to the $2f_1-f_2$ distortion product (DP) in the outer ear canal (otoacoustic emissions) and from an auditory-nerve fiber. Results were essentially negative: no evidence of cochlear amplification was found in that experiment. A variation of that experiment is reported here, where DP responses in the outer ear canal are compared with mechanical responses of the basilar membrane. This variation does not suffer from the major limitation in the original experiment in the choice of possible frequency ratios. Results confirm and extend those of Allen and Fahey entirely. Apparently, the gain of the cochlear amplifier cannot be measured in this way. It is argued that the retrograde wave going to the stapes is most likely reduced in magnitude by wave interference when the two primary frequencies approach each other. Such a reduction does not take place in the forward-going wave to the location tuned to the DP frequency. This explanation is illustrated on the basis of results of earlier experiments on the movements of the basilar membrane. [DOI: 10.1121/1.1856229]

PACS numbers: 43.64 Kc, 43.64 Bt [BLM] Pages: 1260–1266

I. INTRODUCTION

Before the 1970s, it was assumed that a ''second filter'' was necessary to explain the profound difference between mechanical tuning as it had been found on the basilar membrane (BM) and neural tuning as evidenced by auditory-nerve fibers. For a review see Evans (1975). The experiments of Rhode (1971, 1978) and, later, those of Sellick *et al.* (1982), showed that the sharpness of tuning on the BM is well comparable to neural tuning. These basic findings were corroborated and refined by, amongst others, Robles *et al.* (1986) and Nuttall *et al.* (1990, 1991). Accordingly, it has become a problem for theoreticians to explain the sharp tuning of the BM in terms of a mechanical model of the cochlea. An *ad hoc* assumption was posed by Kim *et al.* (1980), involving the BM to be *locally active*. More precisely, the damping of the BM was assumed to be negative over part of the BM, extending basally from the location of largest response, over which segment the BM was assumed to be capable of amplifying the cochlear wave. That *ad hoc* assumption was supported by an analysis by de Boer (1983), who proved that, in the realm of ''short waves,'' it is impossible for a standard cochlear model to show the type of peak response that had been measured (amplitude and phase) when that model is passive. The model must have a BM impedance, of which the real part is negative basally from the point of largest response. In the same year Davis (1983) coined the term ''cochlear amplifier,'' thereby tacitly implying that there is a physiological mechanism that is responsible for local activity and local amplification. This term has come into common usage, although at this point it should be stressed that the concept of cochlear activity is a theoretical one: certain *mechanical* elements of a cochlear model should be ''active'' for that model to show a theoretical response resembling a measured one. Zweig (1991) showed with the ''inverse solution method'' for a one-dimensional model that ''local activity'' is necessary. We further refer to publications in which the inverse solution method for a three-dimensional model (de Boer, 1995a, b) was used to prove that local activity is absolutely necessary for constructing a three-dimensional cochlear model that simulates measured responses accurately (de Boer and Nuttall, 1999, 2000).

Another very important aspect of all experimental results is that the cochlear mechanical response is fundamentally nonlinear, it is characterized by amplitude compression, shows intercomponent suppression, and includes generation of distortion products for multicomponent stimuli. In addition, the response bandwidth is affected by nonlinearity. Tuning and nonlinearity appear to be intrinsically linked. It follows that cochlear activity and nonlinearity are intrinsically linked, too.

An ingenious physiological experiment, designed to es-

a)Electronic mail: e.d.boer@hccnet.nl

timate the ''gain of the cochlear amplifier,'' was performed by Allen and Fahey (1992). It involved two linked measurements of responses of the functioning cochlea, i.e., recording of otoacoustic emissions (OAEs) in the external ear canal along with simultaneous measurement of the response of a primary auditory neuron. The experiments were performed in anesthetized cats. The acoustical stimulus was a pair of pure tones, with frequencies f_1 and f_2, with $f_2 > f_1$, presented by a transducer in the external ear canal. Where the mechanical responses to these tones interact in the cochlea, distortion products (DPs) are generated. One of these DPs, the one with frequency f_{DP} equal to $f_{DP} = 2f_1 - f_2$, was the object of measurement. In what follows, this component will be designated by ''DP.'' The frequency f_{DP} was chosen equal to the characteristic frequency (CF) of the auditory neuron under study so that the neuron was maximally sensitive to that frequency. The frequencies f_1 and f_2 were varied over as wide a range as possible while the DP frequency f_{DP} was kept constant. The stimulus levels of the two tones were equal and chosen such that at every condition a *constant* response was obtained by the auditory neuron—presumably reflecting constant BM response at the location of that neuron. Meanwhile, the f_{DP} component of the OAE in the external ear canal was measured.

In its most simplified form, the reasoning employed was as follows. When the frequencies f_1 and f_2 are varied, the effective locus of the source of the DP varies along the length of the BM. When f_2/f_1 is close to 1, f_{DP} is nearly equal to f_1 and the DP source is close to the location corresponding to f_{DP}, i.e., close to the location of the neuron measured. When f_2/f_1 is appreciably larger than 1, the DP source is near the peak generated by the f_2 tone. From the source of the DP two waves emerge, one going basally and giving rise to an OAE (pathway #1) and the other apically, towards the location of the neuron (pathway #2). Assume now that there exists a bidirectional cochlear amplifier for the frequency f_{DP}. The spatial region of activity would extend basally from the location of the neuron, over a few mm. The total amplification that a wave undergoes as it follows a path through this spatially distributed amplifier is monotonically related to the length of the path. When the source of the DP is near the location of the neuron (f_2/f_1 close to 1), the wave of pathway #1 traverses a large amount of the path and thus undergoes maximal amplification, and that of pathway #2 none. When the source is far from the neuron (f_2/f_1 appreciably larger than 1) the wave of pathway #1 is not amplified but that of pathway #2 undergoes nearly maximal amplification. Since the response of the neuron for the frequency f_{DP} is kept constant, the amplitude of the OAE in the ear canal should vary as *the square of the amplification factor*.

The experimental results were, in a sense, unexpected. When the ratio f_2/f_1 was varied while constant response of the neuron was monitored, almost no variation was observed in the OAE. Allen and Fahey concluded that the amplification of the cochlear amplifier is less than 10 dB, or, more simply, that ''there exists no cochlear amplifier.'' There is at least one limitation in parameters in the Allen–Fahey experiment. When the ratio f_2/f_1 is close to 1, the three frequencies are close together, and it will not be the f_{DP} component but the (stronger) f_1 component that excites the neuron, making it impossible to monitor the DP component's amplitude at that location. Therefore, the minimum usable value of f_2/f_1 is decidedly different from 1. It was reasoned that this limitation would not affect the conclusion. Because it has been proven impossible to generate a cochlear model without amplification which faithfully simulates cochlear responses as they have been measured (see above), the negative result of the Allen–Fahey experiment forms part of a mystery.

With an entirely different technique, the essential elements of the Allen–Fahey experiment were repeated by Shera and Guinan (1997). This very elegant experiment was also conducted in anesthetized cats, and measurement of OAEs was the exclusive technique employed. The basic stimulus consisted, as before, of two sinusoidal signals, with frequencies f_1 and f_2. The strength of the DP at its best location was monitored via generation of a *secondary DP*. This was accomplished when a third tone was added to the stimulus, with a frequency not far from the DP frequency $f_{DP} = 2f_1 - f_2$. The forward-traveling wave at the DP frequency f_{DP} and the traveling wave of the third stimulus tone gave rise to a secondary DP which could be detected in the external ear canal. The results were essentially similar to those of Allen and Fahey: when the ratio f_2/f_1 approaches 1, there was no sign of cochlear amplification. The mystery remains.

Up to now, several attempts at explaining these—and related—results have been made. These are described further on (Sec. V) where we have occasion to comment upon them. In the interest of brevity we forgo here other, secondary, contributing factors such as suppression of the DP by the primary tones. We will come back to these in Sec. IV.

In the present paper we describe an extension to Allen and Fahey's work, a variation that does not suffer from the principal limitation of that work. Instead of monitoring the response of a neuron ''tuned'' to the DP frequency, it is the velocity of the basilar membrane (BM) that is directly measured and kept constant. More precisely, the component with the DP frequency f_{DP} of the BM velocity is selected and kept at a constant amplitude. In most other aspects, our experiments were of a similar type as Allen and Fahey's. Our results fully corroborate the results of Allen and Fahey and those of Shera and Guinan. What is left to explain is the mystery. It will be argued that a process called ''wave interference'' (Shera, 2003) is likely to be the major cause of the negative outcome of the Allen–Fahey experiment and its variations.

II. EXPERIMENTAL PROCEDURES

We have performed our experiments on guinea pigs. Details about experimental animals and surgical techniques have been described in previous publications (Nuttall *et al.*, 1991, 2004; de Boer and Nuttall, 1997, 2000). The protocols of the experiments described in this paper were approved by the Committee on the Use and Care of Animals, Oregon Health & Science University. Two miniature tweeter loudspeakers were connected with narrow tubes to a plastic cou-

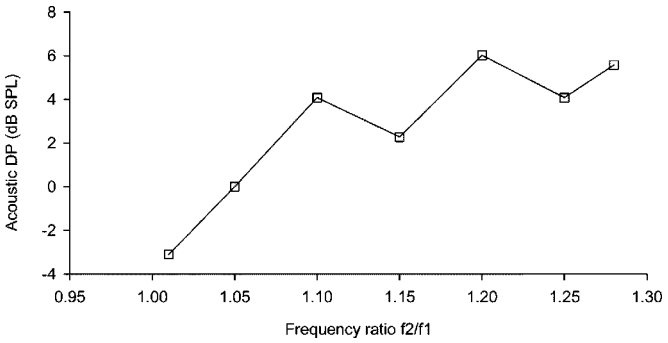


FIG. 1. Experiment #1. DP OAE component in the ear canal versus ratio of primary frequencies. Criterion: DP component of BM velocity at 17-kHz place 8.5 μm/s.

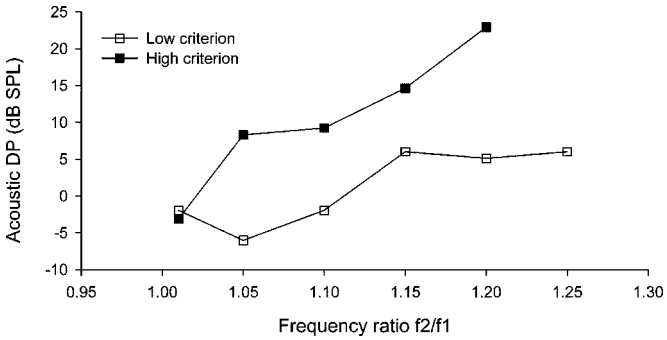


FIG. 2. Experiment #2. DP OAE component in the ear canal versus ratio of primary frequencies. Two criteria are applied to the DP component of BM velocity at 17-kHz place, 21.5 and 62.5 μm/s.

pler placed in the external ear canal. The narrow tubes provided enough damping to ensure a relatively flat frequency response in the region upward from 17 kHz, the frequency region of the primary tones. The same coupler housed an Etymotic 10 B^+ microphone to record the otoacoustic emissions (OAEs). The coupler was not specifically designed to minimize reflection, although the two long and narrow tubes tend to absorb reflected sound.

Basilar-membrane velocity was measured at a location with a best frequency of 16 to 18 kHz. A small opening was made in the first-turn scala tympani bony wall of the cochlea. Glass beads ($\sim$20-μm diameter gold-coated) were placed onto the BM. The laser beam of a Doppler velocimeter (Polytec Corp. OFV 1102) was focused on one of the beads. By measuring compound action potentials (CAPs) we verified that in the region from 15 to 20 kHz the hearing loss due to surgery remained below 10 dB. Pure-tone stimuli were generated by computer; each of the two tones was sent to one of the loudspeakers. All OAEs were recorded with a microphone (Etymotic Research ER-10B^+, Elk Grove Village, IL) from the ear canal. Two real-time FFT analyzers, Stanford Research Systems type SR770 and Hewlett-Packard type 35665A, were used to observe and monitor the signals from the velocimeter and the Etymotic microphone. We ensured that the system's intermodulation distortion was well below the OAE level.

III. EXPERIMENTS AND RESULTS

The principal aim of the experiment was to measure the DP component of the OAE in the ear canal as a function of the frequency ratio f_2/f_1 whereby the amplitude of the DP component of the BM velocity at the measurement site remained constant. In the first experiment the ratio f_2/f_1 was given the values 1.01, 1.05, 1.10, 1.15, 1.20, 1.25, and 1.28. The frequencies f_1 and f_2 of the primary tones were varied whereby the DP frequency f_{DP} was kept constant at 17 kHz, approximately the best frequency at the location where BM velocity was measured. By adjusting the levels of the primary tones the amplitude of the DP component of the BM velocity was kept at the value 8.5 μm/s; for small frequency ratios the level of both primary tones was around 50 dB SPL.

Figure 1 shows the result of this experiment. The figure shows the level of the DPOAE sound pressure in the ear canal for the series of selected values of the ratio f_2/f_1 of the primary tone's frequencies. According to the simplest possible reasoning (see the Introduction), the level of the DPOAE should rise as the ratio f_2/f_1 approaches 1; instead, it decreases. Our experiment does not have a limitation in how near to 1 this ratio can be (within the accuracy of the FFT analyzer, of course), and we observe that the tendency found by Allen and Fahey is confirmed here.

A second experiment was carried out in a different animal. In this case two amplitude criteria were applied to the DP BM velocity component: 21.5 and 62.5 μm/s. For the higher criterion the levels of the primary tones were around 60 dB SPL for small frequency ratios. Results are shown in Fig. 2. The larger ratios could not be used because they would require excessive sound levels. It is clear that the decrease of the DPOAE with decreasing frequency ratio is present in both cases but it is more pronounced for the higher criterion. The tendencies shown in Figs. 1 and 2 were confirmed by two more experiments.

IV. SUPPRESSION

One of the phenomena that could interfere with our simplest interpretation of the procedure is suppression. Two types of suppression should be considered: mutual suppression of the two primary tones and suppression of the DP wave by (one of) the primaries. Kanis and de Boer (1997) indicated that in a very simple nonlinear model of the cochlea mutual suppression of the primaries could give an explanation of the decrease of the DP wave when the two frequencies approach one another. We decided to include mutual suppression in our experiments. Two tones were presented, at the levels used (and varied) in the experiments, and the BM response to each of them was measured with the other tone being switched on and off. This experiment would give an indication of the importance of mutual suppression. Figure 3 shows the result, the percentage of suppression, for both the low and the high criterion, as a function of frequency ratio. Except for the smallest ratio, 1.01, suppression is small to moderate. However, for the ratio 1.01 suppression of the tone with frequency f_1 on that with frequency f_2 is so large that it can at least partially explain the decrease seen in Figs. 1 and 2 in going from ratio 1.05 to 1.01.

To assess suppression of the DP wave by the primary complex is somewhat more complicated. A first-order estimate is obtained by presenting only one tone and measuring

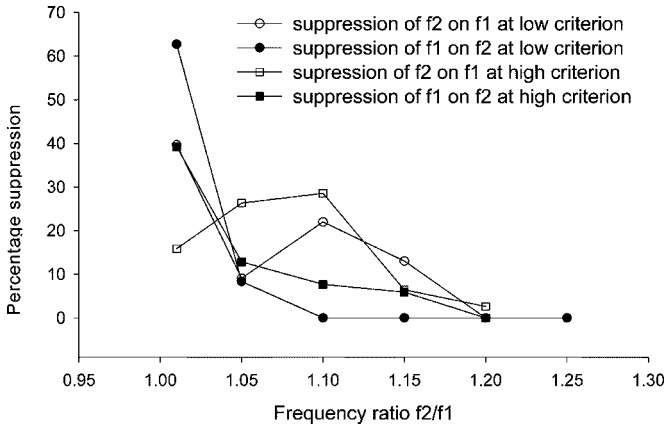


FIG. 3. Mutual suppression of primary components. Levels are varied for different values of the frequency ratio as in the actual experiment #2. Results are shown for both the low and the high criterion for the mechanical DP component.

the suppression of another tone at the DP frequency. Again, the levels of the tone were chosen the same as in the actual experiments. As Fig. 4 shows, suppression of a tone at the DP frequency is substantial, and mainly due to the tone with frequency f_1, the tone that is closest in frequency to the DP. However, somewhat unexpectedly, this suppression does not depend on frequency ratio. A substantial variation of suppression between ratios 1.01 and 1.05, for instance, is not found here (such a variation only occurs when both tones are strong, as Fig. 3 shows). In summary, we do not find a possible cause for the decline of DP acoustic emissions between ratios 1.05 and 1.01 here.

V. POSSIBLE EXPLANATIONS

The most mysterious property of the results described by Allen and Fahey (1992) is that the nearer the frequency ratio f_2/f_1 is to 1, the smaller is the DPOAE amplitude. In this respect we should remember that the DP measured at the level of the BM does not show evidence of a decrease between the ratios 1.1 and 1.01 (Robles *et al.*, 1997). Obviously, there is a fundamental asymmetry in play. Going back to DPOAEs, our results confirm Allen and Fahey's findings in a more general setting, without a limit to how close to 1 the frequency ratio can be. When we concentrate on the region between frequency ratios of 1.01 and 1.1, we may note

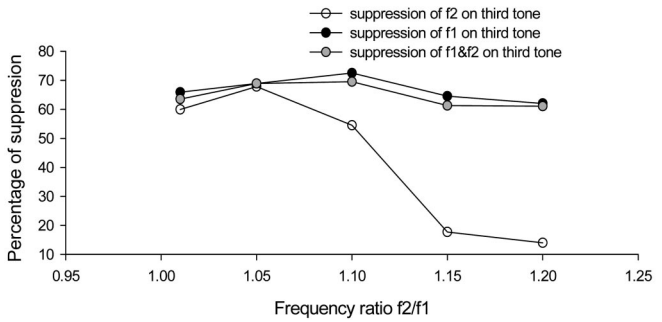


FIG. 4. Suppression of a tone with the frequency of the DP by the primary components (presented separately or together). Levels of the primary components were varied as in experiment #2, high criterion only.

a resemblance with the "tuning" of DPOAEs that has been observed (see, e.g., Brown *et al.*, 1992). In those tuning experiments the DP frequency was not constant but it varied little over this ratio interval. Furthermore, the primary-tone levels were kept constant. Nevertheless, the general tendencies are the same. Several studies have attempted to explain this fundamental finding. For instance, Lukashkin and Russell (2001) described how a no-memory nonlinear device would react to patterns of primary tones with varying amplitudes. For certain amplitude and ratio conditions they found a notch in the amplitude of the DP response and reasoned that this notch and its level dependence could be at the root of the decrease of the DP when f_2/f_1 approaches 1. In addition, they point to mutual suppression of the primary tones as an important contributing factor, a subject we treated in the preceding section. The analysis of Lukashkin and Russell (2001) does not involve the entire region of overlap, and does not include the phase of individual wavelets. A further impediment to considering it in relation to the Allen–Fahey paradigm is that it would apply to both DP waves, of pathway #1 as well as pathway #2. A similar argument can be applied to the idea launched by Allen and Fahey (1993) about a "secondary cochlear map," where it is assumed that DPs are undergoing a "filtering;" there seems to be no reason why the waves in the two directions would be affected differently. In summary, none of the earlier attempts at explaining the variation in the frequency-ratio region between 1.01 and 1.1 would explain why only the basally going wave is affected and not the apically going wave.

Up to now, only one comprehensive attempt at explaining the results of the Allen–Fahey experiment has been made. Shera (2003) computed DP waves generated by a two-tone complex in a cochlear model according to a number of simplifying assumptions. He stressed that the *source* of DPs extends over a certain distance and that this region can encompass an appreciable variation of phase. In particular, when f_2/f_1 approaches 1, a good deal of destructive interference will occur in the wave of pathway #1, whereas no such interference occurs in the wave of pathway #2. Accordingly, the signal in pathway #1 would be reduced in amplitude considerably and cochlear amplification would not be observable at all in the external ear canal. Incidentally, the same mechanism should be responsible for the tuning of DPs. To wit: for ratios close to 1 the interference should decrease the observed OAE amplitude, whereas for ratios above 1.2 the increasing separation between the response peaks corresponding to the frequencies f_1 and f_2 diminishes the overlap between the excitation patterns and thus diminishes the strength of the DP.

Shera's arguments can be extended and applied to realistic responses. Consider the low-level frequency response of the BM as measured in a guinea pig. With the assumption of a cochlea that scales the logarithm of frequency to location along the BM, this response can be transformed to the place domain. The so-derived response is a function of place (let us introduce the variable x) for a particular value of frequency. For the latter frequency we take the best frequency at the location where measurements have been made, in our case 17 kHz. It is now possible to estimate also the response

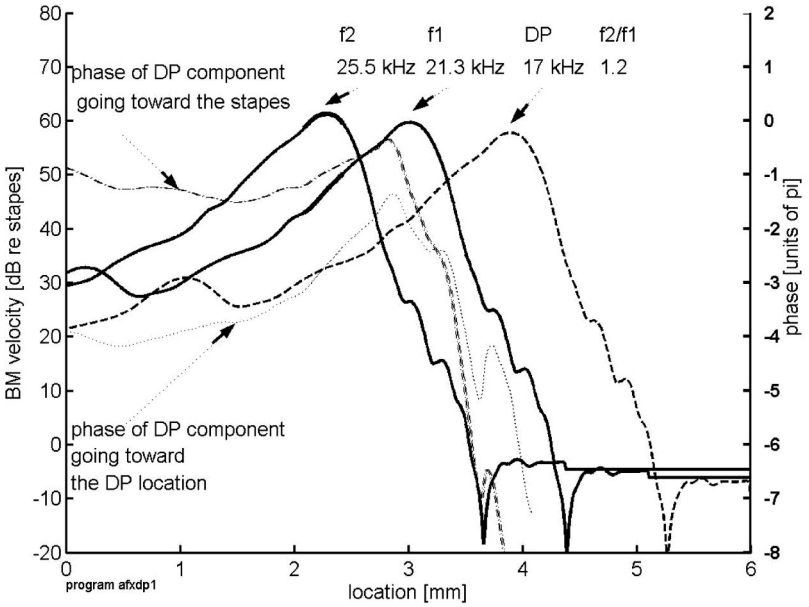


FIG. 5. Cochlear response patterns (in the place domain) produced by two tones with a frequency ratio of 1.2. Data taken from experiment 19 924. Thick solid curves show the amplitude patterns generated by the two primary components. The fat segments of these curves denote the source region of the DP. The thick dashed curve shows the amplitude pattern of a hypothetical stimulus at the DP frequency presented at the stapes. Dash-dot line: phase of DP-frequency wave traveling to the stapes; thin dotted line: phase of DP-frequency wave traveling from source to destination (which is the location of maximal BM response for the DP frequency). In the region of overlap the phase variations are small.

as a function of place at other frequencies. This is done for the frequencies 21.3 and 25.5 kHz, chosen such as to have a ratio of 1.2 and to correspond to a DP of frequency $f_{DP}=2f_1-f_2$, equal to 17.1 kHz. It is stressed that this only involves the transformation from frequency to location (the ''cochlear map'') and not any specific model of cochlear mechanics. The resulting amplitude patterns are shown by the two solid curves in Fig. 5. The thick dashed curve shows the equivalent response of a fictitious tone with the frequency $f_{DP}=2f_1-f_2$. We now can imagine that a DP with this frequency is generated, and certain parts of the two solid curves are made thicker to indicate the x region where the overlap between the response patterns is largest, i.e., the region where the (distributed) source of the DP is assumed to be. From every point in this region two waves emanate, one going towards the stapes and one towards the apex of the cochlea. These waves correspond to the pathways #1 and #2 mentioned earlier. We can estimate the phase course of each of these waves as follows. Let $\varphi_1(x)$ and $\varphi_2(x)$ be the phases of the cochlear patterns of tones 1 and 2, respectively, measured at location x referred to the stapes. Similarly, let $\varphi_{DP}(x)$ be the phase of a (fictitious) tone with the DP frequency, again measured at location x. Then, the phase $\varphi_{st}(x;0)$ given by

$$\varphi_{st}(x;0)=2\varphi_1(x)-\varphi_2(x)+\varphi_{DP}(x)-\varphi_{DP}(0), \tag{1}$$

is the phase of the retrograde wave of pathway #1, originating at location x and measured at location 0 (the stapes). Note that the same phase shift $\varphi_{DP}(x)$ that the DP wave incurs in going from location 0 (the stapes) to location x has to be added for the retrograde wave along pathway #1. In Fig. 5 the dash-dot line shows the phase of the wave of pathway #1 as a function of location x. In contrast, the phase $\varphi_{ap}(x)$ given by

$$\varphi_{ap}(x)=2\varphi_1(x)-\varphi_2(x)-\varphi_{DP}(x)+\varphi_{DP}(x_{DP}), \tag{2}$$

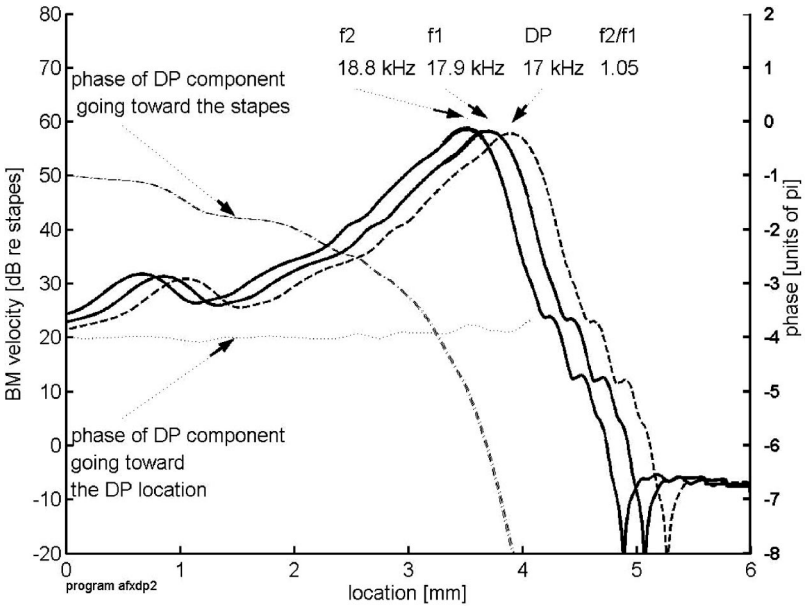


FIG. 6. Cochlear response patterns of two tones with a frequency ratio of 1.05. Dash-dot line: phase of DP-frequency wave traveling to the stapes; thin dotted line: phase of DP-frequency wave traveling from source to destination, the location of maximal BM response for the DP frequency. In this case the wave traveling towards the stapes is much attenuated because the DP-wave phase is varying very fast in the region of overlap and wave interference is pronounced.

is the phase of the apically going wave of pathway #2, which originates at location x and is measured at location x_{DP}, the location of the peak for the frequency f_{DP}. The finely dotted line in Fig. 5 shows the phase $\varphi_{ap}(x)$ of the wave toward the apex (pathway #2), again as a function of x. For each of the two functions shown the zero point is not important; what should interest us is only the *variation* of the phase. The wave of pathway #1 does not show large variations in the overlap region, which means that DP wavelets created there will arrive at the stapes with almost the same phase and will mainly reinforce one another. The phase variations in pathway #2 are slightly larger in the overlap region but the total variation does not exceed π radians (see the phase scale on the right of the figure). For the wave traveling in this direction reinforcement is slightly smaller.

In Fig. 6 we show the condition where the frequency ratio is 1.05. The phase variations of pathway #2 are minimal but those of pathway #1 are much larger than in Fig. 5. This means that in this case the DP wavelets in pathway #1 emanating from the overlap region have a pronounced tendency to cancel; as a result the OAE will be much reduced in amplitude. The other wave is hardly diminished in amplitude. The effect shown by Figs. 5 and 6 (derived from realistic responses) lies at the root of the computations carried out by Shera (2003) and substantiate his theoretical analysis but are more general. At this point it is not possible to go into any deeper detail, for instance, computing the integrated contribution of all wavelets to the total waves, because we would need more refined assumptions and a more detailed three-dimensional model of the excitation of the outer hair cells to do that. This extension falls outside the scope of the present paper.

VI. DISCUSSION

From the experiments described we can draw the conclusion that, in contrast to expectation, the DPOAE does not show evidence of cochlear amplification when the frequency ratio f_2/f_1 approaches 1. On the contrary, the DPOAE level decreases in this case. In this respect we corroborate the experimental results of Allen and Fahey (1992). Our measurement method differed in one respect from that of Allen and Fahey: we did not specifically try to minimize reflection at the stapes. We consider it unlikely that in the frequency ratio range from 1.01 to 1.1 variations of reflectivity could be so large as to explain the course of the results.

The most likely reason for the decrease of the DPOAE when f_2/f_1 approaches 1 lies in the wave-interference effect emphasized by Shera (2003), which we have illustrated in the preceding section. One important question remains. According to the simplest theory, the ratio of the DPOAE to the BM velocity at the DP frequency should vary as the square of the cochlear amplification factor (see the Introduction and the Allen–Fahey paper) when f_2/f_1 is varied over a large range. Would the interference effect be large enough to compensate that factor? In this respect we have to consider the nature of cochlear amplification. In many papers ''amplification'' is taken as the ratio of the response at the peak measured in the live animal to the response at the same frequency in the dead animal. This value may be up to 80 dB. In other work it is taken as the ratio of the peak responses in these two cases; values of up to 50 dB have been quoted. In the physical sense, neither value is correct because it is not the power which is compared. de Boer and Nuttall (2001) published a report in which the actual power amplification was computed in a (three-dimensional) model of the cochlea designed to simulate measured responses. Note that in this work the term ''power amplification'' refers to the actual amplification of the acoustical power that is flowing in the fluid of the model, from the stapes to the destination. The result was that the power amplification was maximally 17.7 dB. It is this power amplification that has to be ''camouflaged'' by wave interference, and it is not difficult to see that this would be possible.

ACKNOWLEDGMENTS

The authors are grateful to Christopher A. Shera for his contributions to our understanding, and to Paul F. Fahey for pointing out and discussing omissions in our manuscript. This study has been supported by NIH NIDCD DC 00141.

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**, 809–816.

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.

Davis, H. (**1983**). ''An active process in cochlear mechanics,'' Hear. Res. **9**, 79–90.

de Boer, E. (**1983**). ''No sharpening? A challenge for cochlear mechanics,'' J. Acoust. Soc. Am. **73**, 567–573.

de Boer, E. (**1995a**). ''The inverse problem solved for a three-dimensional model of the cochlea. I. Analysis,'' J. Acoust. Soc. Am. **98**, 96–903.

de Boer, E. (**1995b**). ''The inverse problem solved for a three-dimensional model of the cochlea. II. Application to experimental data sets,'' J. Acoust. Soc. Am. **98**, 904–910.

de Boer, E., and Nuttall, A. L. (**1999**). ''The inverse problem solved for a three-dimensional model of the cochlea. III. Brushing up the solution method,'' J. Acoust. Soc. Am. **105**, 3410–3420.

de Boer, E., and Nuttall, A. L. (**1997**). ''The mechanical waveform of the basilar membrane. I. Frequency modulations (glides) in impulse responses and cross-correlation functions,'' J. Acoust. Soc. Am. **101**, 3583–3592.

de Boer, E., and Nuttall, A. L. (**2000**). ''The mechanical waveform of the basilar membrane. III. Intensity effects,'' J. Acoust. Soc. Am. **107**, 1497–1507.

de Boer, E., and Nuttall, A. L. (**2001**). ''Power gain of the Cochlear Amplifier,'' in *Physiological and Psychological Bases of Auditory Function*, edited by D. J. Breebaart, A. J. M. Houtsma, A. Kohlrausch, V. F. Prijs, and R. Schoonhoven (Shaker, Maastricht), pp. 1–7.

Evans, E. F. (**1975**). ''The sharpening of cochlear frequency selectivity in the normal and abnormal cochlea,'' Audiology **14**, 419–442.

Kanis, L. J., and de Boer, E. (**1997**). ''Frequency dependence of acoustic distortion products in a locally active model of the cochlea,'' J. Acoust. Soc. Am. **101**, 1527–1531.

Kim, D. O., Neely, S. T., Molnar, C. E., and Matthews, J. W. (**1980**). ''An active cochlear model with negative damping in the partition: Comparison with Rhode's ante- and post-mortem observations,'' in *Psychophysical, Physiological and Behavioural Studies in Hearing*, edited by. G. v. d. Brink and F. A. Bilsen (Delft University Press, Delft), pp. 7–14.

Lukashkin, A. N., and Russell, I. J. (**2001**). ''Origin of the bell-like dependence of the DPOAE amplitude on primary frequency ratio,'' J. Acoust. Soc. Am. **110**, 3097–3106.

Nuttall, A. L., Dolan, D. F., and Avinash, G. (**1990**). ''Measurements of basilar membrane tuning and distortion with laser Doppler velocimetry,''

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, Berlin), pp. 288–295.
Nuttall, A. L., Dolan, D. F., and Avinash, G. (**1991**). ''Laser Doppler velocimetry of basilar membrane vibration,'' Hear. Res. **51**, 203–214.
Nuttall, A. L., Grosh, K., Zheng, J., de Boer, E., Zou, Y., and Ren, T. (**2004**). ''Spontaneous basilar membrane oscillation and otoacoustic emission at 15 kHz in a guinea pig,'' J. Assoc. Res. Otolaryngol. **5**, 337–349.
Rhode, W. S. (**1971**). ''Observations of the vibration of the basilar membrane in squirrel monkeys using the Mössbauer technique,'' J. Acoust. Soc. Am. **49**, 1218–1231.
Robles, L., Ruggero, M. A., and Rich, N. C. (**1986**). ''Basilar membrane mechanics at the base of the chinchilla cochlea. I. Input–output functions, tuning curves, and response phases,'' J. Acoust. Soc. Am. **80**, 1364–1374.
Robles, L., Ruggero, M. A., and Rich, N. C. (**1997**). Two-tone distortion on the basilar membrane of the chinchilla. J. Neurophysiol. **77**, 2385–2399.
Rhode, W. S. (**1978**). ''Some observations on cochlear mechanics,'' J. Acoust. Soc. Am. **64**, 158–176.
Sellick, P. M., Patuzzi, R., and Johnstone, B. M. (**1982**). ''Measurement of basilar membrane motion in the guinea pig using the Mössbauer technique,'' J. Acoust. Soc. Am. **72**, 131–141.
Shera, C. A. (**2003**). ''Wave interference in the generation of reflection- and distortion-source emissions,'' in *Biophysics of the Cochlea: From Molecules to Models*, edited by A. W. Gummer, E. Dalhoff, M. Nowotny, and M. P. Scherer (World Scientific, Singapore), pp. 439–454.
Shera, C. A., and Guinan, J. J. (**1997**). ''Measuring cochlear amplification and nonlinearity using distortion-product otoacoustic emissions as a calibrated intracochlear sound source,'' Abstracts of the Midwinter Meeting of the Association for Research in Otolaryngology 20, 13.
Zweig, G. (**1991**). ''Finding the impedance of the organ of Corti,'' J. Acoust. Soc. Am. **89**, 1229–1254.

Recording depth and signal competition in heterodyne interferometry

Ombeline de La Rochefoucauld,[a] Shyam M. Khanna,[b] and Elizabeth S. Olson[c]
Fowler Memorial Laboratory, Department of Otolaryngology, Head and Neck Surgery, College of Physicians and Surgeons of Columbia University, New York, New York 10032

(Received 14 June 2004; revised 11 November 2004; accepted 22 November 2004)

A common way to measure submicroscopic motion of the organ of Corti is heterodyne interferometry. The depth over which vibration can be accurately measured with heterodyne interferometry is determined by both the optics, which controls to what extent light from nonfocal planes reaches the photodetectors, and demodulation electronics, which determines to what extent signal generated by out-of-focal-plane light influences the measurements. The influence of a second reflecting surface is investigated theoretically and experimentally. By reviewing the theory of FM demodulation and showing tests with a Revox FM demodulator, it is demonstrated that the influence of a secondary signal on a measurement depends on the modulation index. Both high- and low-modulation index signals are encountered in heterodyne interferometry of the cochlea. Using a He–Ne-like diode laser ($\lambda = 638$ nm), the border between low- and high-modulation signals is at a displacement of about 25–100 nm. Confocal interferometry reduces the magnitude of out-of-focus signals, and therefore their effect on vibration measurement. The response of the confocal system to reflected signals from two surfaces separated by distances encountered within the cochlear partition is shown. The results underscore the benefit of steep optical sectioning for intracochlear measurements. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1848177]

PACS numbers: 43.64.Kc, 43.64.Yp Pages: 1267–1284

LIST OF SYMBOLS

f_o frequency of the object beam
f_r frequency of the reference beam
f_c carrier frequency
f_m modulation frequency
Δf frequency deviation
ω radial frequency $= 2\pi f$
λ wavelength of laser light
$V_A(t)$, $V_C(t)$ velocity of surface **A** or **C**
$x_A(t)$, $x_C(t)$ FM signal out of photodetector due to surface **A** or **C**
$\beta = 2\pi\Delta f/\omega_m$ modulation index,
V_o = velocity amplitude,
X_o = displacement amplitude
$\Delta f = 2V_o/\lambda = 2\omega_m X_o/\lambda$ frequency deviation in terms of object motion, due to Doppler shift
ψ_0 relative phase between FM signals $x_A(t)$, $x_C(t)$
a, c amplitudes of the FM signals $x_A(t)$, $x_C(t)$
ρ $= c/a$
$E(t)$ theoretical demodulator output signal
$E_o(t)$ experimental demodulator output signal from optical experiment
$E_g(t)$ experimental demodulator output signal from experiment with signal generators
R reflectivity
m relative interference amplitude (wavefront distortion)
k value the carrier power takes at a given distance from the plane of focus (from optical sectioning curve)
d distance between the two surfaces in the optical experiments
z distance from the focal plane

I. INTRODUCTION

Heterodyne interferometry has been the primary technique for measuring cochlear motion for some time (Willemin *et al.*, 1988, Nuttall *et al.*, 1991, Ruggero and Rich, 1991, Cooper, 1999a). The technique uses interference between two laser beams (object and reference) whose frequencies have been shifted from each other by an amount that can be resolved by standard electronics (455 kHz in our system). The velocity of the test object shifts the frequency of the object beam due to the Doppler effect and is encoded in the frequency modulation of the 455-kHz heterodyne ''carrier'' signal. The heterodyne signal is decoded using a frequency demodulator. (Equivalently, the displacement of the test object changes the distance through which the object beam travels, thus modulating the phase of the object beam. Displacement is decoded with a phase demodulator.) The cochlear tissues are of low reflectivity and, in order to in

[a]Author to whom correspondence should be addressed. Electronic mail: or2107@columbia.edu
[b]Electronic mail: smk3@columbia.edu
[c]Electronic mail: eao2004@columbia.edu

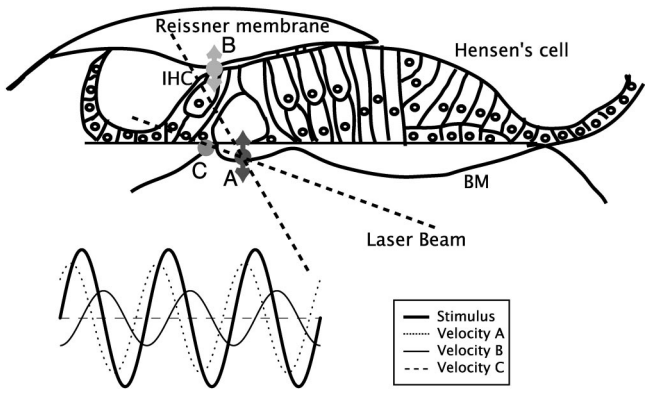


FIG. 1. The organ of Corti consists of several cellular and acellular layers. The measuring laser beam of the interferometer is therefore reflected from several surfaces. When sound is applied to the cochlea, these surfaces (for example, **A**: basilar membrane (BM); **B**: inner hair cell (IHC); **C**: lamina) vibrate with different amplitudes and phases. Individual reflections from these surfaces produce carrier signals of different amplitudes, phases, and Doppler shifts at the photodetector. The relative magnitude of the optical signals reaching the detector depends on the plane of focus of the interferometer, reflectivity and orientation of the surface, and on the optical sectioning characteristics of the interferometer. The frequency demodulator determines how much the out-of-plane reflections affect the vibration being measured.

crease the power in the object beam, reflecting beads are often placed on the sensory tissue and the motion of the beads measured. A bead is a retro-reflector and the light reflected from it comes from a well-defined depth. However, measurements with beads also have disadvantages: The position of measurement is restricted to that of the bead, and holes must be made in the overlying tissue in order to place the bead. The bead might or might not move with the underlying tissue (Khanna *et al.*, 1998, Cooper, 1999b). On the other hand, for measurements made without a bead, the laser beam goes through the tissue and means have to be provided in order to select the depth from which the reflection is accepted by the interferometer. For example, in Fig. 1 different parts of the sensory tissue are imagined to move with different amplitudes and phases. We would like to know under what conditions reflections from out-of-focus surfaces affect the measured velocity of the in-focus surface.

The accuracy with which vibration at a selected plane can be measured in the presence of reflections from secondary surfaces is determined by the optical sectioning characteristics of the interferometer and the signal processing of the FM demodulator. By reviewing the theory of FM demodulation and describing tests with our FM demodulator, we show that the influence of a secondary signal is qualitatively different depending on whether the modulation index is high (greater than 2) or low (less than 0.5). The modulation index, β, is the ratio of the frequency deviation to the modulation (stimulus) frequency ($\beta = \Delta f / f_m$). For perpendicular incidence, the frequency deviation is related to the magnitude of the velocity via a Doppler shift: $\Delta f = 2V_o/\lambda$, where $\lambda = 638$ nm for our laser. When the response to a pure-tone stimulus is linear, $V_o = \omega X_o$, where X_o = displacement and $\omega = 2\pi$ multiplied by the stimulus frequency, and thus $\beta = 4\pi X_o/\lambda$. Therefore, high modulation index corresponds to displacements greater than 100 nm, low modulation index corresponds to displacements less than 25 nm. Both high- and low-modulation-index signals occur in heterodyne interferometry of the cochlea. For high-modulation-index signals, the fundamental component of the output of the FM demodulator is not affected by a secondary signal unless the secondary signal's power is nearly as large as that of the primary signal. However, the output waveform can be distorted. For a low-modulation-index signal, a secondary competing signal can have a relatively large effect on the fundamental component of the output signal, but the output signal waveform is not distorted. Cells and structures in the cochlea have a wide range of reflectivities, and steep optical sectioning is necessary to reduce contamination by out-of-focus signals.

Detailed discussion on the spatial resolution of intracochlear velocity measurements appeared in two 2001 letters to the *Journal of the Acoustical Society*. These are Ren and Nuttall, 2001, ''Recording depth of the heterodyne laser interferometer for cochlear vibration measurement,'' and Dalhoff *et al.*, 2001, ''Remarks about the depth resolution of heterodyne interferometers in cochlear investigations.'' Our contribution concerns optical sectioning and FM demodulation, and continues the discussion of those letters.

In the article by Ren and Nuttall, a heterodyne interferometer was used to measure the velocity of a vibrating mirror. The optical sectioning curve shows the decrease of the power in the carrier signal as the distance between a reflective plane and the focal plane (the defocus distance) is increased. The FWHM is defined as the full width of the carrier power curve at half maximum. The recording depth can be defined as the defocus distance at which the reported velocity of the reflective plane decreases to 50%. Ren and Nuttall reported that the carrier power decreased to 50% of its maximum at a distance of $\sim \pm 9.5$ μm from the focal plane (FWHM $= 19.5$ μm), and to approximately 10% at a distance of ± 25 μm. At larger distances there was no further reduction of carrier power. In contrast, the measured velocity did not change with defocusing distances -40 to $+25$ μm. At distances of -60, $+35$ μm the velocity had dropped to 50% of its maximum. Because this total distance (95 μm) was much larger than the FWHM of 19.5 μm for the carrier, the authors noted that the FWHM for the carrier power could not be used to determine the FWHM for the velocity. Below, we further explore the relationship between depth resolution and optical sectioning.

The letter by Dalhoff *et al.* discussed the effect of a signal from a secondary reflector. As the letter states, the light waves from the primary and secondary reflectors will add at the input to the photodetector. In this contribution we extend the discussion of Dalhoff *et al.* by considering the subsequent processing of the summed signal. When reflections from two objects contribute to the light input to the photodetector, the output signal from the photodetector at the heterodyne frequency that feeds the FM demodulator is the sum of two signals, each with amplitude proportional to the square root of the light power from one of the reflectors. Each signal is frequency modulated by the Doppler shift that is proportional to the velocity of its reflector. The purpose of FM demodulation is to extract the instantaneous frequency of the frequency-modulated signal. Therefore, to analyze the

summed signal analytically, the instantaneous frequency of the summed signal is the quantity of interest. The analytic section of this paper shows the instantaneous frequency under various conditions and leads to the low-modulation-index, high-modulation-index classification of results. Tests with our FM demodulator with input signals produced with two signal generators are shown, and confirm the analytic results. Finally, we show the response of our integrated optical and demodulator system when the two competing signals are produced optically, by reflecting surfaces with realistic cochlea-like separations and motions. First we review the optical sectioning of the split-aperture system developed by Khanna *et al.* (1996), as the conclusions of the demodulation analysis and tests will point to the importance of steep optical sectioning for reducing the contamination from secondary reflectors.

II. MEASUREMENTS AND DISCUSSION

A. Optical sectioning

The optical sectioning capability of an optical system is its ability to select light from one plane in the illuminated sample and reject light from out-of-focus planes. The shape of the optical sectioning curve is governed by several factors: (I) The objective lens's numerical aperture (N.A.) determines the shape of the optical sectioning curve near the plane of focus. (II) Light returning from out-of-focus planes can be reduced by directly blocking it. This is how a slit confocal microscope works (Koester *et al.*, 1994). (III) A low coherence source (such as a low coherence diode laser or a superluminescent diode) can be used to reduce the contribution of light returning from out-of-focus planes. This strategy was described by Dalhoff *et al.* (2001).

The approximate analytic expression for the optical sectioning curve due to the objective lens's N. A. is

$$V(z)=\{\sin(Z)/Z\}^2, \quad Z=\pi(\text{N.A.})^2 z/n\lambda, \tag{1}$$

where $V(z)$ is proportional to the light power as a function of Z, n is the refractive index of the medium, N.A. is the numerical aperture of the lens, λ is the wavelength, and z is the distance from the focal plane. [Dalhoff *et al.* discuss the limitations of Eq. (1) and reference Wilson, 1990.] [As a clarifying note: When the object beam interferes with a reference beam in a heterodyne interferometer, the voltage out of the photodetector at the heterodyne frequency—the heterodyne signal—is proportional to the square root of Eq. (1), as only the object beam is passed through the objective lens. The power in the heterodyne signal, the ''carrier power,'' which is proportional to voltage squared, is proportional to Eq. (1).] In Fig. 2, we compare the theoretical objective lens optical sectioning dictated by Eq. (1) with the optical sectioning realized by the Koester/Khanna interferometer with split-aperture objective.

The split aperture is made by placing an opaque vertical strip close to the back aperture of the objective lens (Koester *et al.*, 1994). The optical system is arranged so that the illumination beam uses one half of the lens and the reflected light uses the other half. Figure 2 shows that without the split aperture, the theoretical optical sectioning curve initially

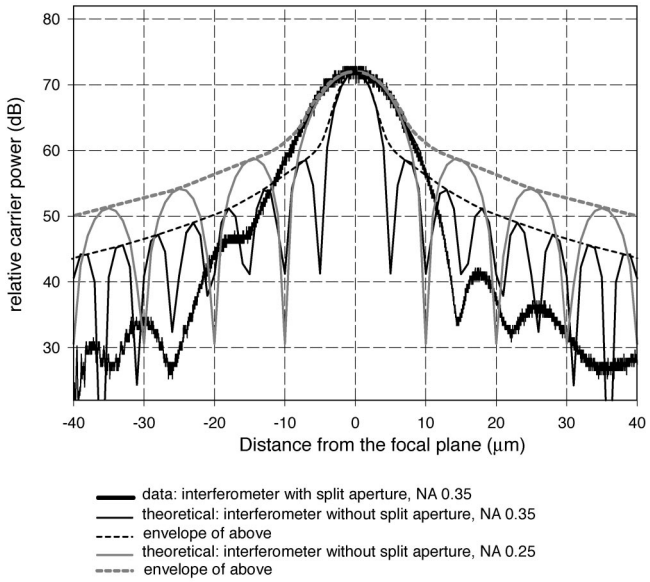


FIG. 2. Optical sectioning curve as a function of the distance of the reflector from the focal plane. The optical sectioning curve was measured by the carrier power in the heterodyne interferometer signal. We show: (i) Measured optical sectioning curve with split-aperture objective (NA=0.35, thick black line). (ii) Calculated optical sectioning curve due to objective lens without split aperture (NA=0.35, solid black line). (iii) Corresponding envelope (dashed black line). (iv) Calculated optical sectioning curve due to objective lens without split aperture (NA=0.25, solid gray line). (v) Corresponding envelope (dashed gray line). Near the focal plane ($d \leqslant \pm 10\ \mu$m) the optical sectioning curve of the split-aperture lens (NA=0.35) matches the theoretical curve for NA=0.25. This is because the split aperture reduces the effective aperture of the lens. However, away from the focal plane ($d \geqslant \pm 10\ \mu$m) the rejection of the split-aperture system is greater than that of the other systems.

falls off smoothly with distance but then rebounds due to diffraction sidelobes. It is the envelope of the curve that is relevant in practice, and therefore the envelopes of the theoretical curves have been included in our Fig. 2. The opaque strip in the split-aperture system provides steep fall-off away from the focal plane. The experimental curve shown in Fig. 2 does not represent a theoretical limit, and improvements of the split-aperture system are expected to extend the steep fall-off even further. A disadvantage of the split aperture is that the effective numerical aperture of the lens is reduced in one dimension. Thus, in the region of the peak (where the N.A. governs the curve shape) the data are very close to the theoretical results for a lens with N.A. 0.25, although the N.A. of the lens when fully illuminated is 0.35. The split-aperture arrangement offers substantial and increasing improvement in sectioning when the defocusing distance is greater than 10–15 μm. Table I compares the optical sectioning properties obtained with the split-aperture arrangement of Khanna *et al.* with that of several other interferometers.

B. FM interferometry

1. The demodulation process

a. Theory. In heterodyne interferometry a laser beam is divided into two beams: object beam (frequency f_o) and reference beam (frequency f_r). One or both beams are frequency shifted so that the frequency difference ($f_o - f_r$) between the two beams is f_c (carrier frequency). In our case

TABLE I. First column: (i) Paper from which the data are quoted. (ii) Magnification and numerical aperture (NA) of the objective lens used. (iii) Full width at half maximum (FWHM), the distance in μm between the two points where the carrier level drops to half power. Columns 2–5: distances from the focal plane in μm where the carrier power drops by factors of 10, 100, 1000, and 10 000, respectively. Symbol $\cdots$ indicates that the level was not reported.

Carrier power/maximum light rejection	10^{-1}	10^{-2}	10^{-3}	10^{-4}
Ren & Nuttall, 2001 20 X, NA 0.42 FWHM=19.5 μm	± 25	$\cdots$	$\cdots$	$\cdots$
Dalhoff *et al.*, 2001 (using the short coherence length laser diode) NA 0.14 FWHM=27 μm	$-28, +26$	$-44, +41$	$-60, +51$	$-75, +63$
Cooper, 1999a 5 X, NA 0.13 FWHM=75 μm (with respect to carrier power)	$-82, +65$	$-146, +156$	$\cdots$	$\cdots$
10 X, NA 0.25 FWHM=16 μm (with respect to carrier power)	$-16, +16$	$-40, +35$	$-83, +97$	$-272, +265$
Khanna *et al.*, 1996 20 X, NA 0.53 FWHM=4.2 μm	$-5, +3$	$-11, +4$	$-16, +10$	$-40, +20$
Figure 2 data above 20 X, NA 0.35 Nikon lens in air FWHM=9 μm	$-9, +8$	$-13, +12$	$-21, +13$	$-33, +30$

$f_o - f_r = 455$ kHz. The object beam is focused on the surface to be measured (for example, **A** in Fig. 1). The object beam reflects from the surface and is combined with the reference beam. The photodetector measures light power—the sum of the object and reference waves, quantity squared. When the sum is squared a term at the difference frequency, $f_o - f_r$, is produced (along with terms at the frequencies $2f_o$, $2f_r$ and $f_o + f_r$). f_r and f_o are too high in frequency to be resolved by the photodetector, and the photodetector's ac voltage output is at the difference frequency ($f_c = f_o - f_r$). The amplitude of this ac signal is proportional to the product of the reference and object wave amplitudes, or equivalently, to the product of the square root of the reference and object beam powers. (See Willemin *et al.*, 1988 and Cooper, 1999a for more details.) When the object (**A**) is moving, f_o is changing in time due to the Doppler shift, and $f_o(t) - f_r = f_c + f(t)$, resulting in the frequency-modulated (FM) signal $x_A(t)$

$$x_A(t) = a\cos\left(2\pi\int (f_o(t') - f_r)dt'\right)$$
$$= a\cos\left(\omega_c t + 2\pi\int f(t')dt'\right), \quad (2)$$

where $x_A(t)$ is the voltage input to an FM demodulator tuned to 455 kHz. Demodulation is accomplished by detecting the phase of the signal, $(\omega_c t + 2\pi\int f(t')dt')$, taking the time derivative to find the instantaneous frequency and subtracting the carrier frequency. For an unchallenged signal, this results in a perfect extraction of $f(t)$. Further along we will consider the effect of applying this simple algorithm to a signal that is challenged by the signal from a secondary object.

When the modulating signal is sinusoidal, $f(t) = \Delta f \sin(\omega_m t + \phi_A)$, the FM signal can be written as

$$x_A(t) = a\cos(\omega_c t - \beta\cos(\omega_m t + \phi_A))$$
$$\text{with} \quad \beta = \frac{2\pi\Delta f}{\omega_m}, \quad (3)$$

where f_m is the modulation frequency and corresponds to the frequency of the object's motion ($\omega = 2\pi f$). ϕ_A is the phase of the object's movement. Δf is the frequency deviation and is directly related to the velocity amplitude, V_o, as $\Delta f = 2V_o/\lambda$, where λ is the laser wavelength. This is the Doppler shift. β, the modulation index, can be expressed as a function of velocity amplitude ($\beta = 2V_o/\lambda f_m$) or displacement amplitude ($\beta = 4\pi X_o/\lambda$).

The spectrum of x_A can be expressed in terms of Bessel's functions $J_n(\beta)$. It contains a carrier component and an infinite set of sidebands located symmetrically on either side of the carrier ($f_c \pm nf_m$). For small modulation index, only the Bessel's coefficients $J_0(\beta)$ and $J_1(\beta)$ have significant values, so that the FM signal is effectively composed of a carrier and a single pair of sidebands at $f_c \pm f_m$.

First, consider the velocity of a single surface. At what defocus value would the demodulator fail to accurately report the velocity? The ability to measure vibration depends on the carrier to noise ratio (C/N) at the input of the demodulator. In the most sensitive detection method, the noise level at the photodetector output is determined by the shot noise N, which is directly proportional to the reference beam power. (The reference beam is made powerful enough so that this condition obtains.) For the FM demodulator to function the C/N must exceed 10 dB with the full 150-kHz bandwidth of the demodulator (Willemin *et al.*, 1988). This is known as the ''threshold effect'' in the FM literature (Panter, 1965). As the surface is moved away from the focal plane the light in the object beam will decrease, and so the carrier power will

decrease while the shot noise will remain constant. As long as the carrier power remains at least 10 dB above the noise power, the demodulator will continue to measure the vibration of the object. For example, when measuring from a naturally bright reflector such as guinea pig Hensen's cell, the maximum carrier power in our system is -10 to -20 dB. Shot noise is -68 dB. With a 0.35 numerical aperture objective lens, the optical sectioning of the split-aperture system (Fig. 2) will decrease the carrier signal to -58 dB at a distance of $\sim +/-25$ μm from the focal plane. Because the FWHM of the optical sectioning curve with this lens is only ~ 9 μm, the recording depth—the defocus value for which the velocity is accurately reported—can be broader than the FWHM. This is the same effect that was discussed by Ren and Nuttall (2001). The defocus distance over which the vibration can be measured depends on the interferometer optical sectioning, the incident light power, and the reflectivity of the surface. The steeper the optical sectioning, the narrower will be the allowed defocus distance. Lower incident light and lower reflectivity also narrow the allowed defocus distance. In summary, the relatively broad allowed defocusing distance can be understood in terms of the FM demodulator's threshold C/N ratio.

The relatively large defocus distance over which velocity can be measured accurately will not influence the result when measuring the motion of a single surface. However, the observation does raise questions about the effect a secondary surface, which is within the allowed defocus distance of a primary surface, would have on a measurement of the primary surface's motion. In brief, when measuring through multiple surfaces, it is important that the interferometer measures the vibration of the surface that is in focus. This depends both on the optical sectioning and how the demodulator responds to competing signals. The optical sectioning determines how much the competing signal is reduced, and the demodulator determines how much the residual competing signal affects the measurements.

Therefore, we take up the question of competing signals. Returning to Fig. 1, assume that point **A** moves with velocity $V_A(t)$ upon sound stimulation, and point **C** on the lamina moves with velocity $V_C(t)$. If **A** is at the focal plane, light reaching the detector from surface **C** will be attenuated according to the optical sectioning characteristics of the interferometer. How will the signal from **C** influence the reported velocity? The photodetector output will contain an ac component from the interference of the **A** wave and the reference wave: $x_A(t)=a\cos(\omega_{cA}t-\beta_A\cos(\omega_m t))$. The photodetector output will also contain an ac component from the interference of the **C** wave and the reference wave: $x_C(t)=c\cos(\omega_{cC}t-\beta_C\cos(\omega_m t+\phi_C)+\psi_0)$. The amplitudes c and a are proportional to the square root of the light power in the **C** and **A** waves. ϕ_C represents the phase of the movement at **C** relative to at **A**. ψ_0 is the relative phase between the **A** and **C** light waves, due to the optical path length difference. ψ_0 depends on the distance at rest between the two surfaces. In order to reduce the number of parameters we will only treat the case $\phi_C=0$. $\omega_{cC}=\omega_{cA}=\omega_c$, but we write them with different symbols to discuss signal competition more generally. For example, ω_{cC} not equal but close to ω_{cA} is the situation with two competing FM radio signals. There will be a third ac component due to interference of the **A** and **C** waves with each other, but its frequency will be far from the carrier frequency the FM demodulator is tuned to and will not influence the demodulation. Therefore, the relevant ac voltage at the demodulator will be $x(t)=x_A(t)+x_C(t)$. The mathematics involved in adding the two signals is just like that used to discuss "beats" in the general case in which the two signals that make up the beat are not of equal amplitude. Using phasor algebra, $x(t)$ is written as

$$x(t)=a[\sqrt{1+\rho^2+2\rho\cos(\varphi)}]\cos(\omega_{cA}t-\beta_A\cos(\omega_m t)+\theta),$$

$$\text{with}\quad \rho=\frac{c}{a};\quad \varphi=(\omega_{cC}-\omega_{cA})t+\psi_0-(\beta_C-\beta_A)\cos(\omega_m t),$$

$$\text{and}\quad \tan(\theta)=\frac{\rho\sin(\varphi)}{1+\rho\cos(\varphi)}. \tag{4}$$

The amplitude of the signal $x(t)$, $a\sqrt{1+\rho^2+2\rho\cos(\varphi)}$, is modulated. To a first approximation, this will not affect the output of the FM receiver. Indeed, during the demodulation process, the signal $x(t)$ is clipped in order to remove any amplitude modulation. This clipping is performed by the "limiter."

The demodulation process consists of the extraction of the instantaneous frequency, $\omega_i(t)\equiv d\psi/dt$, from the phase, $\psi(t)=(\omega_{cA}t-\beta_A\cos(\omega_m t)+\theta)$. Finally, the carrier frequency is subtracted to give the demodulated output signal, $E(t)$

$$E(t)=\omega_i(t)-\omega_{cA}=2\pi\Delta f_A\sin(\omega_m t)+\frac{d\theta}{dt},$$

$$\text{with}\quad \theta=\arctan\left(\frac{\rho\sin(\varphi)}{1+\rho\cos(\varphi)}\right),\quad \rho=\frac{c}{a},$$

$$\varphi=(\omega_{cC}-\omega_{cA})t+\psi_0-(\beta_C-\beta_A)\cos(\omega_m t). \tag{5}$$

[A generalization of Eq. (5) for which $\phi_C\neq 0$ is included in a footnote.[1]] Equation (5) is useful to study the theoretical influence of a competing signal $x_C(t)$ on the output of the FM receiver. The first term gives the output in the absence of the competing signal. The second term shows the effect of competing signals. This "error" in the instantaneous frequency depends on the modulation indices β_C, β_A, on the ratio ρ of the amplitudes of the secondary and primary signals, on their relative phase ψ_0, and on the difference of their carrier frequencies. The time derivative of the angle θ in Eq. (5) can be calculated directly by using MATLAB. Alternatively, an expansion of this term in Bessel's functions is presented by Panter (1965). When $\omega_{cC}\neq\omega_{cA}$ (as for two competing radio stations) the relative phase between the two signals changes with time and the value of the phase ψ_0 doesn't matter—it can be neglected. When both carrier frequencies are the same, as for optical interferometry, the phase ψ_0 depends on the distance between the two surfaces without stimulation and is an important parameter.

In practice, FM receivers include a cascade of nonideal limiters and bandpass filters. In contrast, the theoretical expression Eq. (5) corresponds to an ideal limiter without filter. The cascade of limiters and filters was developed in part to

reduce the effect of competing signals and, in addition to Eq. (5), a computer model has been developed that probes the effect of the cascade.

b. How the FM receiver works—Principle of the model. One step of the FM receiver is to shift the carrier frequency to an intermediate frequency (11 MHz, IF bandwidth $=150$ kHz for our receiver). Then, the signal goes through a limiter: An ideal limiter clips the signal in order to eliminate any amplitude variations. Due to the limiting process, the spectral components of the signal are spread out. A bandpass filter following the limiter passes spectral components centered about the carrier frequency and removes their harmonics. The bandwidth of the bandpass filter is the IF bandwidth. Baghdady (1956, 1961) discusses the competing signal problem, and how the size of the bandwidth should depend on the ratio of signal strengths. The limiter bandwidth must be wide enough to pass a sufficient number of sideband components to add up to a resultant signal whose average frequency at the input of the discriminator is equal to the frequency of the stronger of the two competing signals. During the process of amplitude limiting followed by a filtering, a reduction of the effective amplitude of the competing signal is achieved. For more details see Middleton (1981, 1996). The way we implemented the computer-model FM demodulator was as follows: We started with the signal $x(t)$, clipped it to retain only the zero-crossing times, and then bandpass filtered it. The filter was a window applied in the frequency domain with a bandwidth of 150 kHz. The clipping and filtering was repeated five times. Finally, the instantaneous frequency was found with the zero-crossing times of the processed signal. [Zero crossings were found with software; the way that an electronic circuit figures zero-crossing times was described in Cooper (1999a).] Our objective with this simple model was to demonstrate the basic method by which FM demodulator electronics can reduce the effect of competing signals beyond what Eq. (5) predicts, and for that it was useful.

c. Realistic parameter values. Before delving into the analysis of Eq. (5), it is useful to consider the β, X_o, and ρ values that normally occur in an intracochlear measurement. The division between high- and low-modulation index is at a β value of $\sim 0.5-2$, corresponding to displacements, $X_0 = 25-100$ nm (using a He–Ne-like diode laser, $\lambda = 638$ nm). The measured range of displacement depends on cochlear position (apical vs basal), the structure considered (basilar membrane, Hensen's cell), and on species. For basal basilar-membrane (b.m.) measurements in chinchillas, X_o extends from 0.1 to 200 nm. For guinea pigs, basal b.m. displacements range from 0.05–40 nm. For tectorial membrane measurements in chinchillas, the measured displacements extend from 0.7 to 200 nm [e.g., the reviews of Robles and Ruggero (2001) and Ulfendahl (1997)]. For apical b.m. measurements in guinea pigs, the displacement extends from 6 to 36 nm (Khanna, 1998). Clearly, the intracochlear measurements extend into both high and low modulation cases.

$\rho = c/a$ is the ratio of signal coming from the secondary surface (**C**) to signal coming from the primary surface (**A**). To be precise, the signal strength (carrier level of the interferometer signal) from an object in the cochlea depends on several quantities: the object's reflectivity, R; the object's relative interference amplitude, m [m represents wavefront distortion and takes values less than 1 (Khanna *et al.*, 1996)]; on the distance between the object and the plane of focus (d), and the value (k) the carrier power takes at this distance due to the optical sectioning. R, m, and d depend on the cochlea and k depends on the interferometer optics. The heterodyne signal power from the surface **A** will be proportional to ($m_A^2 R_A k_A$) and the heterodyne signal voltage (a from the above) to $\sqrt{m_A^2 R_A k_A}$. Finally, the ratio $\rho = c/a$ is found by taking the ratio of signal voltages from two surfaces. Consider the following example, in which the motion of the organ of Corti is probed in guinea pig from the scala vestibuli side. Physiologically it is an interesting approach as measurements of several key cochlear structures—Hensen's cells, hair cells, and basilar membrane, can be made (Khanna and Hao, 2000). In order to measure basilar-membrane motion from scala vestibuli adjacent to the Hensen's border region, the incident laser beam passes through the Reissner's membrane, through Claudius's cells, and is focused on the b.m. (Khanna and Hao, 2000). For the b.m., which is in focus, $k=1$. The reflectivities and relative interference amplitudes of the b.m. in the tunnel region, outer hair cells, Hensen's cells, and Reissner's membrane were measured with an apical turn approach, and the values reported in Khanna *et al.* (1996).

The values in the table below are from that reference. In guinea pig, Hensen's cells are quite reflective due the lipid droplets they contain. The relative sizes of the signal from the target (basilar membrane) and the competitors (Reissner's membrane or Hensen's cells) are found by comparing $\sqrt{m^2 R k}$ for the three. The ratio of the $\sqrt{m^2 R k}$ values gives ρ. The relative interference amplitude (m) depends on the state of focusing. The representative m value of a tissue is found with the surface in focus. When the surface is out of focus, the m value might change because the tissue surface is not flat. The change in m with defocus has not been measured, and m is treated as a constant in Table II. Therefore, the ρ values listed and competition indicated are upper bounds. This ratio, listed in the final column of Table II, will guide the interpretation of the studies on competing signals below. In this example the reflective Hensen's cells, with a ρ value that varies between 0.06 and 0.13, pose the most serious challenge to a measurement of b.m. motion from scala vestibuli.

In summary, both low and high β are found in intracochlear motion measurements. Regarding ρ, most major surfaces in the cochlea are separated by at least 100 μm. With this separation, in the special case of two surfaces with equal reflectivity, the ρ values are less than 0.1 for all the systems in Table I. When the surfaces do not have equal reflectivity, steep optical sectioning is needed to reduce ρ, as the example above makes clear. As cochlear measurements begin to probe motion at the cellular level, steep optical sectioning becomes even more essential.

d. Numerical results. In this section the ideal demodulation described by Eq. (5) and the demodulator cascade model are used to study the influence of β_C, β_A, ρ, and ψ_0 on $E(t)$. Results obtained with both methods are compared. We only show results for the case in which $\omega_{cC} = \omega_{cA} = \omega_c$ be-

TABLE II. Calculation of relative signals from b.m, r.m. and H.C. when laser beam is focused on the b.m. d = distance from b.m., k = optical sectioning factor, R = reflectivity, m = wavefront distortion factor, m^2Rk = relative carrier power, ρ = square root of the ratio of carrier power of out-of-focus structure to carrier power of in-focus b.m.

	d	k	$R\times10^{-5}$	m	$m^2Rk\times10^{-7}$	ρ
Basilar membrane	0	1	2.1	0.14	4	
Reissner's membrane	$\sim$290 μm	10^{-4}	3.5	0.13	0.0006	0.012
Hensen's cells (lipid droplets) shoulder region	$\sim$60 μm	10^{-4}	38–89	0.19–0.28	0.014–0.07	0.06–0.13

cause this applies in heterodyne interferometry. We studied in depth the case in which the competing surface was stationary: $\beta_C=0$ and $\beta_A\neq0$. This corresponds to the experimental condition of Fig. 1, in which the stationary, highly reflective bone is the competing signal. We also considered the case that corresponds to the primary surface stationary and the secondary surface moving: $\beta_C\neq0$, $\beta_A=0$. Results from this case can be used to understand the interference that arises when a competing surface moves much more than the in-focus surface.

Figure 3 shows $E(t)$ using Eq. (5) (ideal limiter without filter) for both low and high modulation index (upper curve: $\beta_A=0.2$; lower curve: $\beta_A=4$; $\beta_C=0$ for both curves). Results without competing signal (thin line: $\rho=0$) are compared to those with competition (thick and dashed curves: $\rho=0.6$ and $\psi_0=0$ deg). Based on the section above, the ratio used for this plot is much larger than what would normally occur in the cochlear experiments; the purpose of the plot is to clearly illustrate the differing effects of a competing signal when the modulation index of the primary signal was low compared to when it was high. At low modulation index the output signal was reduced in size but undistorted. The high-modulation-index signal was distorted but the amplitude of the fundamental frequency was unchanged. The dashed curves in Fig. 3 illustrate the effect of the FM demodulator cascade. The cascade of limiters and filters produced no change for the low-modulation-index case (the dashed curve is superimposed on the thick one), whereas at high modulation index the distortion in the time domain was reduced. The characteristics of the spectrum of an FM signal are helpful to explain this observation: The limiter, which spreads the components of the spectrum, has more effect at high modulation index, where the spectrum contains a carrier component and an infinite set of sidebands, compared to the low-modulation case, where the significant components are at $f_c\pm f_m$. The filter removes more sidebands for the high-modulation-index case.

Figure 3 showed the influence of a competing signal on $E(t)$ in the time domain. In Fig. 4 we look at the same results in the frequency domain, and consider the amplitude of $E(t)$ at the stimulus frequency, f_m. Difference in the amplitude of $E(t)$ at frequency f_m is compared to the case without a competing signal (where $\rho=0$). The result is shown as a function of the ratio ρ for different values of phase, ψ_0. Figure 4(A) shows the effect of an unmodulated signal ($\beta_C=0$) of amplitude c on a low-modulation-index signal ($\beta_A=0.2$). The modulation frequency was 9.6 kHz. This panel illustrates the strong dependence of the output on the relative phase between competing signals. When the phase between the signals was equal to 0 deg, the output decreased as the

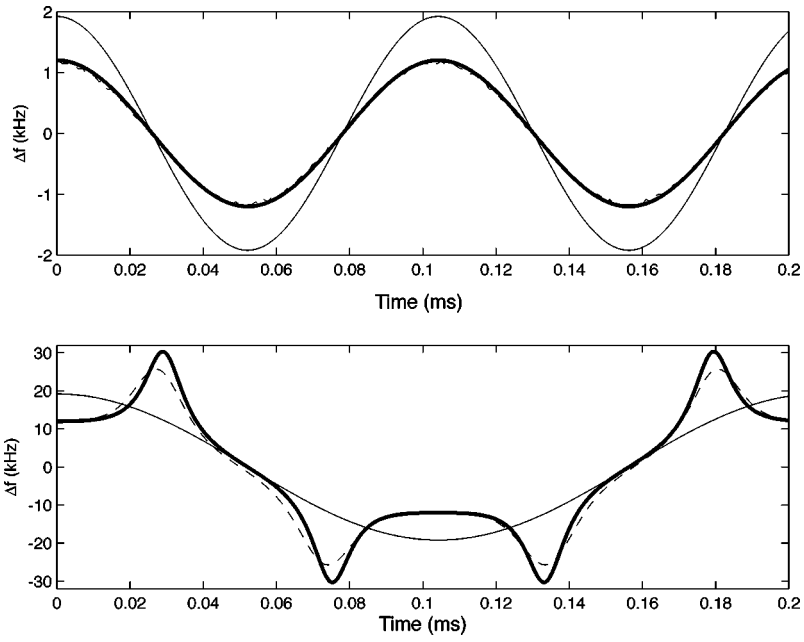


FIG. 3. Comparison of calculated results using Eq. (5) (thin and thick lines) with simulated model FM receiver results (dashed line), with and without a competing signal. The top panel shows demodulator output for low modulation index ($\beta_A=0.2$, $f_m=9.6$ kHz, $\Delta f=1.92$ kHz, $\psi_0=0$ deg), while the bottom panel shows results for high modulation index ($\beta_A=4$, $f_m=4.8$ kHz, $\Delta f=19.2$ kHz, $\psi_0=0$ deg). The thin line on both panels shows output without competition ($\rho=0$). Thick and dashed lines on both panels show demodulator output with competition ($\rho=0.6$). For low modulation index (top), a strong competing signal reduces the amplitude of the output by 4 dB, but the waveform remains unchanged. The results using Eq. (5) are very close to those obtained with the simulated FM receiver. For high modulation index (bottom), the time waveform is distorted by the strong competing signal. The results from Eq. (5) are close to those obtained by simulated FM receiver. The distortion predicted by the latter is slightly lower.

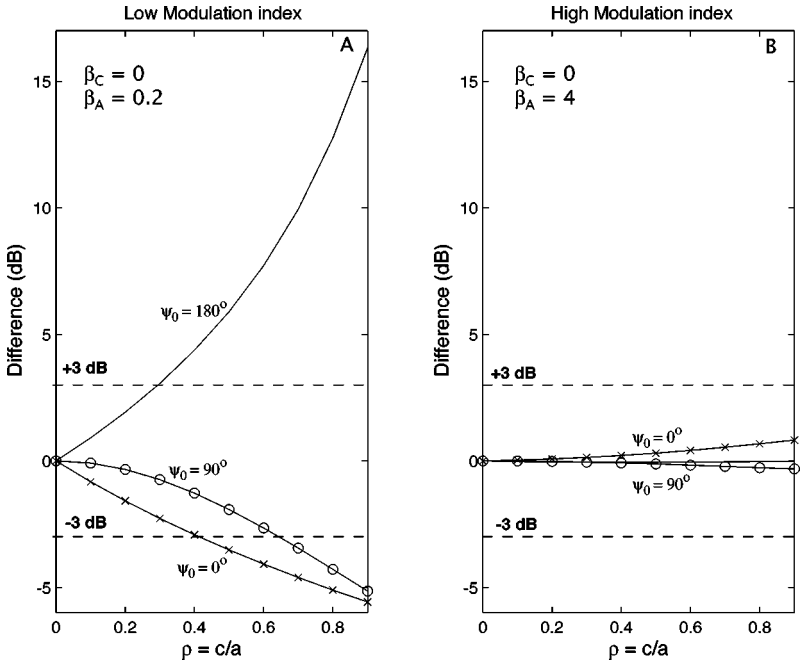


FIG. 4. Changes in the demodulator output amplitude due to the presence of a competing signal as a function of the ratio ($\rho = c/a$) and for different values of phase ψ_0: (x) 0 deg; (o) 90 deg; and (-) 180 deg. Panel (A), low modulation index of 0.2 ($\beta_A = 0.2$, $\beta_C = 0$, $f_m = 9.6$ kHz). Demodulator output can decrease or increase substantially depending on the relative phase of the two competing signals and their relative amplitudes. The worst case is when the two signals are 180 deg out of phase. Panel (B), high modulation index of 4 ($\beta_A = 4$, $\beta_C = 0$, $f_m = 4.8$ kHz). The demodulator output at the stimulus frequency remains substantially unchanged even in the presence of a strong competing carrier.

ratio increased (as the relative size of the unmodulated signal increased). When the ratio was equal to 0.9, a decrease of 5 dB was observed. This result suggests that in the low-modulation-index case the effect of a competing signal can be thought of in terms of a weighted average: The output is approximately equal to the amplitude (strength) of each signal (a or c) multiplied by its own frequency deviation (which in the case of the cochlear experiments is proportional to the velocity, and in the case at hand was zero for signal **C**), and divided by $(a+c)$. An expansion of Eq. (5) for $\beta \ll 1$ confirms this. When the phase between the signals was equal to 180 deg the weighted average idea still applies, except the divisor is equal to $(a-c)$, which produces the observed *increase* in output. The result highlights the strong effect of the phase. In terms of an optical experiment, 180 deg phase will occur when twice the distance between the two surfaces is $(n+\frac{1}{2})\,\lambda$, where n is an integer. The 0 deg phase will occur when twice the distance between the two surfaces is $n\lambda$. With $\psi_0 = 0$ deg, to have an error on the output less than $+/-3$ dB, the ratio of the interfering signal over the modulated signal needed to be less than 0.4 and, when $\psi_0 = 180$ deg, this value was smaller, 0.3. ψ_0 of 0 and 180 deg were worst cases, and intermediate phases showed smaller errors. In general, harmonic distortion was small for low modulation index: For $\beta_A < 0.5$, with a ratio of 0.4 or less, the amplitude of harmonic components was at least 30 dB down (with $\beta_A = 0.2$, $\beta_C = 0$, $\psi_0 = 90$ deg) compared to the fundamental (not shown).

Figure 4(B) shows the change in $E(t)$ with ρ at the frequency f_m due to a competing signal in the high-modulation-index case ($\beta_A = 4$, $\beta_C = 0$, $f_m = 4.8$ kHz). Changes in the output at the stimulus frequency were less than 1 dB for all ratios and phases. However, as shown in Fig. 3, the output was distorted in the time domain. In order to have the harmonic components at least 25 dB down, the ratio needed to be less than 0.1 (not shown). Based on the ρ values in Table II, even with steep optical sectioning a detection of distortion must be treated with caution when competition from secondary signals is a possibility and displacements are relatively large.

To summarize the theoretical section in the case of $\beta_C = 0$: The theoretical effect of competing signals at the input of the receiver was described first in the time domain. At low modulation index, there was a change of the amplitude of the signal but it was not distorted. The design of the receiver (cascade of limiters and filters) did not reduce the effect of a competing signal. At high modulation index the competing signal produced distortion in the output that could be reduced by the processing of the FM receiver. In the frequency domain the theoretical influence of a competing signal on the receiver output at the fundamental frequency was shown in Fig. 4. For low modulation index, an increase or decrease of the output was found, depending on the relative phase of the two signals. The simple notion of a weighted average gave a reasonable estimate of the effect of competing signals. At high modulation index, with respect to the amplitude of the fundamental component the output error was always less than 1 dB, even when the competing signals were nearly equal [Fig. 4(B)]. However, the theoretical output waveform was distorted in the presence of the competing signal.

We began with the case $\beta_C = 0$ and $\beta_A \neq 0$. Equation (5) is now used to explore the opposite case, $\beta_C \neq 0$ and $\beta_A = 0$. When the input is the unchallenged signal **A**, the output is zero. The time-domain plots in Fig. 5 show the effect of either low-modulation (left column) or high-modulation (right column) competing signal **C**. The top panel in each column shows the frequency deviation of the competing signal **C**—the output if signal **C** was the sole input. The middle and bottom panels show the effect of this competing signal when it is combined with signal **A** ($\beta_A = 0$) at ratios of 0.3 and 0.05. The low-modulation/high-modulation classification of results found above still applies. The low-modulation competitor leads to an undistorted, but nonzero output that is like the output for the competing signal **C** alone shown in the

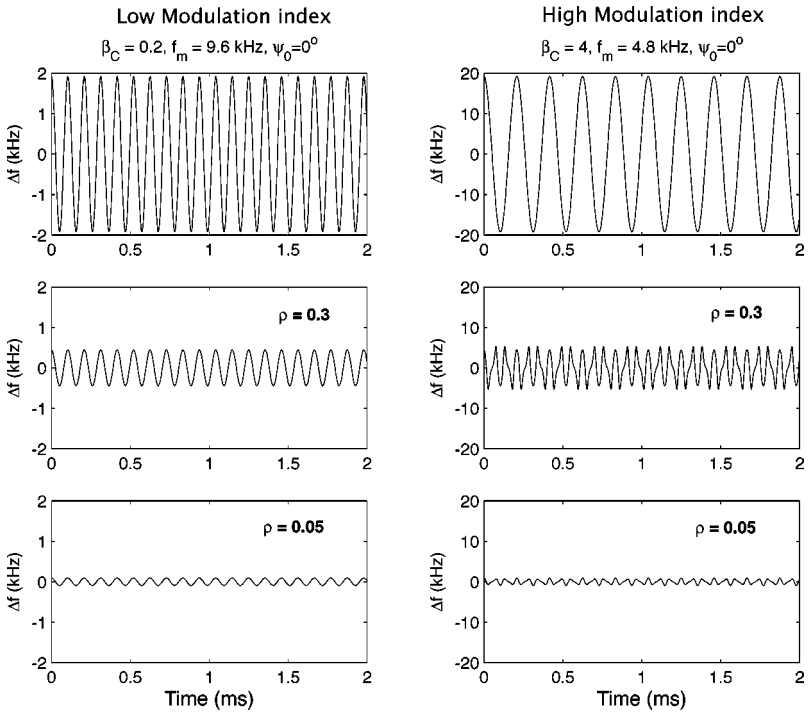


FIG. 5. Results calculated from Eq. (5) with $\psi_0 = 0$ deg. This experiment simulates the influence of a vibrating surface when the interferometer is focused on a stationary surface ($\beta_A = 0$). The unmodulated carrier is larger than the competing modulated carrier signal. The upper curves correspond to the competing signal itself ($a = 0$). The middle and lower panels show the effect of a modulated signal combined with signal **A** ($\beta_A = 0$, $\beta_C \neq 0$) for a ratio ($\rho = c/a$) of 0.3 and 0.05. For the low modulation index ($\beta_C = 0.2$, left column) even when the competition is weak (0.05) the demodulator output contains a modulation component. For the high modulation index case ($\beta_C = 4$, right column) the demodulator output is highly distorted.

top panel but with a substantially reduced amplitude. This is as expected since the unmodulated signal **A** is stronger than **C**. The size of the output scales with ρ, the relative strength of competing signal **C**, and as above, a weighted-average rule of thumb applies for the low-modulation case. The high-modulation competitor leads to a highly distorted output. The distorted output has odd symmetry and possesses only odd harmonics because it was calculated with ψ_0 equal to zero; in general, the output is composed of both even and odd harmonics.

2. Electronic experiments with two generators

To simulate competing signals, x_A and x_C were generated with two signal generators. Again, we begin with the case $\beta_A \neq 0$, $\beta_C = 0$. The first generator produced a frequency-modulated signal with a carrier frequency 455 kHz, corresponding to interference between the reference beam and **A** ($x_A(t) = a\cos(\omega_{cA}t - \beta_A \cos(\omega_m t))$). The second signal generator produced a signal at 455 kHz with no frequency modulation, corresponding to interference between the reference beam and **C** ($x_C(t) = c\cos(\omega_{cC}t)$). To test the above theoretical results, the sum $x(t) = x_A + x_C$ was frequency shifted to 100 MHz and then applied at the input of our demodulator and the output was measured. (The shift in frequency is a processing detail that is not important to the message of this paper.) Although both signal generators were set with $f_c = 455$ kHz, in fact the two generators were not phase locked and their carrier frequencies were slightly different. They were close in frequency, so the resulting sum can be thought of as the sum of two signals with the same carrier frequency and a slowly varying relative phase ($\omega_{cC}t = (\omega_{cA} + d\omega)t = \omega_{cA}t + \psi_0(t)$). Thus, the experimental setup could be compared to the theoretical situation of Eq. (5) with $\omega_{cC} = \omega_{cA}$ and ψ_0 slowly varying. As ψ_0 varied slowly from 0 to 180 deg to 360 deg, the effects of the phase were compared with the predictions of Fig. 4. f_m and Δf were chosen to represent different values of the modulation index ($0.04 < \beta < 4$). f_m values were 4.8 and 9.6 kHz and Δf was from 200 to 38 000 Hz (so that in terms of a cochlear experiment, the velocity amplitude "V_o" was 0.06–12 mm/s and the displacement amplitude "X_o" was 2–200 nm). The output of the FM demodulator, $E_g(t)$ (the subscript g designates the generator experiments) was measured first with just signal **A**, and then as the level of signal **C** was increased (increase of the ratio ρ). For each ratio, $E_g(t)$ was measured and stored. The amplitude of $E_g(t)$ at the modulation frequency f_m was determined by Fourier transform (FT). The FT was performed over a portion of $E_g(t)$ where the two signals were in phase ($\psi_0 = 0$ deg) and out of phase ($\psi_0 = 180$ deg), as well as over the whole signal.

The upper curve of Fig. 6 shows $E_g(t)$ as a function of time ($\beta_A = 0.2$, $f_m = 9.6$ kHz, $\rho = 0.6$). (Note that this ratio is much larger than what is expected to occur in cochlear measurements. As above, the large ratio is used for illustrative purposes.) The signal in the time domain contained regularly spaced peaks. We could interpret these peaks in terms of the time-varying relative phase, ψ_0. At the time of the maxima in $E_g(t)$, the input signals were out of phase and at the time of the minima, the signals were in phase. The time between two maxima was the period of the frequency difference, $1/(f_{cC} - f_{cA})$. The lower curve of Fig. 6 is a zoom on a small portion (0.2 ms) of the upper curve, at a time when x_A and x_C were in phase. The theoretical output was calculated using Eq. (5) (dashed line; lower curve Fig. 6) and compared to the experimental one (solid line). The two figures agreed well, indicating that $E(t)$ resulting from the theory was useful to predict the receiver output, $E_g(t)$.

Figure 7 shows a high-modulation-index case. $E_g(t)$ is shown as a function of time as the bold curve ($\beta_A = 4$, $f_m = 4.8$ kHz, $\rho = 0.6$). The 1-ms periodicity is due to the time-

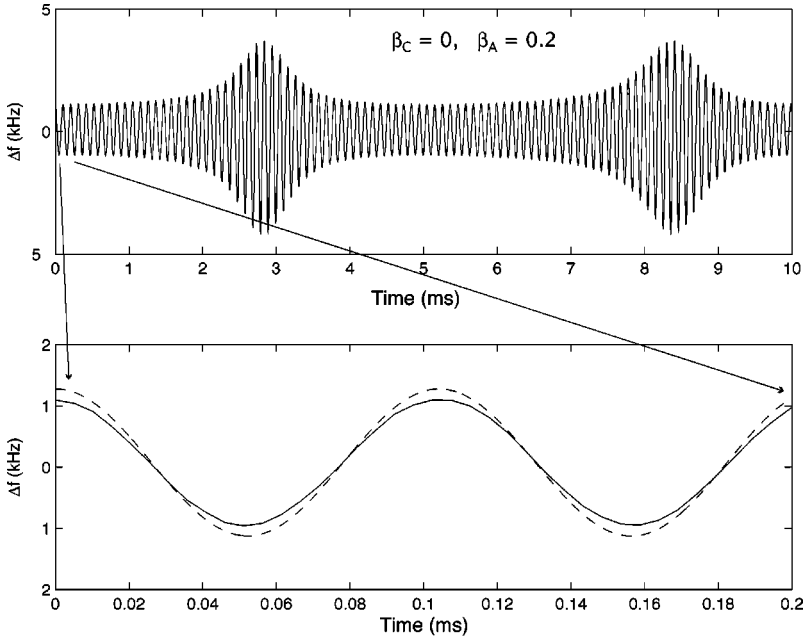


FIG. 6. Upper curve: Experimental output of the FM receiver as a function of time ($\beta_A=0.2$, $\beta_C=0$, $f_m=9.6$ kHz, $\rho=0.6$). Lower curve: superposition of the theoretical instantaneous frequency (dashed line) and the experimental one (solid line) over a short period of time. The generators producing the FM and the competing signals were not phase locked. Their relative phase changed with time and, as a consequence, the demodulator output amplitude fluctuated with time increasing to a peak when the phase difference was 180 deg, and decreasing to a minimum when the phase angle was 0 deg [see Fig. 4(A)].

varying relative phase between the primary and secondary input signals. We include two other curves for comparison. The dashed curve shows the unchallenged output due to signal **A** (when signal **C** was off). The bold curve is distorted relative to the unchallenged signal. During the course of the measurement, the relative phase between the two signals, changes and this is reflected in the changing wave shape of the bold curve. Its overall size does not change. The thinner curve shows the predicted effect of the competing signal using Eq. (5). It is slightly more distorted than the output of the demodulator, confirming that the actual workings of the FM demodulator reduce the effect of competing signals in the high-modulation-index case.

Figure 8 presents an analysis of experimental data like that of Figs. 6 and 7 for a wide range of ratio ρ. In Fig. 8, the change in $E_g(t)$ (amplitude at f_m) due to a secondary signal compared to $E_g(t)$ without competing signal ($\rho=0$) is shown as a function of the ratio ρ. For each of the curves, the Fourier transform is performed in a different part of the time domain of a curve like Fig. 6, which allows us to infer the influence of the phase. Panel (A) corresponds to the low-modulation-index case (analysis of experimental data like that of Fig. 6). When the Fourier transform is performed in the part of the time domain where $E_g(t)$ is peaked at a maximum, corresponding to $\psi_0=180$ deg (plain curve), the competing signal caused an increase in the receiver output of 7 dB for a ratio of 0.6. When the FT was done where $E_g(t)$ was minimum, corresponding to $\psi_0=0$ deg (curve with crosses), then there was a decrease of 4 dB for the same ratio of 0.6. Figure 8(A) can be compared to the theoretical results shown in Fig. 4(A) (low modulation of $\beta_A=0.2$, $f_m=9.6$ kHz). They agree well. When the FT was done over

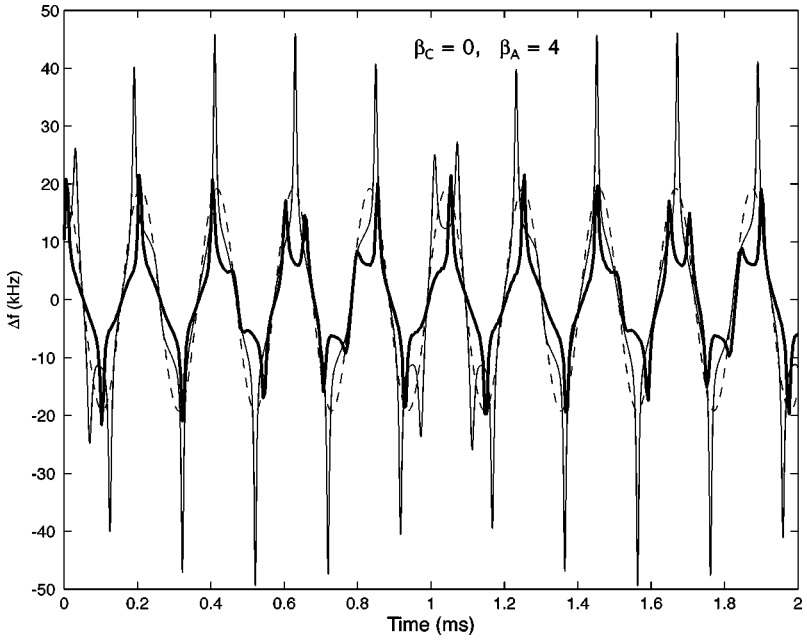


FIG. 7. Effect of a competing signal on the demodulator output (high modulation index: $\beta_A=4$, $\beta_C=0$, $f_m=4.8$ kHz, $\rho=0.6$). The bold curve was measured experimentally and is compared with calculated curves: The thin curve is the theoretical result using Eq. (5), whereas the dashed curve is the expected curve without competing signal. Both experimental and theoretical results show distortion of the sinusoidal waveform. The distortion in the experimental results is slightly less than that predicted by Eq. (5).

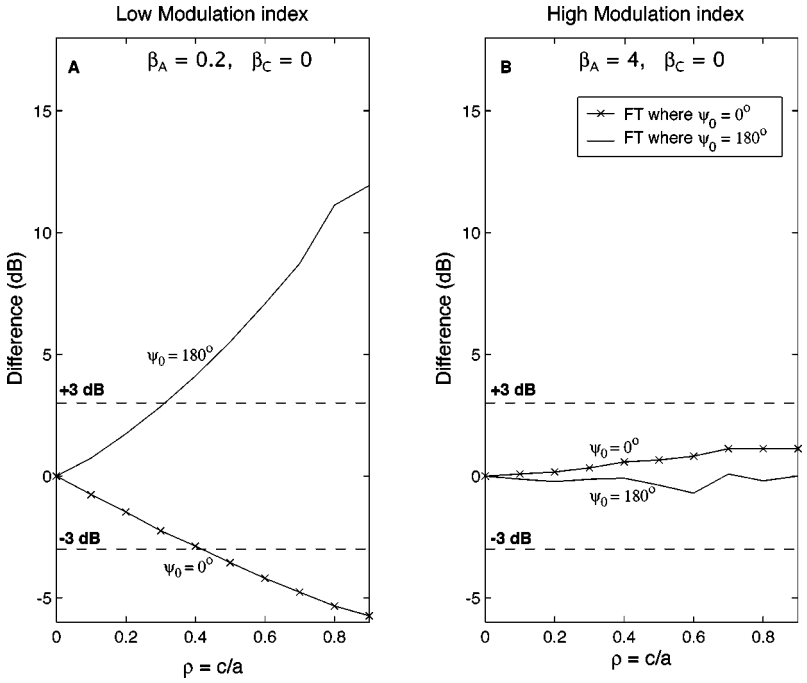


FIG. 8. Change in demodulator output (amplitude at f_m) with different magnitudes of competing carrier levels (ρ). Experiments were similar to those described in Figs. 6 and 7, but with ratio ρ ranging from 0 to 0.9. The values in the solid curve were derived from the peak region of the demodulator time varying output (where $\psi_0 = 180$ deg). Values in the curve with crosses were derived from the valley region of the curves (where $\psi_0 = 0$ deg). Panel (A) corresponds to the low-modulation-index case ($\beta_A = 0.2$, $\beta_C = 0$, $f_m = 9.6$ kHz) and panel (B) to the high-modulation-index case ($\beta_A = 4$, $\beta_C = 0$, $f_m = 4.8$ kHz). Results are similar to those predicted by theory (Fig. 4).

the entire signal, there was an average over the phase, and the output, even for a ratio of 0.9, was close to the output of the signal without competition (not shown). Therefore, when signals have different carrier frequencies, a time average can minimize the effect of competing signals. However, in the case of the interferometer, both secondary and primary signals have the same carrier frequency, and time averaging does not reduce the effect of competition.

Figure 8(B) presents an analysis of experimental data at large β ($\beta_A = 4$, $f_m = 4.8$ kHz). It illustrates the weak dependence of the amplitude at the stimulus frequency on the relative phase between carrier signals for the high-modulation-index case. The error is within +/-1 dB for all phases and ratios. Figure 8(B) can be compared to the theoretical results shown in Fig. 4(B). They agree well.

To summarize the experimental results with generated signals in the case $\beta_A \neq 0$, $\beta_C = 0$ (secondary signal not frequency modulated):

At low modulation index (β_A) when two signals with approximately the same carrier frequency were summed at the input of the receiver, and when they were in phase, the error due to the competing signal **C** produced a decrease of the demodulator output. When they were out of phase, the error produced an increase in the demodulator output. The FM demodulator output depended strongly on the ratio of the amplitude of both signals and their relative phase. The output waveform was not distorted by the competing signal.

For high modulation index (β_A), the output of an FM demodulator at the stimulus frequency was not influenced by the competing signal **C** even when the secondary signal was nearly as large as the primary signal. However, the output signal was distorted.

The most important conclusion from these results is that the theoretical expression for $E(t)$ in Eq. (5) was very useful for predicting the output of the FM demodulator, $E_g(t)$. In the high-modulation-index case, the distortion predicted by Eq. (5) is an upper bound, as the FM demodulator did offer some improvement in reducing distortion.

The study with signal generators and FM demodulator above concerned the case $\beta_C = 0$ and $\beta_A \neq 0$. The opposite case, $\beta_C \neq 0$ and $\beta_A = 0$, has also been explored. Curves similar to those of Fig. 5 were obtained (not shown). These results confirm the low-modulation/high-modulation classification of results that was noted already, and also confirm the usefulness of Eq. (5) for predicting the demodulation.

3. Optical experiments with the interferometer

Above we compared the theoretical predicted effect of competing FM signals to experimental results using generated FM signals and our REVOX FM demodulator. Below we complete the study by measuring the effect of competing signals that were produced by reflective surfaces, using the confocal-heterodyne interferometer developed by Khanna *et al.*, 1996. We explore first the case in which the secondary surface was fixed, $\beta_C = 0$, $\beta_A \neq 0$, and then the case in which both surfaces were moving, $\beta_A \neq 0$, $\beta_C \neq 0$. The setup is shown in Fig. 9. The system was composed of two surfaces: A glass tube with a semitransparent thin plastic membrane at its end was used as a vibrating object. It was fixed to a stiff steel rod. The back surface was a polished steel fiber, fixed rigidly to a micropositioner. The position of this fiber was adjusted by using a calibrated piezoelectric translator. Vibrations were measured for several separations of the front and back surfaces. The fiber was inserted at a distance as close as 22 μm from the front surface. The front surface was vibrated by driving it with a speaker at $\omega_m = 2\pi f_m$. The back surface was stiff and did not vibrate with the sound stimulation. The laser's object beam was focused at the center of the membrane. The primary signal, $x_A(t)$, was produced by interference between the object beam that reflected from the membrane and the reference beam. The secondary signal, $x_C(t)$,

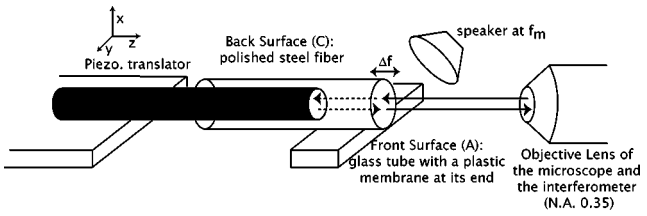


FIG. 9. The setup shown was used to optically simulate the multiple reflection conditions that might be encountered in the cochlea. A thin plastic membrane at the end of a glass tube (diameter=1.2 mm) was used as one of the partially transmitting/reflecting surfaces. This membrane could be vibrated with the sound from a speaker placed near it. By changing the frequency and intensity of sound, the frequency and amplitude of membrane vibration could be controlled. The vibrating membrane produced the frequency-modulated component of the carrier. A polished surface of a steel fiber inserted in the tube provided the second reflecting surface. This surface was considered stationary and provided the unmodulated component of the carrier. Piezoelectric and mechanical translators allowed the steel fiber to be positioned precisely with respect to the plastic membrane. Mechanical translators in the interferometer were used to position the whole assembly together with respect to the objective lens. This allowed the position of the focal plane to be changed and measured with 1-μm resolution. Carrier level (measured using a spectrum analyzer) and demodulator output (proportional to the velocity) were measured as a function of distance from the front surface.

was produced by interference between the object beam reflected from the fixed back surface and the reference beam. The relative phase between the two signals was constant and depended on the distance (d) between the two surfaces without stimulation: $\psi_0 = 2d2\pi/\lambda$, with λ the laser wavelength. Measurements were made at different positions of the plane of focus: from 10 μm in front of the membrane to 10 μm behind the back surface. For each position, the carrier level and the demodulator output were measured.

First, the vibration of the membrane alone (front surface) was measured. The measured carrier level was always at least 30 dB above the shot-noise level (-68 dB), and, consistent with the discussion of single-surface vibration above, there was no change in the reported vibration amplitude as a function of the distance from the membrane (from -32 to $+32$ μm).

The predictions of Eq. (5) were compared to our experimental data obtained with two surfaces. The values needed for Eq. (5) were the modulation index (determined from the vibration of the front surface when focused on it), the amplitude ratio of the two signals (which varies as the laser is focused at different planes) and their relative phase, ψ_0. The distance between both surfaces was known within a precision of $+/-1$ μm, which was not accurate enough to determine the phase. The phase was a free parameter in the analysis, chosen in order to fit the experimental data.

The ratio c/a was found using the measured carrier levels from the two surfaces together, and the measurement of the optical sectioning curve (carrier level as a function of distance from the focal plane) from a single surface (Fig. 2). The ac signal at the demodulator is given by $x(t)$ of Eq. (4), and the carrier level corresponds to the power in this signal [proportional to $x(t)^2$]. In the case at hand, $\omega_{cC} = \omega_{cA}$. Therefore, the theoretical two-surface carrier power as a function of distance from the focal plane (theoretical two-surface optical sectioning curve) is

$$\text{carrier power (dB)} = \text{constant} + 10\log_{10}(a^2 + c^2 + 2ac\cos(\psi_0)), \quad (6)$$

where a and c were functions of distance, $a = a(z)$, $c = c(z)$. We have assumed that each could be described by the square root of the carrier power vs distance curve with shape as in Fig. 2. ψ_0 was fixed, depending only on the distance between the surfaces. The theoretical two-surface carrier power was calculated using Eq. (6), and compared to the measured two-surface carrier level vs distance curve. The positions of the front and back surfaces were known roughly (± 1 μm) but could be adjusted slightly to improve the fit of Eq. (6) to the measured data. The phase ψ_0 was a fixed but unknown value that was freely adjusted for the best fit. From this fitting procedure $a(z)$, $c(z)$, and ψ_0 were found. Equation (5) was then used to predict the receiver output, $E_o(t)$, where the subscript "o" designates the optical experiments.

Ten experiments with a fixed back surface were performed with different distances between the two surfaces (from 22 to 66 μm). For half of them the back surface was more reflective than the front one (carrier level higher when focused on the back surface than on the front surface). Vibrations were measured for eight frequencies (from 1 to 8 kHz). Theoretical results from Eq. (5), $E(t)$, were compared to the data, $E_o(t)$, and agreed well. Results from several experiments are presented below. The first two correspond to low-modulation-index cases, one with a back surface of higher reflectivity than the front surface and one with the front surface more reflective. We have seen that the effect of competition in the low-modulation-index case is on the fundamental frequency component, without introducing distortion. This held true in the optical experiments, and therefore we emphasize the fundamental component in the analysis of these two experiments.

In the first, the front surface was positioned on $z = 0$ μm and the back surface was at 23 μm behind ($z = -23$ μm). The measured carrier level as a function of the distance is presented in Fig. 10(A) (thick curve). It is the composite curve due to the two surfaces. The back surface had higher reflectivity than the front surface. From this curve and knowing what the curve is like for a single surface (measured data in Fig. 2), the amplitudes a and c were extrapolated. The dashed line is the theoretical carrier level obtained from Eq. (6) with $\psi_0 = 100$ deg. The ratio ρ changed from 0.04 at the front surface ($z = 0$ μm) to 0.9 at $z = +13$ μm and to 1 at $z = -12$ μm. For a ratio of 1, the amplitudes of the two signals are equal.

When driven with the loudspeaker at 7 kHz, the motion of the front surface was measured and the modulation index β_A of the front surface was determined to be 0.043. As $\rho(z) = c(z)/a(z)$, $\beta_C = 0$, and ψ_0 and β_A were known, the theoretical expression [Eq. (5)] could be used in order to predict the demodulator output, and this prediction could be compared with the actual demodulator output. Figure 10(B) assembles experimental (line with circles) and theoretical (crosses) demodulator output. Theory and data were similar.

In the second experiment, the distance between the front surface ($z = 0$ μm) and the back surface was 31 μm. The measured carrier level as a function of distance is presented

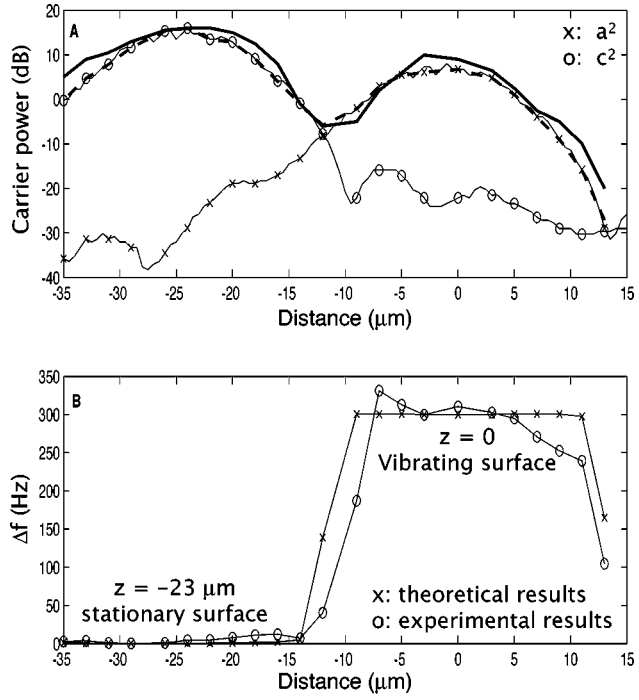


FIG. 10. Interferometer experiment with a vibrating membrane and a stationary back-reflecting surface. The membrane was located at $z=0\ \mu$m and the back surface at $z=-23\ \mu$m. The reflectivity of the back surface was higher. Panel (A): Measured carrier level as a function of distance is shown with a thick line. The carrier level rises and reaches a peak as the plane of the membrane reaches the focal plane. It decreases between the surfaces and reaches a maximum again as the back surface reaches the focal plane. The measured optical sectioning curve (Fig. 2) was used to extrapolate the amplitude of the signals from the two surfaces individually. (The crosses correspond to the extrapolated amplitude of the modulated signal a^2 and the circles to the amplitude of the unmodulated signal c^2). The phase angle ψ_0 between the two carriers could not be measured experimentally and it was selected for best fit. The dashed line is the theoretical carrier level obtained from Eq. (6). Panel (B): Experimentally measured change in the frequency deviation at the modulation frequency as a function of distance from the focal plane (line with circles). Theoretically fitted curve [using Eq. (5)] is shown (line with crosses). ($\beta_A=0.043$, $f_m=7$ kHz, $d=23\ \mu$m, $\psi_0=100$ deg). The theoretical curves fit the measured data quite well.

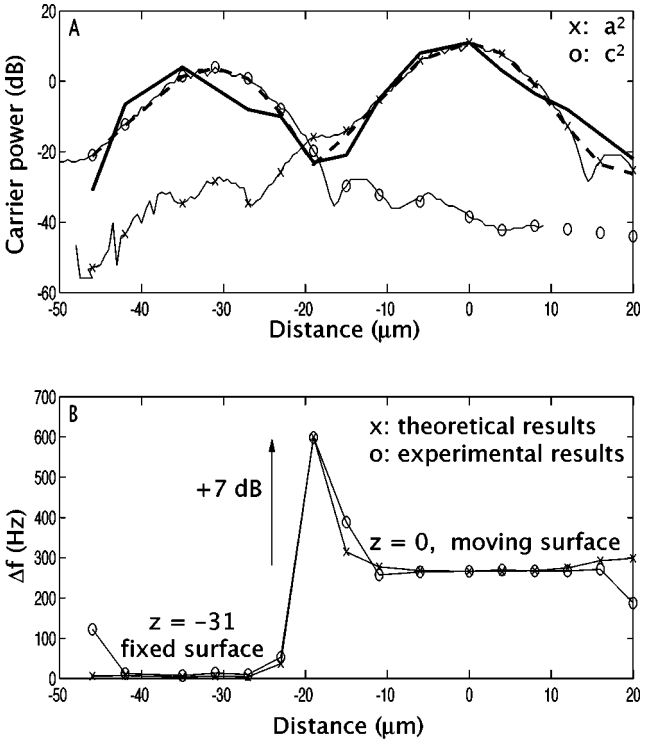


FIG. 11. A second interferometer experiment with a vibrating membrane and a stationary back-reflecting surface. The membrane is located at $z=0\ \mu$m and the stationary back surface is located at $z=-31\ \mu$m. The reflectivity of the front surface was higher. Panel (A): Measured carrier level as a function of distance is shown with a thick line. The dashed line is the theoretical carrier level obtained from Eq. (6). The crosses correspond to the extrapolated amplitude of the modulated signal a^2 and the circles to the amplitude of the unmodulated signal c^2. For this experiment, $\beta_A=0.038$, $f_m=7$ kHz, $d=31\ \mu$m, $\psi_0=166$ deg. Panel (B): Experimental (o) and theoretical (x) frequency deviation (Hz) as a function of distance ($\beta_A=0.038$, $f_m=7$ kHz, $d=31\ \mu$m, $\psi_0=166$ deg). The theoretical results fit the experimental data well, and in particular can account for the peak at $-19\ \mu$m.

in Fig. 11(A). The back surface had lower reflectivity than the front surface. From the extrapolated amplitudes a and c, the theoretical carrier level [Eq. (6)] was determined with $\psi_0=166$ deg (dashed line). The ratio went from 0.004 on the front surface ($z=0\ \mu$m) to 0.1 at $z=+16\ \mu$m and to 0.64 at $z=-19\ \mu$m. When driven with sound, the front surface moved with a small displacement, and produced a low-modulation-index signal ($\beta_A=0.038$, $f_m=7$ kHz). Figure 11(B) compares experimental (line with circles) and theoretical (crosses) data obtained from Eq. (5), knowing ψ_0, the modulation index β_A, $\beta_C=0$, and the values of ρ. An increase of the output of 7 dB (compared to the vibration of the front surface, $z=0$) is observed on $z=-19\ \mu$m. This increase is well explained by the theory. It is due to the relative phase of 166 deg between the signals coming from the two surfaces. Also of note is that even when focused on the back surface, the reported frequency deviation, both theoretical and experimental, although small, was not zero. Therefore, the mobile surface exerts an influence on the measurement at the fixed surface even when the fixed surface has a much larger signal strength. This result is consistent with what was shown in Fig. 5, and reinforces the finding that both relative signal strength and relative β size determine the degree of competition.

With the membrane at the end of a narrow glass tube as a front surface, the modulation index was always low ($\beta_A<0.1$). In order to study a high-modulation-index case, a plastic tube with a bigger diameter membrane was used. We showed above that when the modulation index of the signals is high a competing signal introduces distortion while leaving the fundamental component unaffected. This held true in the optical experiments, and therefore we emphasize the distortion in the analysis of these experiments. When the surfaces were alone, the motion (measured frequency deviation) was almost undistorted (distortion was at least 50 dB down relative to the fundamental), and therefore we attribute the distortion measured when both surfaces were present to the competitive interaction of the two signals. For this experiment, the distance between the two surfaces was $\sim 66\ \mu$m; the modulation index of the front surface, measured at a stimulus frequency of 4 kHz, was 1.77. Figure 12 shows the time-domain signals measured when focused on the mobile front surface [panel (A)] and on the fixed back surface [panel (C)]. The ratio of back/front signal strengths was 0.016 when

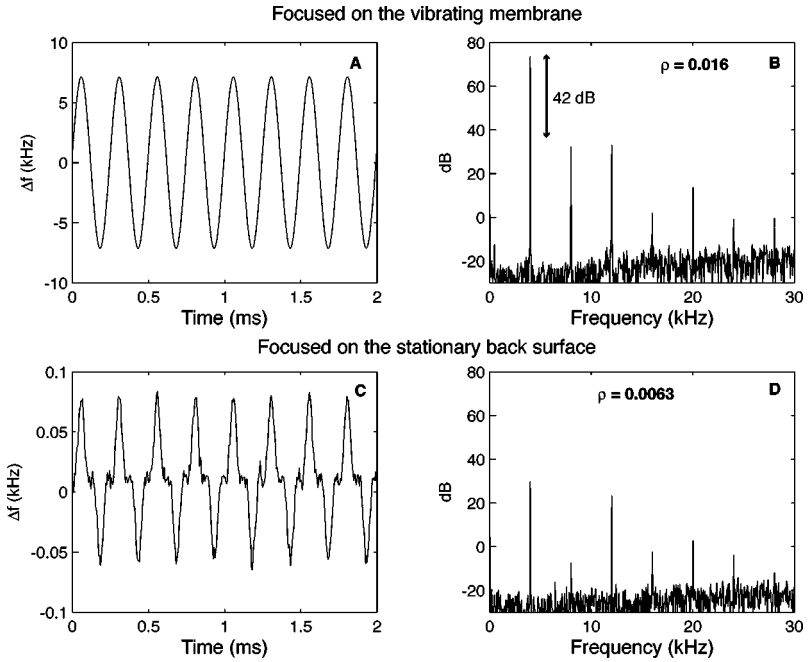


FIG. 12. High-modulation-index case ($f_m=4$ kHz, $\beta_A=1.77$). A larger membrane was used in order to increase its vibration amplitude. The distance between the membrane and the stationary back surface was 66 μm. The upper left curve [panel (A)] is the signal obtained when focused on the front surface, as a function of time; panel (B) is its Fourier transform. The lower curve is the signal obtained with the back surface in focus in the time domain [panel (C)] and in the frequency domain [panel (D)]. The stationary back surface, even though it is far away and its contribution to the total carrier level is small ($\rho=0.016$), can introduce distortion in the demodulator output. Measuring through the vibrating membrane, the stationary surface also appears to vibrate with a distorted waveform.

focused on the front surface, and the ratio of front/back signal strengths was 0.0063 when focused on the back surface. Again, we see that the relatively large motion of the front surface is able to influence the back surface measurement even with a small ratio of signal strengths. The spectra for these two signals are shown in panels (B) and (D). The back surface introduces only a small amount of distortion to the front surface measurement. This is consistent with our expectations, since the back surface is of $\sim$zero β, and the ratio of signal strengths is small.

We conclude the optical experiment section by showing one case in which both front and back surfaces were moving. This was accomplished by threading a glass tube with membrane into the larger plastic tube with membrane and stimulating at 2 kHz. There are no surprises with the results under these conditions, and they serve to round out our experimental results by including a more general case. In Fig. 13, the back surface had $\beta_C=1.14$. The front surface had $\beta_A=1.05$. The ratio of back/front signal strengths was 0.014 when focused on the front surface, and the ratio of front/back signal strengths was 0.032 when focused on the back surface. Panel (A) shows the front surface alone; panel (C) shows the back surface alone (to illustrate the low level of mechanical distortion). Panels (B) and (D) show the distortion introduced by the competing signal of the other surface, when they were separated by 108 μm. [The small change in the fundamental component is difficult to interpret, as the exact position where the motion is measured on the membrane may have changed between panels (C) and (D).] The degree of distortion introduced by the competing signals depends on both the strength of the competing signal and the modulation β of that signal.

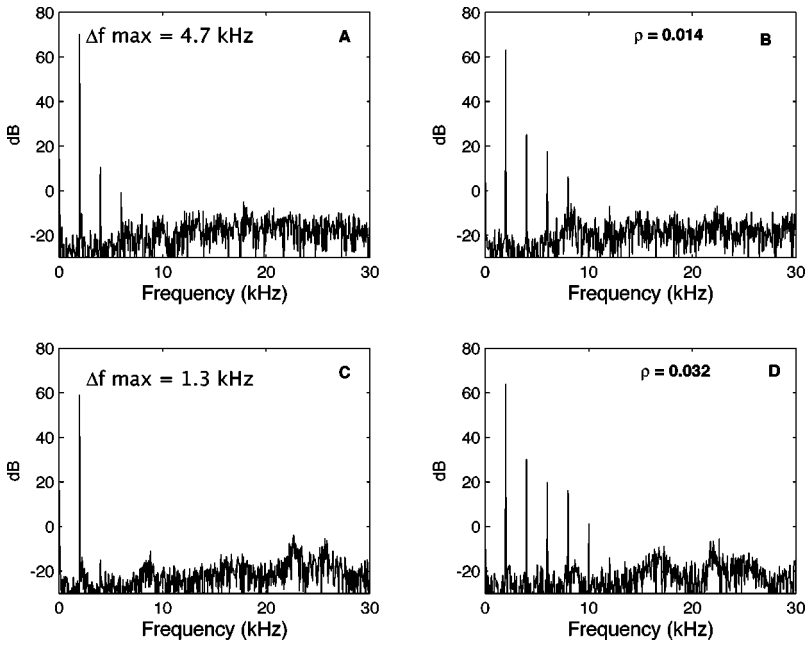


FIG. 13. Two vibrating membranes separated by 108 μm were used in this experiment. Panel (A) shows the Fourier transform of the signal from the front surface alone; panel (C) shows the back surface alone. When measured alone the front membrane vibrated with approximately 11 dB higher amplitude than the back membrane. The second harmonic distortion was low, 60 dB below the fundamental. It was not measurable on the second surface. Panels (B) and (D) show the distortion introduced by the competing signal of the other surface ($\beta_A=1.05$, $\beta_C=1.14$, $f_m=2$ kHz). Panel (B): The front surface is in focus, $\rho=0.014$; panel (D): The back surface is in focus, $\rho=0.032$. When the vibration of each membrane was remeasured with the two membranes in place, the measured amplitude of the first membrane was lowered by about 7 dB (perhaps due to not measuring the same location on the membrane) and its second harmonic distortion increased by 15 dB. The measured amplitude of the second membrane increased by 5 dB and its second harmonic distortion was increased to a level 34 dB below the fundamental. These observations are qualitatively consistent with the theoretical predictions.

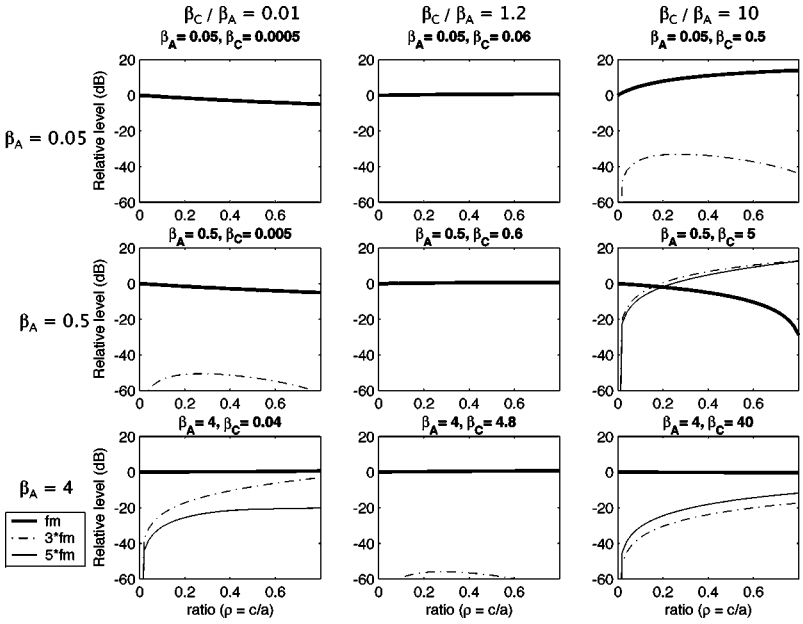


FIG. 14. Effect of a competing signal on the fundamental and harmonic output of the FM demodulator. Change in demodulator output in dB is calculated for various $\rho = c/a$ values and for nine different β combinations. The calculations have been made for $\psi_0 = 0$ deg. The demodulator output at the fundamental frequency can increase, remain unaffected, or decrease sharply depending on the parameters. Only odd harmonics are present when $\psi_0 = 0$ deg. Harmonics are plotted in dB relative to the fundamental level in the absence of competition.

III. BACK TO THE THEORY

From the above, Eq. (5) is very useful to predict the output signal of the FM demodulator when two signals contribute to its input. We conclude the Results section by using Eq. (5) to generate some useful experimental guidelines regarding signal competition. Above, realistic ρ and β values were gleaned from the literature. β values extended from very low modulation index (for low to moderate stimulus levels) to modulation indices at high stimulus levels up to ~ 10. Chinchilla had larger β values than the other species looked at, and guinea pig measurements were always within a low-modulation case, although getting close to the boundary at the highest stimulus levels. Many surfaces in the cochlea are of low reflectivity, the Hensen's cells of guinea pig being a notable exception. If the reflectivities of different surfaces are close to being equal, then the ratio, ρ, is determined largely by the optical sectioning curve and the distance between surfaces. Most surfaces in the cochlea are separated by at least 100 μm, so from Table I (which reports carrier power, so the square root is the relevant quantity) we can see expected ρ values depend on the interferometer and will span values as high as 0.3, and as low as 0.01, with the lower value possible with the low-coherence diode laser of Dalhoff *et al.* and the divided aperture system of Koester *et al.* When the competing surface is of relatively high reflectivity (bone, for example) or is closer, ρ will be accordingly higher. Future measurements that attempt to separate the motion of, for example, tectorial membrane and reticular lamina, will encounter more serious signal competition. The other important parameter is the modulation index ratio. Khanna *et al.* (2000) measured vibrations of Hensen's cell and basilar membrane in an *in vivo* preparation: The basilar-membrane velocity was 10 times smaller, $\beta_{\text{Hensen}}/\beta_{\text{BM}} = 10$

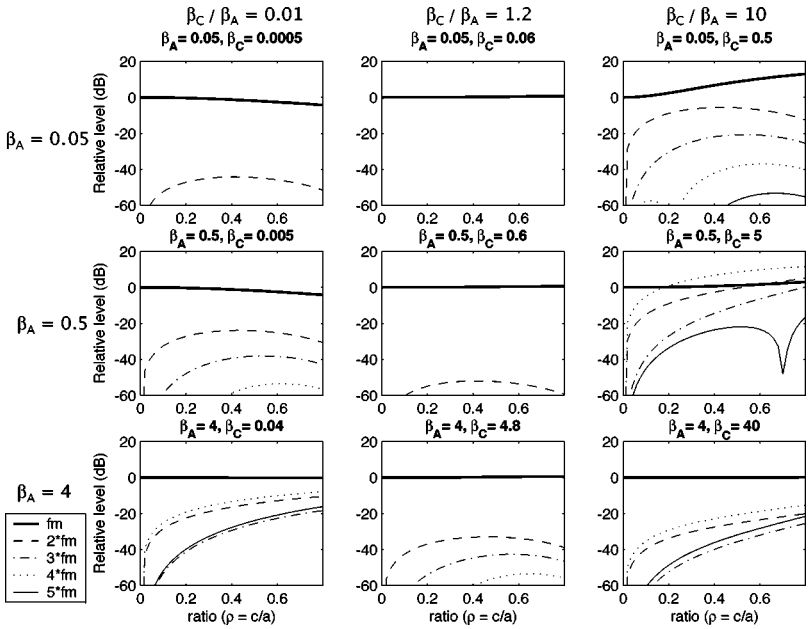


FIG. 15. As in Fig. 14 except calculations have been made for $\psi_0 = 90$ deg. The fundamental component is not affected appreciably except when the modulation index of the competing signal is much higher than that of the primary signal ($\beta_C/\beta_A = 10$). Distortion contains both even and odd harmonics.

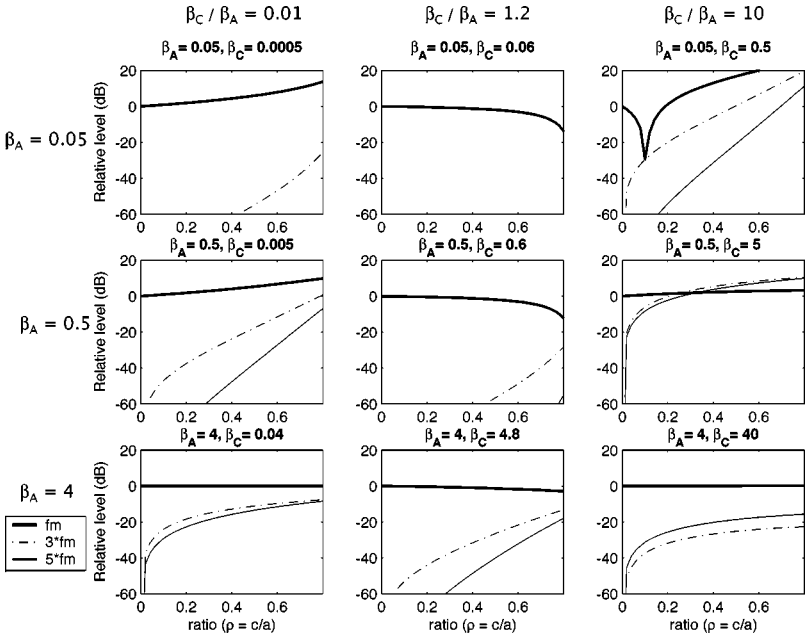


FIG. 16. As in Fig. 14 except calculations have been made for $\psi_0 = 180$ deg. The competing signal introduces the biggest changes for this phase.

(their Fig. 1). In order to address a full range of experimental conditions, we present guidelines that cover a large range of ρ values and three different β_C/β_A ratios.

In Figs. 14, 15, and 16 we show the effect of a competing signal as a function of ρ, and for three different β_C/β_A ratios (left: $\beta_C/\beta_A = 0.01$; middle, $\beta_C/\beta_A = 1.2$; and right $\beta_C/\beta_A = 10$). This generates three plots. These three plots are shown for three values of β_A (top: $\beta_A = 0.05$; middle: $\beta_A = 0.5$; and bottom $\beta_A = 4$). What is plotted is the change in output, $E(t)$, due to the competing signal, reported in dB. The harmonics were zero without competition, and what is plotted is their level relative to the fundamental (without competition). We have seen the strong effect of phase, ψ_0, on the results. Recall that the phase is not under experimental control; it depends on the distance between surfaces. Therefore, we present results for three phase values, 0 deg (Fig. 14), 90 deg (Fig. 15), and 180 deg (Fig. 16). Note that the even harmonics are absent when the phase is 0 deg or 180 deg. This is due to the symmetry of this condition. The 90 deg case shows the more general result, in which both even and odd harmonics are present.

The results from Figs. 14–16 have been distilled in Figs. 17–19, by presenting results in terms of nominally acceptable levels of error in the fundamental and distortion. The acceptable levels chosen were ± 3 dB for the fundamental response, and harmonic levels at least 30 dB down relative to the unchallenged fundamental. These ''threshold'' curves are shown for the same three values of β_A as in Figs. 14–16. On the left axis is ρ_{lim}, the ρ value for which acceptable levels are obtained. (To have a smaller error, ρ has to be less than ρ_{lim}.) If we examine Figs. 17–19 for a ratio of 0.1, we see that for low modulation index ($\beta_A = 0.05$) the errors are ac-

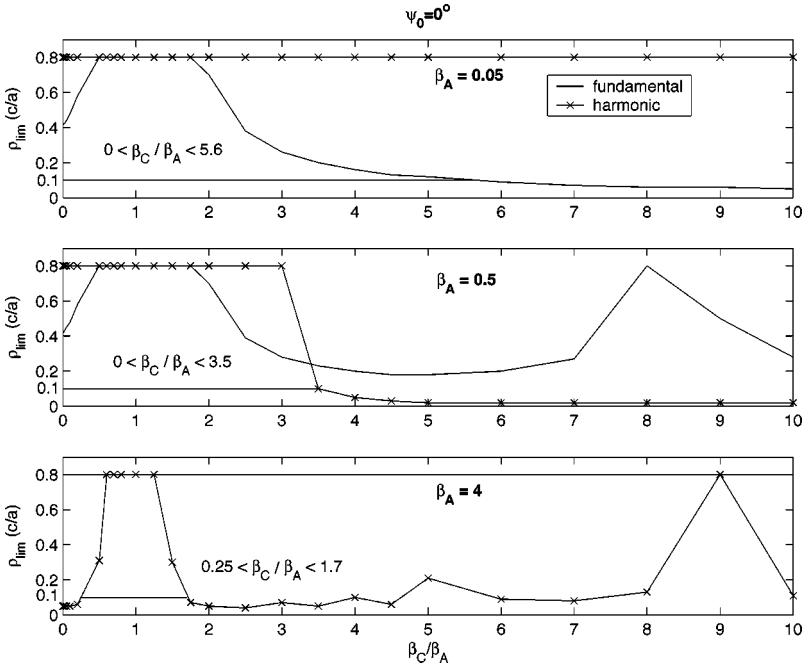


FIG. 17. Ratio (ρ_{lim}) for which the change in the fundamental is $+/-3$ dB and the level of the harmonic components is -30 dB *re*: fundamental. ρ_{lim} is shown as a function of β_C/β_A for three values of β_A (upper panel, $\beta_A = 0.05$; middle panel $\beta_A = 0.5$; and lower panel, $\beta_A = 4$). The phase is equal to $\psi_0 = 0$ deg. To have an error less than $+/-3$ dB for the fundamental or to have the harmonic component more than 30 dB down compared to the fundamental, ρ needs to be less than ρ_{lim}.

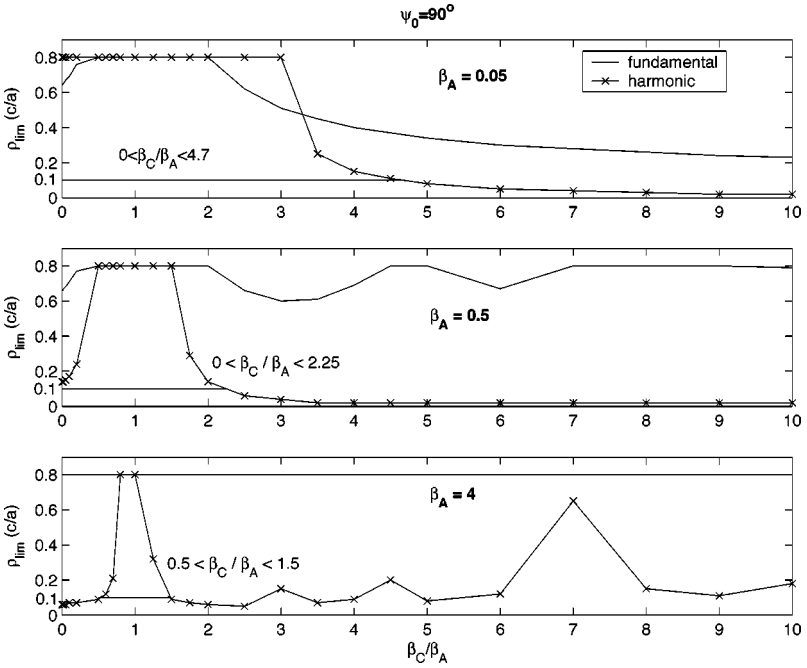


FIG. 18. As in Fig. 17 except $\psi_0 = 90$ deg.

ceptable as long as $\beta_C < 3.5\beta_A$, and for high modulation index ($\beta_A = 4$) the errors are acceptable if β_C is between $0.5\beta_A$ and $1.5\beta_A$, and again are acceptable for large values of β_C ($\beta_C > 7.5\beta_A$). Comparing Figs. 17–19, it is clear that errors in the fundamental component are much greater in the case of $\psi_0 = 180$ deg. Note that if two surfaces vibrate with an arbitrary ρ, with approximately the same amplitude ($\beta_C/\beta_A \sim 1$), and the distance between surfaces is such that $\psi_0 = 0$ deg, no errors or distortion are observed (Fig. 17). In most cases the phase will not be exactly 0 deg, and the phase, being related to distances between cochlear structures, is not under experimental control. However, the physiological basis of measured harmonics could be confirmed by their reproducibility. Comparing Figs. 14–16, the harmonic structure is very sensitive to the ψ_0 value. The ψ_0 value will vary from measurement to measurement, due to slightly different sites and distances. The distortion pattern, as well as the value of the fundamental, will change between the measurements if they are produced by competing signals from out-of-focus planes.

IV. CONCLUSIONS

One factor that governs the optical sectioning capability of an optical system is the objective lens's numerical aperture. Its effect can be described analytically and compared to the split-aperture objective developed by Koester *et al.* The optical sectioning is greatly improved by the split-aperture system.

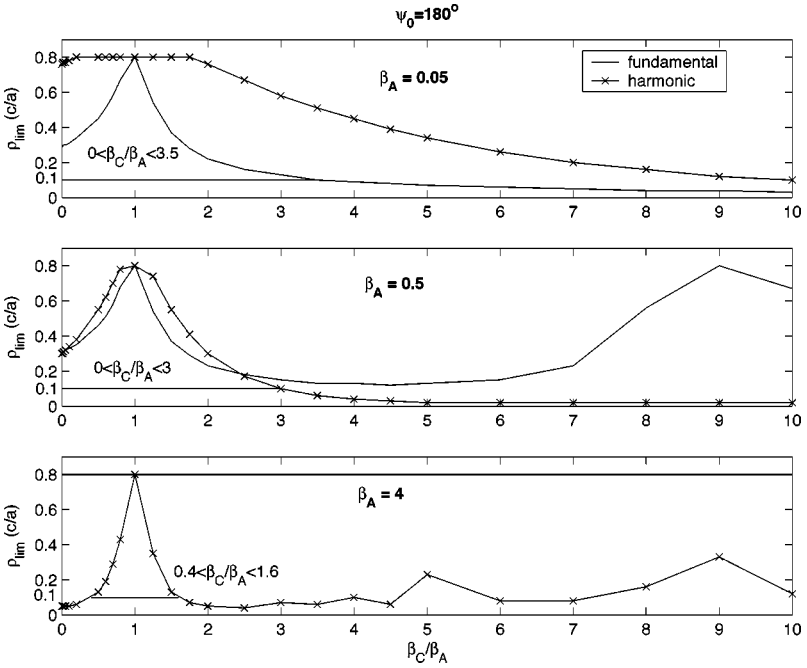


FIG. 19. As in Fig. 17 except $\psi_0 = 180$ deg.

We investigated the effect of a competing signal on the measured vibration. The investigation used both a purely theoretical treatment of FM demodulation [resulting in Eq. (5)] and a simulation of an FM receiver, as well as tests with our FM receiver using competing FM signals that were electronically generated, and finally tests with our entire interferometer/receiver system using optical signals that mimicked a true intracochlear measurement. The most basic result was that the theoretical expression Eq. (5) proved to be useful in predicting actual effects of competition. Beyond that, we found that the qualitative effect of a competing signal depended on the modulation index. When the modulation index was low (corresponding to displacements <25 nm) the competing signal altered the magnitude of the FM receiver output but did not affect its wave shape. When the modulation index was high (displacements>100 nm) the magnitude of the output at the stimulus frequency was resistant to the competing signal even when the ratio was large. However, distortion was present in the receiver output. For both low and high modulation indices the magnitude of the change depended on the phase difference between the two competing signals.

These findings point out that in cochlear vibration measurements made without beads, in order to be sure that the vibration of the selected structure in focus is measured it is important to reduce reflections from out-of-focus structures. A high numerical aperture objective lens attenuates the strength of the reflections from surfaces close to the focal plane, but this attenuation levels off with distance from the focal plane at a level that is not very restrictive. Other means provide attenuation at greater distances from the focal plane. The specialized confocal optics provided by the split-aperture interferometer developed by Koester *et al.* (1994) is one such system. Interferometers using incoherent light, such as that discussed in Dalhoff *et al.* (2001), also provide steep optical sectioning.

ACKNOWLEDGMENTS

This work was supported by the Emil Capita foundation and the NIDCD. Thanks to the reviewers for helpful comments on the manuscript, especially their suggestion that the modulation index be succinctly expressed as a function of displacement amplitude.

[1]When **A** and **C** move with different phases ($\phi_C \neq 0$), the angle φ in Eq. (5) is $\varphi=(\omega_{cC}-\omega_{cA})t+\psi_0-\beta_C\cos(\omega_m t+\phi_C)+\beta_A\cos(\omega_m t)$.

Baghdady, E.J. (**1956**). ''Interference rejection in FM receivers,'' M.I.T. Res. Lab. Electronics, Cambridge, MA, Tech. Rept., **252**, September 24.

Baghdady, E.J. (**1961**). *Lectures on Communication System Theory* (McGraw-Hill, New York), Chap. 19.

Cooper, N.P. (**1999a**). ''An improved heterodyne laser interferometer for use in studies of cochlear mechanics,'' J. Neurosci. Methods **88**, 93–102.

Cooper, N.P. (**1999b**). ''Vibration of beads placed on the basilar membrane in the basal turn of the cochlea,'' J. Acoust. Soc. Am. **106**, L59–L64.

Dalhoff, E., Gärtner, R., Zenner, H.P., Tiziani, H.J., and Gummer, A.W. (**2001**). ''Remarks about the depth resolution of heterodyne interferometers in cochlear investigations,'' J. Acoust. Soc. Am. **110**, 1725–1728.

Khanna, S.M., and Hao, L.F. (**2000**). ''Amplification in the apical turn of the cochlea with negative feedback,'' Hear. Res. **149**, 55–76.

Khanna, S.M., Ulfendahl, M., and Steele, C.R. (**1998**). ''Vibration of reflective beads placed on the basilar membrane,'' Hear. Res. **116**, 71–85.

Khanna, S.M., Koester, C.J., Willemin, J.-F., Dändliker, R., and Rosskothen, H. (**1996**). ''A noninvasive optical system for the study of the function of inner ear in living animals,'' SPIE **2732**, 64–81.

Koester, C.J., Khanna, S.M., Rosskothen, H.D., Tackaberry, R.B., and Ulfendahl, M. (**1994**). ''Confocal slit divided-aperture microscope: Applications in ear research,'' Appl. Opt. **33**, 702–708.

Middleton, D. (**1996**) *An Introduction to Statistical Communication Theory* (IEEE, New York), Chap. 15, pp. 636–642.

Middleton, D., and Spaulding, A.D. (**1981**) ''Signals and Interference in FM reception,'' NTIA Report: 81–87, U.S. Dept. of Commerce.

Nuttall, A.L., Dolan, D.F., and Avinash, G. (**1991**). ''Laser Doppler velocimetry of basilar membrane vibration,'' Hear. Res. **51**, 203–214.

Panter, P.F. (**1965**). *Modulation, Noise and Spectral Analysis: Applied to Information Transmission* (McGraw-Hill, New York).

Ren, T., and Nuttall, A.L. (**2001**). ''Recording depth of the heterodyne laser interferometer for cochlear vibration measurement,'' J. Acoust. Soc. Am. **109**, 826–829.

Robles, L., and Ruggero, M.A. (**2001**). ''Mechanics of the mammalian cochlea,'' Physiol. Rev. **81**, 1305–1350.

Ruggero, M.A., and Rich, N.C. (**1991**). ''Application of a commercially manufactured Doppler-shift laser velocimeter to the measurement of basilar membrane vibration,'' Hear. Res. **51**, 215–230.

Ulfendahl, M. (**1997**). ''Mechanical responses of the mammalian cochlea,'' Prog. Neurobiol. **53**, 331–380.

Willemin, J.F., Dändliker, R., and Khanna, S.M. (**1988**). ''Heterodyne interferometer for submicroscopic vibration measurements in the inner ear,'' J. Acoust. Soc. Am. **83**, 787–795.

Wilson, T. (Editor) (**1990**). *Confocal Microscopy* (Academic, London).

Integration and segregation in auditory scene analysis

Elyse S. Sussman[a)]
Department of Neuroscience, Albert Einstein College of Medicine, 1410 Pelham Parkway South, Bronx, New York 10461

(Received 12 January 2004; revised 29 November 2004; accepted 10 December 2004)

Assessment of the neural correlates of auditory scene analysis, using an index of sound change detection that does not require the listener to attend to the sounds [a component of event-related brain potentials called the mismatch negativity (MMN)], has previously demonstrated that segregation processes can occur without attention focused on the sounds and that within-stream contextual factors influence how sound elements are integrated and represented in auditory memory. The current study investigated the relationship between the segregation and integration processes when they were called upon to function together. The pattern of MMN results showed that the integration of sound elements within a sound stream occurred *after* the segregation of sounds into independent streams and, further, that the individual streams were subject to contextual effects. These results are consistent with a view of auditory processing that suggests that the auditory scene is rapidly organized into distinct streams and the integration of sequential elements to perceptual units takes place on the already formed streams. This would allow for the flexibility required to identify changing within-stream sound patterns, needed to appreciate music or comprehend speech. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1854312]

PACS numbers: 43.66.Ba, 43.66.Mk, 43.64.Bt [SEM] Pages: 1285–1298

I. INTRODUCTION

From infancy we experience a complex auditory environment made up of several simultaneously active sound sources that often overlap in many acoustic parameters. Yet we experience a coherent auditory environment made up of identifiable auditory events. Auditory scene analysis (ASA) involves the ability to disentangle the mixture of sound input, integrating sensory inputs that belong together and segregating those inputs that originate from different sources (Bregman, 1990). Accordingly, segregation and integration are two fundamental aspects of ASA. The focus of the current paper is on the interaction between these two important auditory processes in ASA.

A. Auditory memory

A crucial part of the auditory scene analysis process is the ability to connect elements over time (integration processes), allowing us to recognize a series of footsteps or to understand spoken speech. Transient memory plays a critical role in this process especially because auditory processing relies so heavily on the temporal domain. This is easily demonstrated by thinking about how we understand a spoken sentence. Once each word of a sentence is spoken (i.e., the physical source of auditory information is completed), only the neural trace of the physical sound information is retained. To get to the meaning of the sentence we have to access the memory of the previous words that were spoken.[1] Consider, however, that processing a speech stream most often occurs within the context of other simultaneous sound streams, whether on a city street, at the office, or in a department store. The focus of the current study is on understanding the relationship between the integration and segregation processes, specifically, to focus on the order in which the following processes are hypothesized to occur: (a) the segregation of input into distinct streams (perceived environmental sources) and then (b) the integration of elements necessary for making sense of the sequential patterns within a particular sound stream. In understanding how this memory operates in facilitating ASA, it is important to understand the relationship between these processes in terms of how auditory events are represented and stored in memory. Event-related brain potentials (ERPs) were used in the current investigation to explore the contents of transient auditory memory and to take advantage of the ability to use this method to observe the relationship between these processes when the listener has no task involving the sounds.

B. Event-related brain potentials

ERPs are noninvasive electrophysiological measures of cortical responses to sensory events that provide distinctive signatures for stimulus change. Change is an important cue for the auditory system. Because ERPs provide such high temporal resolution (in the order of milliseconds) and are time-locked to stimulus events, we can gain information about the timing of certain cognitive processes evoked by a given sound. Of particular importance for observing unattended sound processing is the mismatch negativity (MMN) component of ERPs (Näätänen, 1992). MMN reflects sound change detection and is elicited even when the sounds have no relevance to ongoing behavior. The change detection process underlying MMN generation is dependent upon auditory memory. It uses sensory representations of the acoustic regularities extracted from the sound sequence (often called the ''standard''). These sensory representations of the standard form the basis for the change detection process. Incoming

[a)]Electronic mail: esussman@aecom.yu.edu

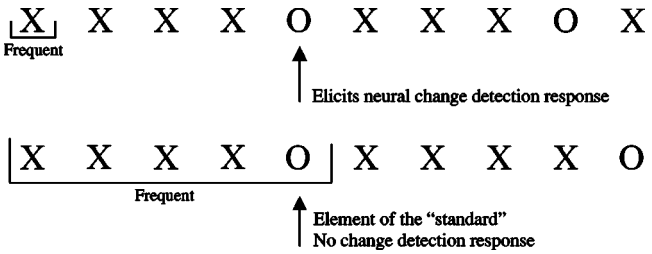


FIG. 1. *Contextual effects on MMN*. The top row displays a typical ''auditory oddball paradigm,'' with an ''oddball'' tone (represented by the letter ''O'') presented randomly (20%) amongst a frequent tone (represented by the letter ''X;'' 80%). The bottom row displays the same percentage of oddballs (O) together with the frequent tones (80%) except that the O tone is presented in a recurring pattern—it occurs every fifth tone. See text for further discussion.

sounds that deviate from the neural trace of the standard elicit MMN, which is generated in the auditory cortex and is usually evoked within 200 ms of change detection. Thus, MMN represents an early process of change detection based upon a memory of the previous sound stimulation. An indication that the mechanisms underlying the MMN system involve memory processes is found in evidence that MMN generation relies on current flow through active N-methyl-D-aspartate (NMDA) channels (Javitt *et al.*, 1996). Javitt *et al.* showed that blockage of NMDA, which has been associated with the cellular mechanisms involved in memory processes (Compte *et al.*, 2000; Kauer *et al.*, 1988; Pulvirenti, 1992), prevented MMN generation without disturbing elicitation of the prior obligatory cortical response to sound onsets. This suggests a link between working memory processes and the cortical auditory information-processing network involved in MMN elicitation.

The main neural generators of the MMN are located bilaterally in the supratemporal plane, as determined by dipole-modeling of electric (Scherg *et al.*, 1989; Ha *et al.*, 2003) and magnetic responses (Sams and Hari, 1991; Woldorff *et al.*, 1998), scalp-current density maps of scalp-recorded ERPs (Giard *et al.*, 1990), functional magnetic resonance imaging (fMRI; Opitz *et al.*, 1999), and intracortical ERP recordings (Halgren *et al.*, 1995; Kropotov *et al.*, 1995). The location of the generators in the auditory cortex accounts for the observed scalp topography of the waveform, which is maximally negative over the fronto-central scalp locations and inverts in polarity below the Sylvian fissure.

C. Assessing neural representations of sound organization

MMN can be used to probe the neural representations extracted from the ongoing sound input because the response to a particular sound is based upon the memory of the previous sounds. This is illustrated in Fig. 1. In the well-known auditory oddball paradigm, an ''oddball'' or infrequent sound is presented randomly among instances of a frequently repeating sound. The *oddball* elicits MMN because it is detected as deviating, in some sound feature (frequency, intensity, duration, spatial location), from the frequently repeating sound. In the top panel of Fig. 1, a regular auditory oddball paradigm is shown. The letter ''X'' represents a tone of one frequency and the letter ''O'' represents a tone of a different frequency. The oddball (**O**) would elicit MMN because it has a different frequency than that of the standard (X). In the bottom panel of Fig. 1, the same ratio of **O** to X sounds is presented, but instead of presenting the O tone randomly, it occurs every fifth tone in the sequence. *If* the brain detects the regularity, in this case the five-tone repeating pattern (X-X-X-X-**O**-X-X-X-X-**O**$\cdots$), then no MMN will be elicited by the **O** tone (the ''oddball'' in the randomized context) because the **O** tone is part of the standard repeating (five-tone) regularity. It is not deviant. Thus, it is important to note that MMN is not *ipso facto* elicited when there is a situation of a frequent and an infrequent tone presented in a sound sequence. What is crucial to its elicitation is what is detected as the regularity in the sound sequence and stored in auditory memory. Thus, MMN is highly dependent upon the context of the stimuli, whether that context is detected without attention focused on the sounds (Sussman *et al.*, 1999a, 1998a) or whether the context is influenced by attentional control (Sussman *et al.*, 1998b; 2002a).

Thus, the change detection process is far more complex than is superficially evident from the auditory oddball paradigm, the most common paradigm used to elicit MMN. Simple sensory discrimination of two tones (i.e., being able to detect whether one tone is different from another in any sound feature) requires only a single presentation of each sound. Repetition is not necessary for this type of discrimination. However, the MMN process is based upon *repetition* and the detection of the regularities in the acoustic input. Even with the seemingly simple auditory oddball paradigm, deviance detection is based upon the memory of the larger context of the sounds and not simply on the detection of two different tones in the sound sequence. Specifically because MMN relies on the larger auditory context (sequential processing of auditory events), it can be used to assess how the acoustic information is structured and stored in memory.

D. Segregation of sounds to create distinct sources

Segregation of sounds to distinct auditory sources is a key function of the auditory system. Consider that the acoustic information entering one's ears is a mixture of all the sounds in the environment, without separation. Decomposing the auditory input is therefore a crucial step, allowing us to detect a single voice in a crowd or to distinguish a voice coming from the left or right in the proverbial ''cocktail party'' setting. For that reason, the process of disentangling the sound into distinct sources plays a critical role in how we experience the auditory environment. Although there remains controversy over the role of attention in the stream segregation process (Botte *et al.*, 1997; Bregman, 1990; Brochard *et al.*, 1999; Carlyon *et al.*, 2001; Jones *et al.*, 1978; 1999; Macken *et al.*, 2003; Sussman *et al.*, 1999a; Winkler *et al.*, 2003), there is considerable ERP evidence to suggest that the segregation of auditory input to distinct sound streams can occur without attention focused on the sounds, and that auditory memory can hold information about multiple sound streams independently (Sussman *et al.*, in press, submitted, 1999a; Ritter *et al.*, 2000; Winkler *et al.*, 2003). This basic finding has been replicated in various paradigms using dif-

ferent segregation cues, such as frequency proximity (Sussman *et al.*, 2001; Winkler *et al.*, 2003), temporal proximity (Sussman *et al.*, 1999a), and a combination of spatial location, frequency, and intensity (Ritter *et al.*, 2000). These studies provide evidence in humans that auditory stream segregation mechanisms are part of an early, primitive process. This is consistent with evidence from animal studies demonstrating that the basic stream segregation mechanisms are part of vertebrates' auditory systems (Fay, 2000; Hulse *et al.*, 1997; Fishman *et al.*, 2001). It is functionally parsimonious that the basic stream segregation process would occur automatically (calculated on the basis of the acoustic characteristics of the sensory input), so that attention, a limited human resource, could be used to identify patterns within the already organized streams, needed for understanding speech or listening to music.

E. Integration of sound elements within a single stream

Auditory integrative processes have predominantly focused on how successive elements influence a single percept of sound, such as with the phenomena of loudness summation (Zwislowski, 1969) and auditory recognition masking (Massaro, 1975). The perception of a sound event is often determined by the sounds that surround it. Recently, however, it was demonstrated that the larger context of a sequence of sounds could affect processing of the individual sound components even when not in direct temporal proximity (Sussman *et al.*, 2002b; Sussman and Winkler, 2001). It is this context-dependent integration process that is the focus of the current study and described in detail below.

The paradigm of the current study was based on previous work (Sussman *et al.*, 1999b, 2002b; Sussman and Winkler, 2001) showing that two different deviations from the same repetitive standard, occurring successively within a less than 200-ms interval (called a ''double deviant''), elicit only one MMN. However, when double deviants were presented within a stimulus sequence that also contained single deviants (i.e., sounds that differ from the repetitive standard in one of the two ways of the double deviants), the double deviants elicited two discrete MMNs, which were separated by the temporal distance of the successive deviations (a schematic of the basic paradigm used in these studies is shown in Fig. 2). Initially, we hypothesized that the successive deviants were integrated into a single deviant event because they were indistinguishable as discrete events when subjects ignored the sounds. This led to a prediction that if subjects could *actively* discriminate the two successive deviants in the blocked context, then the double deviants would elicit two MMNs. However, this turned out not to be the case. We found that even when subjects were trained to discriminate the two successive deviants, only one MMN was still elicited by them (Sussman *et al.*, 2002b). The integration phenomenon occurred solely as a function of changing the stimulus-driven context (i.e., blocked versus mixed) irrespective of whether attention was focused on the sounds.

The results of these studies are consistent with a context-based interpretation of auditory event formation in that the relevant information influencing event formation can be calculated over a longer time. In the blocked context, the second deviant is fully predicted by the first. Every time a deviant occurs another one follows it. Thus, the second of the successive deviants does not provide new information about ''deviancy.'' In contrast, when the double deviants are mixed with single deviants in the sequence, in the mixed context, the second successive deviant is not fully predicted; it carries new information such that the second deviant distinguishes the two types of deviants presented in the block. The presence or absence of single deviants in the sound sequence creates different contexts for the evaluation of the double deviants and, therefore, the double deviants are processed either as unitary events or as two successive events depending on which context they occur in. Additionally, the MMN process appears to be predominantly influenced by the context of a sequence of sounds in a longer-term period than previously thought [estimated to be within 20 s in Sussman and Winkler (2001)]. This global context effect on MMN has also been shown in Sussman *et al.* (2003).

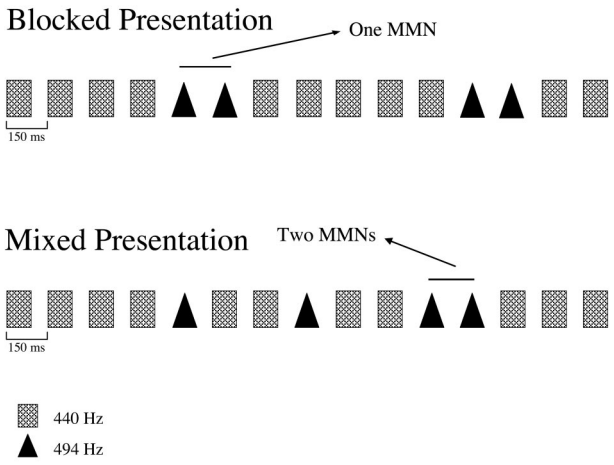


FIG. 2. *A schematic representation showing the blocked and mixed sequences* used in Sussman *et al.* (1999b, 2002b). In the blocked presentation, every time a deviant occurs another deviant follows it (termed ''double deviants''). In the mixed presentation, both single and double deviants occur randomly in the sequence. See text for further discussion.

These data, individually assessing the neural correlates of segregation and integration in ASA, suggest that (1) segregation can occur without attention focused on the sounds and that (2) within-stream contextual factors influence how events are represented in auditory memory.

In the current study, three experiments were conducted to investigate the relationship between the segregation and integration processes when they were called upon to function together—segregation of sounds, by spectral (experiments 1 and 2) or nonspectral (experiment 3) cues, and integration of double deviants within a stream. The blocked and mixed condition paradigms of Sussman *et al.* (2002b; see Fig. 2) were used within an alternating sequence of high- and low-frequency sounds (experiments 1 and 2) and within an alternating sequence of bandpass-filtered noise and harmonic complexes (experiment 3). The sounds were presented at a fast rate that would induce stream segregation and the integration paradigm was imposed within the low-frequency sound stream (experiment 1), within both high- and low-

frequency streams (experiment 2), and within the complex tone stream (experiment 3). I hypothesized that the acoustic characteristics of the input (the stimulus-driven cues) would be used to separate sounds into distinct sources *prior to* the integration of elements and the formation of sound events. If this were true, then the context of the within-stream sounds should influence the response to the double deviants such that one MMN would be elicited in the blocked within-stream context and two MMNs should be elicited in the mixed within-stream context. If integration operates at a global level, then this pattern of results should not be observed because the intervening high tones between the double deviants should interfere with the integration process (due to the alternation of the high and low tones—tested in the control conditions of experiment 1).

II. EXPERIMENT 1. SEGREGATION AND INTEGRATION (LOW STREAM)

Four conditions were presented in experiment 1, in which the context was manipulated within the low tones stream. In one condition the low stream of sounds had only double deviants (*alternating-blocked*) and in another condition double deviants were presented along with single deviants (*alternating-mixed*). Two control conditions were used to ascertain the effect of an intervening sound within a single stream (*low only—standard intervening condition*) and to determine whether the integration of elements was initiated by a time constant or by the number of elements occurring successively (*low only—triple deviant condition*).

A. Participants

Thirteen adults (five males, eight females) between the ages of 23 and 42 years, with no reported hearing loss or neurological disorders, were paid to participate in the study. All subjects passed a hearing screening. Participants were informed of the experimental protocol before they signed a consent form. The study was conducted at the Albert Einstein College of Medicine.

B. Stimuli

Three tones (one high frequency and two low frequency) were used to create the stimulus sequences. They were 50 ms in duration with an intensity of 80 dB SPL (calibrated with Brüel & Kjær 2209 sound-level meter). The high-frequency tone was 1568 Hz (H) and the low-frequency tones were 440 Hz (L) and 494 Hz (LD). Onset and offset ramps of 7.5 ms were created using a Hanning window for all tones. Stimuli were presented in four conditions: (*a*) *alternating-blocked*, (*b*) *alternating-mixed*, (*c*) *low only—standard intervening*, *and* (*d*) *low only—triple deviant*. In the two alternating conditions, H and L alternated in a regular fashion at a rate of 75 ms SOA (onset to onset), except when the deviant (LD) occurred. Thus, a low tone (or a high tone) occurred once every 150 ms. In the *alternating blocked condition*, the deviant (LD)always occurred twice in succession in the low stream (e.g., H L H L H L H **LD** H **LD** H L H L); these will be called ''double deviants'' (see Fig. 3). Double deviants occurred randomly 7.5% of the time. In the *alternating mixed condition*, LD was randomly presented both singly (these will be called ''single deviants'') 5% and twice in succession 5% each (e.g., H L H L H L H **LD** H **LD** H L H L H **LD** H L; see Fig. 3). LD was presented 15% of the time overall in both conditions. Thus, the ''blocked'' and ''mixed'' nomenclature refers to the low stream only, specifically to the deviants and whether there are only double deviants in the sequence (blocked) or whether there are mixed single deviants with double deviants (mixed) in the same sequence.

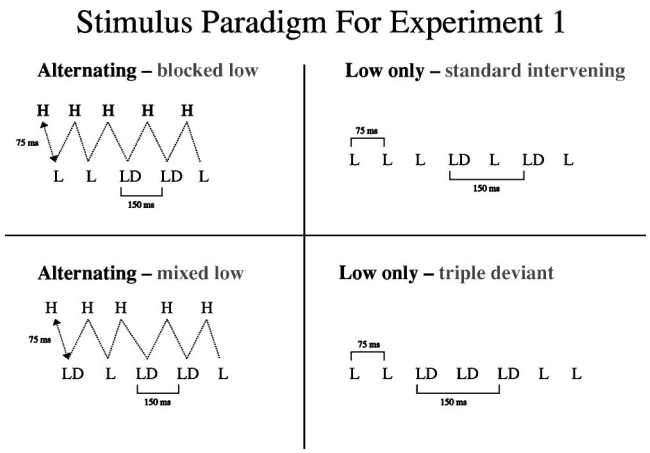


FIG. 3. *A schematic diagram of the stimulus paradigm used in experiment 1*. The alternating pattern of low (L) and high (H) tones that was used to create the alternating conditions is shown in the left column. ''L'' represents a tone of one frequency and the ''LD'' represents a tone of a slightly higher pitch than the L tones. The ''H'' represents tones of much higher frequency than the L and LD tones. The critical variable across conditions is the response to the ''double deviants,'' the two successive ''LD'' tones occurring in the low stream of the alternating conditions. Two control conditions are displayed in the right column. The low-only conditions were presented using the same frequency tones as in the alternating conditions; however, no H tones were presented in these conditions.

C. Procedures

Participants sat in a comfortable chair approximately 1 m from a TV monitor in a sound-attenuated booth. They were instructed to watch a closed-captioned video, itself containing no sounds, and to ignore the sounds. Three thousand stimuli were presented in each of the four conditions ($3000 \times 4 = 12\,000$ stimuli in all). There were two runs of 1500 stimuli presented, totaling eight runs. The eight runs were presented randomly, counterbalanced across subjects to avoid any possible order effects.

D. Electroencephalogram (EEG) recording and data analysis

EEG was recorded with an electrode cap using the following scalp locations (10–20 system): Fz, Cz, Pz, F3, F4, C3, C4, and both mastoids (LM and RM, for the left and right mastoids, respectively). Horizontal eye movements were measured by recording the horizontal electro-oculogram (HEOG) between electrodes F7 and F8 using a bipolar montage and vertical using a bipolar montage between FP1 and an external electrode placed beneath the left eye (VEOG). The common reference electrode was attached to the tip of the nose. Eye movement was monitored with VEOG and HEOG. HEOG was monitored for eye saccade

movement to ensure that participants were reading the captions on the video. The EEG was digitized (Nicolet amplifiers) at a sampling rate of 250 Hz (bandpass 0.05–100 Hz) and then off-line filtered between 1 and 30 Hz. Epochs of 600 ms, with a 100-ms prestimulus and 500-ms poststimulus period, were averaged separately for the standards and double deviants. The double deviants (and triple deviant) were averaged from the onset of the first of the successive deviants, thus the response to the successive deviants could be observed within this period. Epochs with an electrical change exceeding 100 μV were excluded from further averaging to remove the trials contaminated by artifacts of noncortical origin.

For each participant, the artifact-free epochs were then averaged separately for the four conditions and two stimulus types (standard and deviant). The mean amplitude in the 100-ms prestimulus period was subtracted from each point of the averaged ERP responses, which served as the reference from which the evoked responses were measured.

The difference waveforms were obtained by subtracting the ERP response to the standard (e.g., L-H-L in the alternating condition) from the ERP response to the deviants. In the mixed conditions, double deviants (LD-H-LD) were averaged separately from single deviants (LD-H-H). The intervals used for statistically evaluating for the presence of the potential MMNs were determined from the group mean deviant-minus-standard waveforms at Fz (the site of greatest signal to noise ratio), taking a 40-ms window centered on the peak of the MMNs from the conditions in which two successive MMNs were elicited (the *low only—standard intervening* and *alternating-mixed* conditions). The mean amplitude obtained in the intervals used for the two low-only conditions (range 1: 116–156 ms and range 2: 304–344 ms) and for the two alternating conditions (range 1: 140–180 ms and range 2: 320–360 ms) were measured for each subject individually and were then used to verify the presence of MMN, for each range and condition separately with one-sample, one-tailed *t*-tests using the data from the Fz electrode site. Analysis of variance (ANOVA) for repeated measures with factors of condition (blocked versus mixed) and range (range 1 versus range 2) was used to compare the mean voltage. Greenhouse–Geisser correction was reported when appropriate. Tukey HSD *posthoc* tests were calculated.

E. Results and discussion

The ERP results for the low-only and alternating conditions are displayed in Figs. 4 and 5, respectively (also see Table I).

1. Low-only conditions

In the *standard-intervening condition*, the two deviants elicited separate MMNs ($t_{12}=6.14$, $p<0.01$, $t_{12}=2.11$, $p<0.03$; ranges 1 and 2, respectively) approximately 150 ms from each other (see Fig. 4, top). This demonstrates that the two deviants were processed as separate events when there was just one standard tone intervening between them within the single stream. In the *triple-deviant condition*, the triple-deviant elicited only one MMN ($t_{12}=6.87$, $p<0.01$, $t_{12}<1$, $p>0.72$; ranges 1 and 2, respectively; see Fig. 4, bottom),

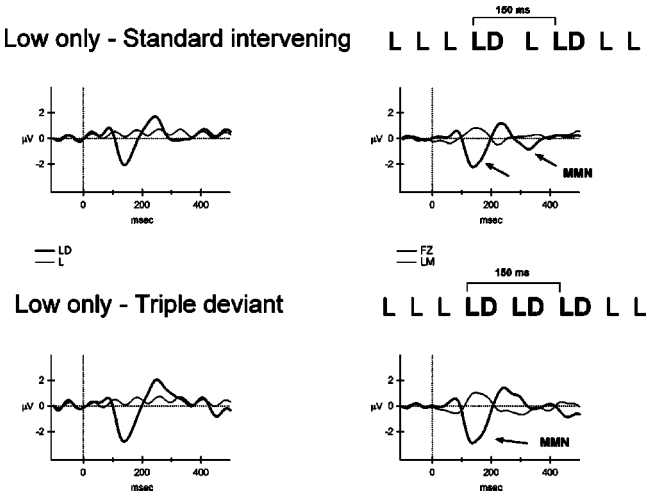


FIG. 4. *Experiment 1—Low tones only conditions*. Grand mean event-related brains potentials elicited by the L (thin line) and LD (thick line) tones in the low-only conditions are displayed in the left column. The deviant-minus-standard difference curves are displayed in the right column with Fz (thick line) overlaid with the left mastoid (LM; thin line). Arrows point to significant MMNs. These two traces (Fz and LM) provide the recording sites with the largest signal-to-noise ratio for the MMN component as well as indicating the topographical information needed to demonstrate that the source of the response was generated in the auditory cortex (see the Introduction). The response to the successive deviants can be observed within the epoch.

suggesting that event formation within the integrating period operates independently of the number of elements falling within the window. The single presentation of a standard tone between the two deviants (e.g., LD-L-LD) resulted in the deviants being processed as two separate events in the standard-intervening condition, compared to the triple-deviant condition in which all three of the elements presented in the same time window were processed as a single deviant event (LD-LD-LD). This provides evidence that event formation operates on the basis of the *type of elements* falling within the integrating period, not simply integrating

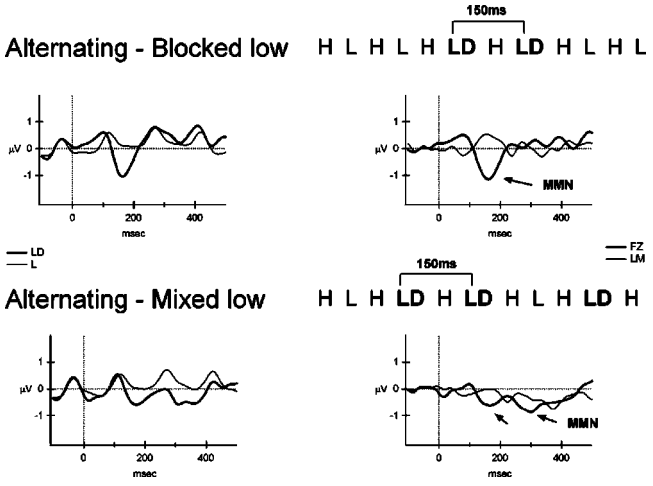


FIG. 5. *Experiment 1—Low tones in the alternating conditions*. Grand mean event-related brains potentials elicited by the L (thin line) and LD (thick line) tones in the alternating conditions are displayed in the left column. The deviant-minus-standard difference curves are displayed in the right column with Fz (thick line) overlaid with the left mastoid (LM; thin line). Arrows point to significant MMNs.

TABLE I. Experiment 1. Mean amplitude values (in μV) and standard deviations (in parentheses) of the grand-averaged deviant-minus-standard (difference) waveforms in ranges 1 and 2 for all conditions at all electrode sites.

	Condition							
	Low only—Standard intervening		Low only—Triple deviant		Alternating-Blocked		Alternating-Mixed	
Electrode	Range 1	Range 2	Range 1	Range 2	Range 1	Range 2	Range 1	Range 2
Fz	−1.90	−0.72	−2.59	0.10	−0.77	0.05	−0.75	−0.82
	(1.36)	(1.05)	(1.12)	(1.22)	(1.47)	(1.20)	(0.99)	(0.58)
Cz	−1.46	−0.50	−1.95	−0.19	−0.62	0.09	−0.54	−0.91
	(1.15)	(1.08)	(0.84)	(1.04)	(1.20)	(0.90)	(0.84)	(0.59)
Pz	−0.66	−0.12	−0.87	−0.32	−0.26	0.08	−0.38	−0.82
	(0.92)	(0.95)	(0.58)	(0.78)	(0.99)	(0.87)	(0.82)	(0.67)
F3	−1.54	−0.58	−2.21	0.05	−0.60	0.19	−0.53	−0.80
	(1.22)	(0.93)	(0.95)	(1.03)	(1.34)	(1.18)	(0.90)	(0.57)
F4	−1.83	−0.51	−2.46	0.21	−0.85	0.08	−0.85	−0.85
	(1.39)	(1.09)	(1.18)	(1.28)	(1.51)	(1.06)	(1.07)	(0.70)
C3	−1.10	−0.27	−1.45	−0.11	−0.34	0.25	−0.43	−0.79
	(1.05)	(0.85)	(0.65)	(0.88)	(1.20)	(0.89)	(0.83)	(0.51)
C4	−1.33	−0.23	−1.60	0.07	−0.57	0.14	−0.60	−0.84
	(1.21)	(1.05)	(0.89)	(0.98)	(1.07)	(0.81)	(0.77)	(0.60)
LM	0.66	0.19	0.76	−0.52	0.58	0.02	−0.03	−0.43
	(0.60)	(0.69)	(0.76)	(0.65)	(0.62)	(0.64)	(0.54)	(0.61)
RM	0.85	0.32	1.26	−0.30	0.65	0.13	−0.01	−0.33
	(0.86)	(0.82)	(0.60)	(0.45)	(0.56)	(0.58)	(0.59)	(0.74)

all elements by a time constant. The integration or segregation of successive elements within the low stream was governed by the nature of the information: elements carrying the *same* information were integrated into a single auditory event, whereas elements that *differed* within that same time period were distinguished from each other. This suggests that the underlying time constant that tends to integrate information operates on the basis of the distinctiveness of the information in the larger context.

2. Alternating conditions

In the *blocked condition*, double deviants in the low stream elicited one MMN ($t_{12}=2.52$, $p<0.02$, $t_{12}<1$, $p>0.50$; ranges 1 and 2, respectively; see Fig. 5, top row). The double deviants were integrated into a single deviant event even though there was a tone intervening between them. This indicates that the streams were segregated prior to the integration process. In contrast, in the *mixed condition*, the same double deviant events elicited two MMNs ($t_{12}=2.33$, $p<0.02$, $t_{12}=5.05$, $p<0.01$; ranges 1 and 2, respectively; Fig. 5, bottom row) when single deviants also occurred within the low stream. ANOVA comparing the mean voltages of the difference waveforms in ranges 1 and 2 revealed an interaction between condition (blocked versus mixed) and range [$F(1,12)=6.48, p<0.03$]. *Posthoc* analyses showed that this interaction was due to larger mean amplitude for range 1 than range 2 in the blocked condition and no difference between mean amplitude in ranges 1 and 2 of the mixed condition. This confirms that one MMN was elicited in the blocked condition and two MMNs were elicited in the mixed condition. This result demonstrates that the presence of the single deviants in the low stream altered the MMN response to the double deviants in the same way as occurs when no intervening high tones are presented in the stimulus block [i.e., when the low stream is presented alone at the same stimulus rate (e.g., Sussman *et al.*, 2002b)].

The double deviants in the *alternating-blocked condition* elicited one MMN but they elicited two MMNs in the *alternating-mixed condition*. Double deviants in both *alternating* conditions should have elicited two MMNs if the successive deviant-standard-deviant pattern had been processed as a single stream, as demonstrated by results of the *low-only* conditions. However, the presence of the single deviants within the low stream of the mixed condition altered the MMN response to the double deviants. This difference in the MMN response to the same double deviants for the blocked and mixed conditions demonstrates that the response was dependent upon the context of the other tones within the low stream, exactly as when the low stream is presented alone (e.g., Sussman *et al.*, 2002a, b). That is, this contextual effect occurred despite the fact that there were intervening high tones between the low sounds. Thus, the results indicate that the segregation into distinct frequency streams occurred *prior* to the integration of elements within streams.

III. EXPERIMENT 2. SEGREGATION AND INTEGRATION IN BOTH HIGH AND LOW STREAMS

A. Rationale

The within-stream contextual effects observed in experiment 1 occurred within only one of the two possible streams. This leaves open the question of whether integration effects occurred within-stream because there was only one stream with changing information. Would the complexity (having

deviants in both streams) created by simultaneously changing contexts within the high and low streams modify processing of the sounds? If complexity could negate the integration effects mediated by changing the contextual information, then a global effect on sound processing could account for the results and could not be excluded as an interpretation of the results of experiment 1. If, however, the same pattern of integration effects were found for double deviants occurring within both the low and high streams (double deviants eliciting one MMN in blocked and two MMNs in mixed contexts) then the conclusions of experiment 1 would be expanded and strengthened. To test this, the same basic paradigm of experiment 1 was used except that deviants occurred in both streams.

B. Participants

Twelve adults (five males, seven females) between the ages of 22 and 37 years, with no reported hearing loss or neurological disorders, were paid to participate in the study. All subjects passed a hearing screening. Participants were informed of the experimental protocol before they signed a consent form. The study was conducted at the Albert Einstein College of Medicine. Two subjects' data were removed due to excessive eye artifact. One subject's data were removed due to experimental error. The data from the remaining nine subjects are reported. None of the subjects that participated in experiment 2 participated in experiment 1.

C. Stimuli

Four tones (two high frequency and two low frequency), 50 ms in duration (tone intensity was calibrated with Brüel & Kjær 2209 sound-level meter at a level of 80 dB SPL), were used to create the stimulus sequences. The high-frequency tones were 1568 Hz (H) and 1760 Hz (HD) and the low-frequency tones were 440 Hz (L) and 494 Hz (LD). Onset and offset ramps of 7.5 ms were created with a Hanning window for all tones. Stimuli were presented in two conditions: (a) *alternating-blocked* and (b) *alternating-mixed*. H and L alternated in a regular fashion at a rate of 75 ms SOA (onset-to-onset). Thus, a low tone (or a high tone) occurred once every 150 ms. In the *alternating-blocked* conditions, double deviants (LD or HD) always occurred twice in succession (presented randomly 6% of the time) in their respective stream but deviants never overlapped across streams. In the *alternating-mixed conditions*, HD and LD were presented randomly as single deviants (4%) and twice in succession (4%) each. Thus, HD and LD were presented 12% overall in both conditions. The ''blocked'' and ''mixed'' nomenclature refers to the presentation of the deviants in both the high- and low-frequency streams.

D. Procedures

Participants sat in a comfortable chair approximately 1 m from a TV monitor in a sound-attenuated booth. They were instructed to watch a closed-captioned video, itself containing no sounds, and to pay no attention to the sounds. Six thousand stimuli were presented in the blocked condition in two runs of 3000 stimuli, and 15 000 stimuli in five runs of 3000 stimuli were presented in the alternating condition, totaling seven runs. The seven runs were presented randomly, counterbalanced across subjects to avoid any possible order effects.

E. Electroencephalogram (EEG) recording and data analysis

Recording parameters matched those of experiment 1 with the following exceptions. The EEG was digitized (Neuroscan Synamps amplifiers) at a sampling rate of 500 Hz (bandpass 0.05–100 Hz) and then off-line filtered between 1 and 15 Hz. Epochs with an electrical change from baseline exceeding $\pm 75\ \mu$V were excluded from further averaging.

For each participant, the artifact-free epochs were then averaged separately for two conditions (blocked and mixed), two frequency ranges (high and low), and two stimulus types (standards and double deviants).

The difference waveforms were obtained by subtracting the ERP response to the standard from the ERP response to the deviant. The intervals used for statistically evaluating for the presence of the potential MMNs were determined from the group mean deviant-minus-standard waveforms at Fz (the site of greatest signal-to-noise ratio), taking a 40-ms window centered on the peak of the MMNs for range 1 and range 2, which can be observed in the epoch. Range 2 intervals for the blocked conditions were obtained from the corresponding mixed conditions, in which two MMNs were elicited. The mean amplitude obtained in the intervals used for the blocked—high condition was range 1: 106–146 ms and range 2: 270–310 ms; for the blocked—low condition was Range 1: 126–166 ms and range 2: 304–344 ms; for the mixed—high condition was range 1: 100–140 ms and range 2: 270–310 ms; and for the mixed–low condition was range 1: 126–166 ms and range 2: 304–344 ms. The intervals were measured for each subject individually and were then used to verify the presence of MMN. Because the predictions for the MMNs were *a priori*, one-sample one-tailed Student's *t*-tests were used to determine whether the mean amplitudes obtained in these latency ranges were significantly greater than zero, separately for each range and condition. ANOVA for repeated measures with factors of condition (blocked versus mixed), frequency (high versus low) and range (range 1 versus range 2) was used to compare mean voltage. Greenhouse–Geisser correction was reported when appropriate. Tukey HSD *posthoc* tests were calculated.

F. Results and discussion

The ERP results for the *alternating-blocked* and *alternating-mixed* conditions are displayed in Figs. 6 and 7, respectively (also see Table II).

1. Alternating-blocked condition

Double deviants elicited one MMN in the low stream ($t_9=5.1$, $p<0.01$, $t_9<1$, $p>0.43$; ranges 1 and 2, respectively; see Fig. 6, top row) and one MMN in the high stream ($t_9=2.0$, $p<0.05$, $t_9=1.0$, $p>0.16$; ranges 1 and 2, respectively; see Fig. 7, top rows). Consistent with the results of experiment 1, blocked condition double deviants elicited one MMN.

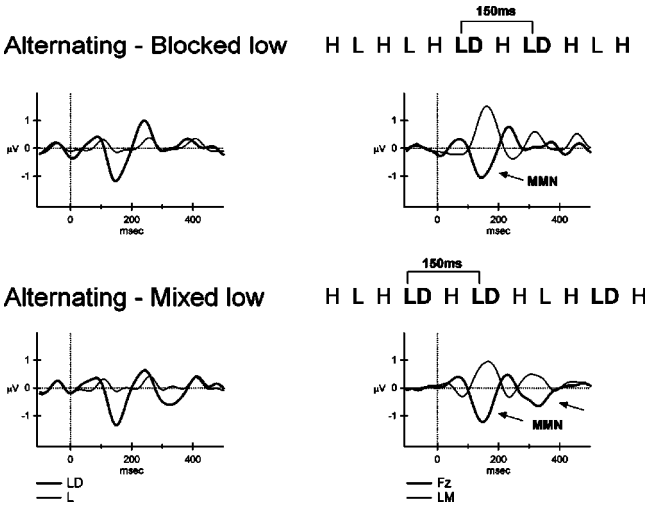


FIG. 6. *Experiment 2—Low tones in the alternating conditions*. Grand mean event-related brains potentials elicited by the L (thin line) and LD (thick line) tones in the blocked conditions are displayed in the left column. The deviant-minus-standard difference curves are displayed in the right column with Fz (thick line) overlaid with the left mastoid (LM; thin line). Arrows point to significant MMNs.

2. Alternating-mixed condition

When single deviants also occurred within respective streams, double deviants elicited two MMNs in the low stream ($t_9=4.8$, $p<0.01$, $t_9=4.8$, $p<0.01$; ranges 1 and 2, respectively; see Fig. 6, bottom row) and two MMNs in the high stream ($t_9=1.8$, $p=0.05$, $t_9=2.0$, $p<0.05$; ranges 1 and 2, respectively; see Fig. 7, bottom row). The ANOVA comparing the mean voltage revealed an interaction of condition and range [$F(1,8)=6.37, p<0.04$] and no interaction with frequency. *Post hoc* analyses showed that the mean voltage of range 1 was larger than that in range 2 for the blocked conditions but the mean voltages of the two ranges were not significantly different from each other in the mixed conditions. The presence of the single deviants in the mixed

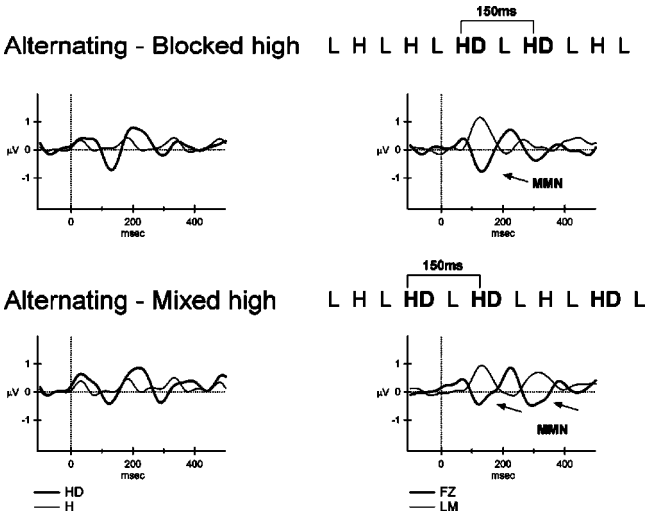


FIG. 7. *Experiment 2—High tones in the alternating conditions*. Grand mean event-related brains potentials elicited by the H (thin line) and HD (thick line) tones in the alternating conditions are displayed in the left column. The deviant-minus-standard difference curves are displayed in the right column with Fz (thick line) overlaid with the left mastoid (LM; thin line). Arrows point to significant MMNs.

condition altered the MMN response to the double deviants, as compared to the blocked conditions, independently in each stream.

The results of experiment 2 rule out complexity as an explanation for the within-stream integration effects found in experiment 1, and extend the results of experiment 1 to show that integration processes are carried out concurrently on multiple streams of auditory information.

IV. EXPERIMENT 3. SEGREGATION BY NONSPECTRAL CUE

A. Rationale

Experiments 1 and 2 used pure tone stimuli with a large (15 semitone) difference between high and low sounds. It is therefore possible to consider that the integration effects may have taken place by frequency range, and were not mediated by streaming. An alternative interpretation would be that with no overlapping excitation of hair cells within the cochlea, frequency effects rather than segregation effects may explain the integration effects of experiments 1 and 2. This interpretation of the current results would characterize frequency as the dominant factor influencing the integration processes, independent of segregation. Thus, another experiment was conducted to assess the integration effect when stream segregation was cued by a factor other than spectral information. The same basic paradigm of experiment 1 was used, except that the alternating sequence of sounds included complex tones and narrowband noise bursts with overlapping spectral composition. The noise and tones were expected to segregate to separate streams (Singh and Bregman, 1997; Iverson, 1995; Cusack and Roberts, 1999, 2000). If the integration were a frequency-specific phenomenon in the previous two experiments, that is, if the integration process did not depend on the initial segregation of the information, then two MMNs should be elicited in both blocked and mixed contexts, because of the effect of the intervening tones (see experiment 1, discussion of low-only control conditions). This result may be consistent with a peripheral channeling explanation (Beauvois and Meddis, 1991; Hartmann and Johnson, 1991; Rose and Moore, 2000). On the other hand, if a similar pattern of MMN results occurs as did in experiments 1 and 2, then it could be concluded that the segregation process preceded the integration process. Or, stated differently, we can conclude that the integration of sequential sound elements takes place within the distinct sound streams.

B. Participants

Twelve adults (five males, seven females) between the ages of 22 and 37 years, with no reported hearing loss or neurological disorders, were paid to participate in the study. All subjects passed a hearing screening. Participants were informed of the experimental protocol before they signed a consent form. The study was conducted at the Albert Einstein College of Medicine. Two subjects' data were removed due to excessive eye artifact. One subject's data were removed due to experimental error. The data from the remaining nine

TABLE II. Experiment 2. Mean amplitude values (in μV) and standard deviations (in parentheses) of the grand-averaged deviant-minus-standard (difference) waveforms in ranges 1 and 2 for all conditions at all electrode sites.

	Condition							
	Alternating-Blocked high		Alternating-Blocked low		Alternating—Mixed high		Alternating—Mixed low	
Electrode	Range 1	Range 2	Range 1	Range 2	Range 1	Range 2	Range 1	Range 2
Fz	−0.67	−0.25	−0.97	0.04	−0.35	−0.43	−1.13	−0.59
	(1.01)	(0.74)	(0.57)	(0.73)	(0.62)	(0.64)	(0.70)	(0.37)
Cz	−0.09	−0.17	−0.60	0.16	−0.21	−0.17	−0.89	−0.32
	(1.15)	(1.05)	(0.43)	(0.77)	(0.76)	(0.81)	(0.58)	(0.59)
Pz	0.68	0.04	0.29	0.46	0.26	0.23	−0.13	0.08
	(0.95)	(0.96)	(0.68)	(0.60)	(1.06)	(0.88)	(1.11)	(1.00)
F3	−0.54	−0.13	−0.75	0.12	−0.32	−0.27	−0.89	−0.39
	(0.98)	(0.65)	(0.45)	(0.71)	(0.53)	(0.60)	(0.69)	(0.43)
F4	−0.55	−0.21	−0.87	0.20	−0.21	−0.21	−0.98	−0.52
	(0.86)	(0.56)	(0.69)	(0.77)	(0.67)	(0.63)	(0.63)	(0.42)
C3	0.02	−0.04	−0.35	0.31	−0.07	−0.05	−0.67	−0.17
	(0.88)	(0.87)	(0.31)	(0.58)	(0.65)	(0.82)	(0.70)	(0.70)
C4	0.13	−0.05	−0.35	0.35	0.06	0.04	−0.62	−0.13
	(1.03)	(0.79)	(0.60)	(0.81)	(0.87)	(0.83)	(0.60)	(0.62)
LM	1.07	0.20	1.24	0.52	0.86	0.56	0.79	0.45
	(0.77)	(0.66)	(0.91)	(0.31)	(0.69)	(0.55)	(1.01)	(0.84)
RM	0.85	0.31	1.11	0.54	0.68	0.55	0.75	0.42
	(0.50)	(0.50)	(0.62)	(0.58)	(0.79)	(0.60)	(0.77)	(0.73)

subjects are reported. Eight of these subjects also participated in experiment 2. None of the subjects participated in experiment 1.

C. Stimuli

Two types of complex sounds were used (created with *Adobe Audition 1.0* software): harmonic complexes (called ''tones'') and bandpass-filtered noise bursts (called ''noise''). The harmonic complexes had the same fundamental frequency ($f0$) as the low tones of experiments 1 and 2 (440 Hz for the standard sound and 494 Hz for the deviant sound) and were composed of the first three harmonics with equal amplitudes. Tone 1 (T) consisted of 440, 880, and 1320 Hz, and tone 2 (TD) consisted of 494, 988, and 1482 Hz. The other type of complex tone (N) was created by narrow-band fast Fourier transform (FFT) filter performed on broadband noise between 440 and 1320 Hz. Thus, it overlapped the same spectral region (440–1320) as the standard complex tone (T). Tone duration for all tones was 50 ms, with onset and offset ramps of 7.5 ms made with a Hanning window. Sound intensity of the tones and noise were equated and presented at 77 dB SPL (Brüel & Kjær 2209 sound-level meter).

Stimuli were presented in four conditions: *tones-only blocked*, *tones-only mixed*, *alternating-blocked* and *alternating-mixed*. The *tones-only conditions* were control conditions to verify that the auditory context would have a similar effect on complex tones as they did with the pure-tone stimuli in previous studies (i.e., double deviants eliciting one MMN in blocked context and two MMNs in mixed context). In the event that the effects of context on integration of sequential sounds were specific to pure tones, the control conditions were needed to determine this. The sound sequence was presented at 150 ms onset-to-onset pace, the pace that the complex tones would occur in the alternating conditions (*alternating-blocked* and *alternating-mixed*) of the main experiment. Three thousand stimuli were presented in two runs of 1500 stimuli each, for the *tone-only blocked condition* and 6000 stimuli in four runs of 1500 stimuli for the *tone-only mixed condition*.

In the alternating conditions, tone and noise stimuli were alternated (T-N-T-N-T-N···) at an onset-to-onset pace of 75 ms. The alternating conditions used the same basic paradigm as experiment 1, with the context changes occurring only in the tones stream. Six thousand stimuli were presented in two runs of 3000 stimuli for the alternating-blocked condition and 24 000 stimuli in four runs for the alternating-mixed condition. Deviants (TD) occurred 15% overall in all conditions (7.5% occurrence of double deviants in the blocked conditions and 8% single deviants and 7% double deviants in the mixed conditions).

D. Procedures

Participants sat in a comfortable chair approximately 1 m from a TV monitor in a sound attenuated booth. They were instructed to watch a silent closed-captioned video and to pay no attention to the sounds. Three thousand stimuli were presented in two runs of 1500 stimuli each, for the *tone-only blocked condition* and 6000 stimuli in four runs of 1500 stimuli for the *tone-only mixed condition*. For the alternating conditions, 6000 stimuli were presented in two runs of 3000 stimuli for the *alternating-blocked condition* and 24 000 stimuli in four runs for the *alternating-mixed condition*. Deviants occurred 15% overall in all conditions (7.5% occurrence of double deviants in the blocked conditions and 8% single deviants and 7% double deviants in the mixed

conditions). The order of the four conditions was counterbalanced across subjects to avoid any possible order effects.

E. Electroencephalogram (EEG) recording and data analysis

Data recording parameters were the same as those used in experiment 2. For each participant, the artifact-free epochs were averaged separately for the complex tones by stimulus type [standards (T) and double deviants (TD)], separately, in the four conditions (tones-only blocked, tones-only mixed, alternating-blocked, and alternating-mixed).

The difference waveforms were obtained by subtracting the ERP response to the standard from the ERP response to the deviant. The intervals used for statistically evaluating for the presence of the potential MMNs were determined from the group mean deviant-minus-standard waveforms at Fz (the site of greatest signal-to-noise ratio), taking a 40-ms window centered on the peak of the MMNs where elicited. For the blocked conditions, range 2 intervals from the mixed conditions were used to measure the presence of MMN in range 2 of the blocked conditions because MMN was expected to be elicited in the mixed conditions. For the tones-only blocked condition the mean amplitude was obtained in the following intervals, range 1: 104–144 ms and range 2: 282–322 ms; for the tones-only mixed condition the intervals were for range 1: 100–140 ms and range 2: 282–322 ms; for the alternating-blocked condition the intervals were for range 1: 104–144 ms and range 2: 258–298; and for the alternating-mixed condition the intervals were for range 1: 100–140 ms and range 2: 258–298. Intervals were measured for each subject individually and then used to verify the presence of the MMN using a repeated measures ANOVA with factors of condition (alternating-blocked, tones-only blocked, alternating-mixed, tones-only mixed) $\times$peak (range 1 versus range 2) and stimulus type (deviant versus standard) to determine in which conditions and which peaks the standard and deviant waveforms differed from each. ANOVA with factors of condition (blocked versus mixed), control (alternating versus tones-only), and range (range 1 versus range 2) was calculated on the difference waveforms to compare mean voltages. Greenhouse–Geisser correction was reported when appropriate. Tukey HSD *post hoc* tests were calculated.

F. Results and discussion

The ERP results for the *tones-only* and *alternating* conditions are displayed in Figs. 8 and 9, respectively (also see Table III).

The ANOVA revealed a three-way interaction between condition, stimulus type, and range [$F(3,24)=6.25, p<0.01$], showing that the significant difference between the standard and deviant (MMN) was dependent upon the condition and range. Tukey HSD *post hoc* comparisons revealed that MMNs were elicited in both range 1 and range 2 of the **mixed conditions** [*alternating mixed* and *tones-only mixed*; see Figs. 8 (bottom row) and 9 (bottom row)], whereas MMNs were elicited only in range 1 of the **blocked conditions** [*alternating blocked* and *tones-only blocked*; see Figs. 8 (top row) and 9 (top row). This pattern of results is consistent with experiments 1 and 2, in which two MMNs were elicited by double deviants in the mixed context and one MMN was elicited in the blocked context.

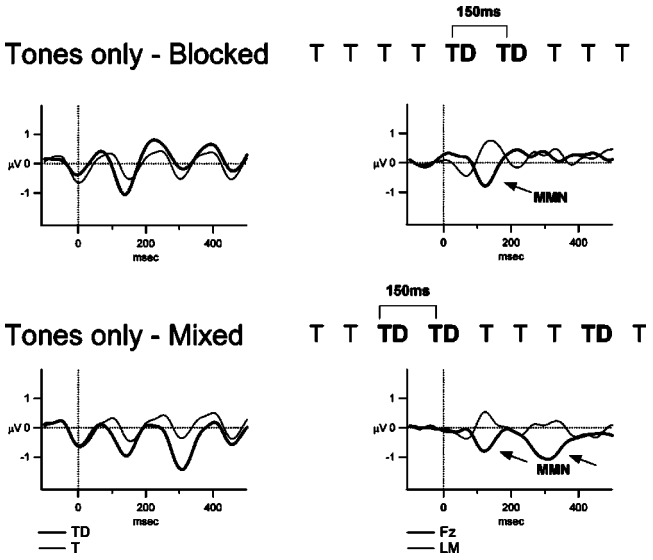


FIG. 8. *Experiment 3—Tones-only conditions*. Grand mean event-related brains potentials elicited by the T (thin line) and TD (thick line) tones in the tones-only conditions are displayed in the left column. The deviant-minus-standard difference curves are displayed in the right column with Fz (thick line) overlaid with the left mastoid (LM; thin line). Arrows point to significant MMNs.

The ANOVA comparing mean voltage of the deviant-minus-standard difference waveforms revealed an interaction between condition and range with no other interactions [$F(1,8)=21.11, p<0.002$]. *Post hoc* analyses show that the mean voltage of range 1 was larger than that of range 2 for the blocked conditions and there was no difference in mean voltage between ranges 1 and 2 in the mixed conditions. This substantiates the finding that one MMN was elicited in the blocked conditions and two MMNs were elicited in the mixed conditions.

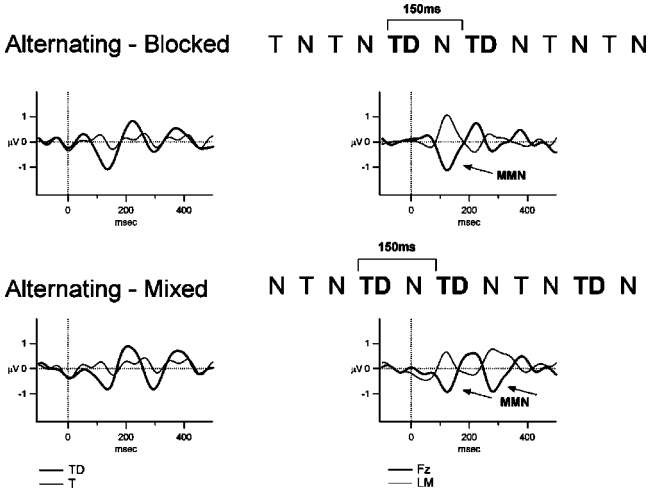


FIG. 9. *Experiment 3—Alternating conditions*. Grand mean event-related brains potentials elicited by the T (thin line) and TD (thick line) tones in the alternating conditions are displayed in the left column. The deviant-minus-standard difference curves are displayed in the right column with Fz (thick line) overlaid with the left mastoid (LM; thin line). Arrows point to significant MMNs.

TABLE III. Experiment 3. Mean amplitude values (in μV) and standard deviations (in parentheses) of the grand-averaged deviant-minus-standard (difference) waveforms in ranges 1 and 2 for all conditions at all electrode sites.

	Condition							
	Tones-only—Blocked		Tones-only—Mixed		Alternating—Blocked		Alternating—Mixed	
Electrode	Range 1	Range 2	Range 1	Range 2	Range 1	Range 2	Range 1	Range 2
Fz	−0.72	0.13	−0.63	−0.81	−1.02	−0.34	−0.76	−0.64
	(0.84)	(0.50)	(0.31)	(0.39)	(0.46)	(0.47)	(0.76)	(0.44)
Cz	−0.40	0.28	−0.42	−0.78	−0.71	−0.26	−0.66	−0.39
	(0.58)	(0.51)	(0.41)	(0.56)	(0.64)	(0.52)	(0.88)	(0.64)
Pz	0.09	0.31	0.10	−0.36	−0.14	−0.14	−0.10	−0.05
	(0.55)	(0.59)	(0.88)	(0.96)	(0.82)	(0.63)	(0.92)	(0.90)
F3	−0.64	0.14	−0.49	−0.59	−0.87	−0.23	−0.71	−0.49
	(0.76)	(0.42)	(0.33)	(0.43)	(0.44)	(0.35)	(0.83)	(0.34)
F4	−0.66	0.18	−0.53	−0.81	−0.89	−0.23	−0.60	−0.45
	(0.81)	(0.59)	(0.37)	(0.45)	(0.40)	(0.44)	(0.72)	(0.49)
C3	−0.34	0.23	−0.32	−0.61	−0.60	−0.22	−0.50	−0.21
	(0.58)	(0.49)	(0.45)	(0.59)	(0.66)	(0.55)	(0.88)	(0.63)
C4	−0.24	0.33	−0.13	−0.55	−0.61	−0.11	−0.28	−0.12
	(0.53)	(0.54)	(0.50)	(0.71)	(0.55)	(0.54)	(0.75)	(0.55)
LM	0.56	0.23	0.48	0.23	0.92	0.21	0.58	0.76
	(0.82)	(0.65)	(0.73)	(0.71)	(0.66)	(0.65)	(0.67)	(0.72)
RM	0.46	0.28	0.50	0.32	0.76	0.30	0.70	0.73
	(0.67)	(0.45)	(0.62)	(0.69)	(0.51)	(0.53)	(0.66)	(0.70)

The results extend the findings of our previous studies using pure tones (e.g., Sussman *et al.*, 2002b), now demonstrating with complex tones the same contextual effects on within-stream integration processes observed with pure tones. Additionally, the results demonstrate stream segregation by a nonspectral cue. This suggests that the stream segregation process involves more than simple frequency resolution of peripheral mechanisms. This finding is consistent with other studies showing that stream segregation is not solely dependent upon peripheral channeling (Bey and McAdams, 2003 [Experiment 1]; Cusack and Roberts, 2000; Iverson, 1995; Moore and Glockel, 2002; Roberts *et al.*, 2002; Vliegen and Oxenham, 1999). The results provide strong support for the notion that segregation occurs prior to integration and that integration effects operate within already segregated sound streams. The interplay between these two processes is thus observed in complex sound situations, such as would be needed for processing speech and other environmental sounds.

V. GENERAL DISCUSSION

Three experiments were conducted to investigate the timing of two organizational processes of ASA, segregation of streams and integration of within-stream elements, when they were called upon to function concurrently and the listener had no task to perform with the sounds. The within-stream contextual effects, the integration of successive elements to a single or double deviant event as indexed by MMN, followed the same pattern of results as has been demonstrated for temporal integration processes when only one sound stream was present (Sussman *et al.*, 2002b). This indicates that the segregation to independent streams occurred prior to the integration and formation of the auditory units within streams. The results demonstrate that (1) multiple sound streams are maintained independently, (2) stream segregation occurs prior to within-stream integration or event formation, and (3) auditory context effects act on the already formed streams.

Experiments 1 and 2 demonstrated that the integration process operates on two streams independently, as though each stream occurred alone. These results show that the integration occurred on the already segregated streams, the segregation process taking place at an earlier level of processing than the integration process. Experiment 3 confirmed that the segregation process preceded the integration of information within streams. Given that the spectral compositions of the two sets of sounds were similar in experiment 3, if the sounds were processed as part of the same stream, then the effects of intervening tones on within stream processes would have been observed (experiment 1, control conditions). Thus, the results of experiment 3 provided strong support to conclude that the segregation processes mediated the integration processes.

The timing of the segregation and integration processes that have been demonstrated in the current and previous studies suggests the following model. A neural representation of the sound input is initially made regarding how many sources are present in the signal. This is calculated according to the acoustic characteristics of the input (the frequency, intensity, duration, and spatial location features as well as the timing of the input). Extraction of the acoustic regularities is then established on the already formed streams (i.e., the context is determined). The long-term effects of the auditory context mediate local, within-stream, event formation (Sussman and Winkler, 2001). The integration processes (or sound event formation) operate after the within-stream context has

been determined. Thus, whether a sound element is processed as one or two events depends on the surrounding information within the same stream. The MMN process uses this information—the representation of the segregated streams and the detected sound regularities within streams—in determining what is deviant in the sound input.

A. Implications for speech processing

Extending the notion that the integration of sound elements into perceptual units proceeds on the already segregated information, the data discussed here support the view that sound elements are integrated into linguistic units (phonemes, syllables, and words) after the initial segregation of the input into distinct sound sources. Speech processing, according to this model, would rely, at least in part, on the primitive ASA processes (Brown and Cooke, 1994; Rosenthal and Okuno, 1998).

When more than one voice occurs simultaneously, segregation by acoustic parameters (e.g., $f0$) serves to distinguish the separate voices in the auditory environment, whereas the integration of successive sound elements within the separate voices serves to distinguish speech units. There are additional segmentation processes that are also necessary for within-stream comprehension, those that allow us to distinguish words within a speech stream or individual footsteps within a series. For example, hearing the difference between ''topic'' and ''top pick'' is determined by the duration of the silent /p/ closure [less than 150 ms was needed to hear one word and between 150 and 300 ms was needed to hear two words; Pickett and Decker (1960)]. In fact, this timing that made the difference between hearing one or two words is consistent with our finding in which the same timing made the difference between tones (pure or complex) being represented as one or two deviant events. Two successive deviants elicited one MMN at 150-ms interval between them, whereas the two successive deviants elicited two MMNs when they were separated by 300 ms (Sussman *et al.*, 1999b). Further research would be needed, however, to clarify whether pure acoustic cues function similarly as acoustic cues embedded in speech for defining auditory events.

Within-stream segmentation also takes place within the ''integrating'' period. The standard intervening tone between the two successive deviant tones (in the *low-only—standard intervening condition*) influenced the integration process as reflected by the MMN response (i.e., affected what became integrated into one event). When a single tone intervened between the two successive deviant tones, falling within the less than 200-ms integrating window, each of the deviants elicited its own MMN. What otherwise might have been responded to as one deviant event (two successive deviants occurring within 150 ms) was separated into two deviant events (eliciting two separate MMNs) by the presence of the intervening tone occurring between them when they were processed in the same low-tone stream. This result indicates that the integration of elements occurring within less than 200 ms operates according to the sequential organization of the unit or auditory event that falls within the window. The temporal proximity of the elements, on its own, is not enough for integration—this should already be evident because the phenomenon of streaming can occur when sounds are 100 ms apart—but neither is the frequency proximity sufficient (the low tone standards are near in frequency to the deviants but are not integrated just by the fact that they fall within this window). Thus, the result indicates that the integration mechanism is ''searching'' for elements that likely form a perceptual unit. The auditory system operates on the basis of the informational units falling within the window and does not just integrate all the elements that fall within a window of less than 200 ms.

This is further confirmed with the triple deviants in the *low-only—triple deviants condition*, in which three successive deviants occurring within the integrating period elicited one MMN. Elements that fall within the integrating window are being integrated to form a single auditory event. These data lend credence to the notion that the integrating period operates independently of the sheer number of elements falling within the window and focuses on the formation of events out of closely occurring sound elements. The integration of elements across time occurred as a function of the role of the stimuli that fell within the integrating period (e.g., deviant vs. standard). Whereas three deviant sounds occurring successively were responded to as a single deviant event, the successive deviant-standard-deviant pattern was not integrated, even though the tones occurred within the same time period. This demonstrates that auditory context influences the integrating window in terms of auditory event formation. This mechanism could be important in speech processing when determining which phonemic units are to be integrated and which segregated, such as those cues that distinguish ''topic'' from ''top pick.''

It remains to be determined whether the segmentation processes that operate within streams to facilitate identification of perceptual units (and which are also part of integration processes necessary to link acoustic elements to perceptual units) differ from the procedures that are initially used to determine how many sources are in the environment. Additionally, the role of attention in these processes remains to be determined.

Context effects on MMN have now been demonstrated in a number of different studies using a wide range of paradigms (Sussman *et al.*, 1998a, 1999a, 2001, 2002a, b; Sussman and Winkler, 2001; Winkler *et al.*, 2003). Interestingly, MMN has also been reported to reflect categorical perception of consonants and vowels (Dehaene-Lambertz *et al.*, 2000; Näätänen *et al.*, 1997; Winkler *et al.*, 1999), voice onset time (Sharma and Dorman, 2000), and length (Nenonen *et al.*, 2003). The phonetic context effects (the shift in category boundary that is altered by surrounding speech characteristics) observed with these MMN studies of categorical perception might actually represent another form of context effect on MMN and not a speech-specific effect (Pisoni *et al.*, 1983; Sussman *et al.*, 2004).

B. Concluding remarks

The results of the current study demonstrate context effects on auditory event formation that occur after the segregation of sounds to distinct streams when the listener is not focused on the sound input. Moreover, integration of within-

stream elements is influenced by the larger context of the within-stream sounds. The auditory system does not just integrate all information that falls within a 150-ms window; the informational content of the elements determines what is integrated within this period and what is not. These data, taken together with previous results, also showing that segregation occurs prior to other processes (e.g., Sussman *et al.*, 1998b; 1999b; Yabe *et al.*, 2001), provide strong evidence that segregation is an earlier, primitive process initially driven by stimulus characteristics of the acoustic input. Functionally, the integration of sequential elements to perceptual units taking place on the already segregated streams would facilitate the ability to identify within-stream sound patterns, thus making it possible to appreciate music or comprehend speech.

ACKNOWLEDGMENTS

This research was supported by the National Institutes of Health (Grant No. R01 DC04263). The author is grateful to Dr. Albert S. Bregman for his advice on experiment 3 and for his comments on an earlier version of the manuscript. The author thanks Dr. Yonathan Fishman and three anonymous reviewers for their helpful comments on an earlier version of the manuscript. The author is also grateful to Rosie Wong and Fakhar Kahn for collecting the data and for their assistance with data processing and manuscript preparation.

[1]This transient form of auditory memory has been estimated to store information for a period of time of at least 30 s (Cowan, 2001).

Beauvois, M., and Meddis, R. (**1991**). ''A computer model of auditory stream segregation,'' Q. J. Exp. Psychol. A **43A**, 517–541.

Bey, C., and McAdams, S. (**2003**). ''Postrecognition of interleaved melodies as an indirect measure of auditory stream formation,'' J. Exp. Psychol. Hum. Percept. Perform. **29**(2), 267–279.

Botte, M.-C., Drake, C., Brochard, R., and McAdams, S. (**1997**). ''Perceptual attenuation of nonfocused auditory streams,'' Percept. Psychophys. **59**, 419–425.

Bregman, A. S. (**1990**). *Auditory Scene Analysis* (MIT, Cambridge, MA).

Brochard, R., Drake, C., Botte, M.-C., and McAdams, S. (**1999**). ''Perceptual organization of complex auditory sequences: Effect of number of simultaneous sub-sequences and frequency separation,'' J. Exp. Psychol. Hum. Percept. Perform. **25**, 1742–1759.

Brown, G. J., and Cooke, M. P. (**1994**). ''Computational auditory scene analysis,'' Comput. Speech Lang. **8**(4), 297–336.

Carlyon, R. P., Cusack, R., Foxton, J. M., and Robertson, I. H. (**2001**). ''Effects of attention and unilateral neglect on auditory stream segregation,'' J. Exp. Psychol. Hum. Percept. Perform. **27**(1), 115–127.

Compte, A., Brunel, N., Goldman-Rakic, P. S., and Wang, X.-J. (**2000**). ''Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model,'' Cereb. Cortex **10**(9), 910–923.

Cowan, N. (**2001**). ''The magical number 4 in short-term memory: a reconsideration of mental storage capacity,'' Behav. Brain Sci. **24**(1), 87–114.

Cusack, R., and Roberts, B. (**1999**). ''Effects of similarity in bandwidth on the auditory sequential streaming of two-tone complexes,'' Percept. Psychophys. **28**, 1281–1289.

Cusack, R., and Roberts, B. (**2000**). ''Effects of differences in timbre on sequential grouping,'' Percept. Psychophys. **62**, 1112–1120.

Dehaene-Lambertz, G., Dupoux, E., and Gout, A. (**2000**). ''Electrophysiological correlates of phonological processing: a cross-linguistic study,'' J. cogn. Neurosci. **12**, 635–647.

Fay, R. R. (**2000**). ''Spectral contrasts underlying auditory stream segregation in goldfish (Carassius auratus),'' J. Assoc. Res. Otolaryngol. **1**, 120–128.

Fishman, Y. I., Reser, D., Arezzo, J., and Steinschneider, M. (**2001**). ''Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey,'' Hear. Res. **151**, 167–187.

Giard, M. H., Perrin, F., Pernier, J., and Bouchet, P. (**1990**). ''Brain generators implicated in processing of auditory stimulus deviance: A topographic event-related potential study,'' Psychophysiology **27**, 627–640.

Halgren, E., Baudena, P., Clarke, J. M., Heit, G., Liegeois, C., Chauvel, P., and Musolino, A. (**1995**). ''Intracerebral potentials to rare target and distractor auditory and visual stimuli. I. Superior temporal plane and parietal lobe,'' Electroencephalogr. Clin. Neurophysiol. **94**, 191–220.

Ha, K. S., Youn, T., Kong, S. W., Park, H. J., Ha, T. H., Kim, M. S., and Kwon, J. S. (**2003**). ''Optimized individual mismatch negativity source localization using a realistic head model and the Talairach coordinate system.'' Brain Topography **15**(4), 233-238.

Hartmann, W. M., and Johnson, D. (**1991**). ''Stream segregation and peripheral channeling,'' Music Percept. **9**, 155–184.

Hulse, S. H., MacDougall-Shackleton, S. A., and Wisniewski, A. B. (**1997**). ''Auditory scene analysis by songbirds: stream segregation of birdsong by European starlings (Sturnus vulgaris),'' J. Comp. Psychol. **111**, 3–13.

Iverson, P. (**1995**). ''Auditory stream segregation by musical timbre: effects of static and dynamic acoustic attributes,'' J. Exp. Psychol. Hum. Percept. Perform. **21**, 751–763.

Javitt, D. C., Steinschneider, M., Schroeder, C. E., and Arezzo, J. C. (**1996**). ''Role of cortical N-methyl-D-aspartate receptors in auditory sensory memory and mismatch negativity generation: Implications for schizophrenia,'' Proc. Natl. Acad. Sci. USA **93**, 11962–11967.

Jones, D. M., Alford, D., Bridges, A., Tremblay, S., and Macken, W. J. (**1999**). ''Organizational factors in selective attention: The interplay of acoustic distinctiveness and auditory streaming in the irrelevant sound effect,'' J. Exp. Psychol.: Learn. Mem. Cogn. **25**, 464–473.

Jones, M. R., Maser, D. J., and Kidd, G. R., (**1978**). ''Rate and structure in memory for auditory patterns,'' Mem. Cognit. **6**, 246–258.

Kauer, J. A., Malenka, R. C., and Nicoll, R. A. (**1988**). ''NMDA application potentiates synaptic transmission in the hippocampus,'' Nature (London) **334**, 250–252.

Kropotov, J. D., Näätänen, R., Sevostianov, A. V., Alho, K., Reinkainen, K., and Kropotova, O. V. (**1995**). ''Mismatch negativity to auditory stimulus change recorded directly from the human temporal cortex,'' Psychophysiology **32**, 418–422.

Macken, W. J., Tremblay, S., Houghton, R. J., Nicholls, A. P., and Jones, D. M. (**2003**). ''Does auditory streaming requie attention? Evidence from attentional selectivity in short-term memory,'' J. Exp. Psychol. Hum. Percept. Perform. **29**(1), 43–51.

Massaro, D. W. Experimental Psychology and Information Processing. Chicago: Rand McNally, 1975.

Moore, B. C. J., and Gockel, H. (**2002**). ''Factors influencing sequential stream segregation,'' Acust. Acta Acust. **88**, 320–332.

Näätänen, R. (**1992**). *Attention and Brain Function* (Erlbaum Hillsdale, NJ).

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R. J., Luuk, A., Allik, J., Sinkkonen, J., and Alho, K. (**1997**). ''Language-specific phoneme representations revealed by electric and magnetic brain responses,'' Nature (London) **385**, 432–434.

Nenonen, S., Shestakova, A., Huotilainen, M., and Näätänen, R. (**2003**). ''Linguistic relevance of duration within the native language determines the accuracy of speech-sound duration processing,'' Cognitive Brain Res. **16**, 492–495.

Opitz, B., Mecklinger, A., von Cramon, D. Y., and Kruggel, F. (**1999**). ''Combining electrophysiological and hemodynamic measures of the auditory oddball.'' Psychophysiology **36**(1), 142-147.

Pickett, J. M., and Decker, L. R. (**1960**). ''Time factors in perception of a double consonant,'' Lang. Speech **3**, 11–17.

Pisoni, D. B., Carrell, T. D., and Gans, S. J. (**1983**). ''Perception of the duration of rapid spectrum changes in speech and nonspeech signals,'' Percept. Psychophys. **34**, 314–322.

Pulvirenti, L. (**1992**). ''Neural plasticity and memory: towards an integrated view,'' Funct Neurol. **7**(6), 481–490.

Ritter, W., Sussman, E., and Molholm, S. (**2000**). ''Evidence that the mismatch negativity system works on the basis of objects,'' NeuroReport **11**, 61–63.

Roberts, B., Glasberg, B. R., and Moore, B. C. (**2002**). ''Primitive stream segregation of tone sequences without differences in fundamental frequency or passband,'' J. Acoust. Soc. Am. **112**, 2074–2085.

Rose, M. M., and Moore, B. C. (**2000**). ''Effects of frequency and level on auditory stream segregation,'' J. Acoust. Soc. Am. **108**, 1209–1214.

Rosenthal, D. F., and Okuno, H. G. (Eds.) (**1998**). *Computational Auditory Scene Analysis* (Erlbaum, Mahwah, NJ).
Sams, M., and Hari, R. (**1991**). ''Magnetoencephalography in the study of human auditory information processing,'' Ann. N.Y. Acad. Sci. **620**, 102–117.
Scherg, M., Vajsar, J., and Picton, T. W. (**1989**). ''A source analysis of the late human auditory evoked potentials,'' J. Cogn. Neurosci. **1**, 336–355.
Sharma, A., and Dorman, M. F. (**2000**). ''Neurophysiologic correlates of cross-language phonetic perception,'' J. Acoust. Soc. Am. **107**, 2697–2703.
Singh, P. G. and Bregman, A. S. (**1997**). ''The influence of different timbre attributes on the perceptual segregation of complex-tone sequences,'' J. Acoust. Soc. Am. **102**(4), 1943-1952.
Sussman, E., and Winkler, I. (**2001**). ''Dynamic process of sensory updating in the auditory system,'' Cognitive Brain Res. **12**, 431–439.
Sussman, E., Horváth, J., and Winkler, I. (**submitted**). ''The role of attention in the formation of auditory streams,'' Percept. Psychophys.
Sussman, E., Ritter, W., and Vaughan, Jr., H. G. (**1998a**). ''Stimulus predictability and the mismatch negativity system,'' NeuroReport **9**, 4167–4170
Sussman, E., Ritter, W., and Vaughan, Jr., H. G. (**1998b**). ''Attention affects the organization of auditory input associated with the mismatch negativity system,'' Brain Res. **789**, 130–138.
Sussman, E., Ritter, W., and Vaughan, Jr., H. G. (**1999a**). ''An investigation of the auditory streaming effect using event-related brain potentials,'' Psychophysiology **36**, 22–34.
Sussman, E., Bregman, A. S., Wang, W., and Kahn, F. J. (**in press**). ''Attentional modulation of electrophysiological activity in auditory cortex for unattended sounds in multi-stream auditory environments,'' Cognitive, Affective, and Behavioral Neurosci.
Sussman, E., Sheridan, K., Kreuzer, J., and Winkler, I. (**2003**). ''Representation of the standard: stimulus context effects on the process generating the mismatch negativity component of event-related brain potentials,'' Psychophysiology **40**, 465–471.
Sussman, E., Čeponien, R., Shestakova, A., Näätänen, R., and Winkler, I. (**2001**). ''Auditory stream segregation processes operate similarly in school-aged children as adults,'' Hear. Res. **153**(1-2), 108–114.
Sussman, E., Winkler, I., Huoutilainen, M., Ritter, W., and Näätänen, R. (**2002a**). ''Top-down effects on the initially stimulus-driven auditory organization,'' Cognitive Brain Res. **13**, 393–405.
Sussman, E., Winkler, I., Ritter, W., Alho, K., and Näätänen, R. (**1999b**). ''Temporal integration of auditory stimulus deviance as reflected by the mismatch negativity,'' Neurosci. Lett. **264**, 161–164.
Sussman, E., Kujala, T., Halmetoja, J., Lyytinen, H., Alku, P., and Näätänen, R. (**2004**). ''Automatic and controlled processing of acoustic and phonetic contrasts,'' Hear. Res. **190**(1-2), 128–140.
Sussman, E., Winkler, I., Kreuzer, J., Saher, M., Näätänen, R., and Ritter, W. (**2002b**). ''Temporal integration: Intentional sound discrimination does not modify stimulus-driven processes in auditory event synthesis,'' Clin. Neurophysiol. **113**, 909–920.
Vliegen, J., and Oxenham, A. J. (**1999**). ''Sequential stream segregation in the absence of spectral cues,'' J. Acoust. Soc. Am. **105**, 339–346.
Winkler, I., Sussman, E., Tervaniemi, M., Ritter, W., Horvath, J., and Näätänen, R. (**2003**). ''Preattentive auditory context effects,'' Cogn., Affect., and Behav. Neurosci. **3**(1), 57–77.
Winkler, I., Lehtokoski, A., Alku, P., Vainio, M., Czigler, I., Csépe, V., Aaltonen, O., Raimo, I., Alho, K., Lang, A. H., Iivonen, A., and Näätänen, R. (**1999b**). ''Preattentive detection of vowel contrasts utilizes both phonetic and auditory memory representations,'' Cognitive Brain Res. **7**, 357–369.
Woldorff, M., Hillyard, S., Gallen, C. C., Hampson, R. S., and Bloom, F. E. (**1998**). ''Magnetoencephalographic recordings demonstrate attentional modulation of mismatch-related neural activity in human auditory cortex,'' Psychophysiology **35**, 283–292.
Yabe, H., Winkler, I., Czigler, I., Koyama, S., Kakigi, R., Sutoh, T., Hiruma, T., and Kaneko, S. (**2001**). ''Organizing sound sequences in the human brain: the interplay of auditory streaming and temporal integration,'' Brain Res. **897**, 222–227.
Zwislowski, J. J. (**1969**). ''Temporal summation of loudness: An analysis,'' J. Acoust. Soc. Am. **46**, 431–440.

Combined representations for frequency and duration in detection templates for expected signals

Beverly A. Wright[a),b)]
Department of Communication Sciences and Disorders and Northwestern University Institute for Neuroscience, Northwestern University, 2240 Campus Drive, Evanston, Illinois 60208-3550

(Received 12 July 2004; revised 28 November 2004; accepted 13 December 2004)

When trying to detect a tonal signal in a continuous broadband noise, listeners attend selectively to both the frequency and the duration of the expected signal. However, it is not known whether they monitor separate or combined representations of these two attributes. To investigate this question, a probe-signal method was used to measure the detectability of signals of expected and unexpected durations at two expected frequencies. The four listeners expected only one of two signals to be presented at random: a brief tone at one frequency or a long tone at another frequency. For each signal frequency, the detectability of the signals of unexpected duration decreased to near chance as the difference between the expected and unexpected duration, at that frequency, increased. The frequency specificity of this duration tuning indicates that both the frequency and the duration of an expected stimulus are represented in a single template. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1855771]

PACS numbers: 43.66.Dc, 43.66.Mk [GDK] Pages: 1299–1304

I. INTRODUCTION

One of the primary purposes of sensory systems is to inform organisms of the presence of external stimuli. In humans, at least, stimulus detection is implicitly defined as conscious awareness of the stimulus. This awareness requires attention. Therefore, to fully understand the process of stimulus detection, we must know how this attention is distributed across different stimulus attributes in a variety of circumstances. Here, for reasons described below, we examined the distribution of attention in an auditory signal-detection task in which listeners were uncertain as to which of two signals, that differed in both frequency and duration, was to be presented.

Many investigations of the distribution of attention in auditory signal-detection tasks have employed the probe-signal method (Greenberg and Larkin, 1968). With this method, listeners are led to expect a particular target signal by presenting it on the majority of trials. However, they occasionally are presented unexpected probe signals that typically differ from the target signal along only one stimulus dimension. Each signal is detectable on ~85% of trials when it is the only one presented. The detectability of the unexpected signals is taken to indicate the selectivity with which listeners attend to the particular value of the expected stimulus along the varied stimulus dimension.

From such experiments it is clear that listeners trying to detect a tonal signal in a continuous broadband noise attend selectively to both the frequency and the duration of the expected signal, but it is not known whether they monitor separate, or combined, representations of these two attributes. Demonstrating focused attention to the signal frequency, listeners led to expect a signal of a particular frequency detect that signal on a higher percentage of trials than signals of unexpected frequency (Greenberg and Larkin, 1968; Yama and Robinson, 1982; Scharf *et al.*, 1987; Dai *et al.*, 1991; Wright and Dai, 1994a). Indicating duration-specific attention, listeners who expect a signal of a particular duration detect that signal more frequently than signals that differ from it in duration (Wright and Dai, 1994a; Dai and Wright, 1995). The primary purpose of the present experiment was to determine whether the selective attention to both frequency and duration reflects (a) the intersection of two separate attention templates, one tuned to the expected frequency, and the other tuned to the expected duration, or, as we (Wright and Dai, 1994a; Dai and Wright, 1995) have claimed but not tested (b) the product of a single template, tuned to both the expected frequency and duration.

It also is not known whether listeners selectively attend to two or more expected signal durations when they are uncertain as to which duration is to be presented, or, if so, whether the uncertainty influences the duration tuning of those templates. When listeners are uncertain as to which of several expected signal *frequencies* is to be presented, signal detectability is highest for signals presented at the expected frequencies, and lower for signals at unexpected frequencies, even when the unexpected frequencies are between two expected ones (Macmillan and Schwartz, 1975; Howard *et al.*, 1986; Schlauch and Hafter, 1991; Hafter *et al.*, 1993). Listeners therefore appear to monitor multiple frequency-tuned templates, each of which is centered on an expected signal frequency. However, the frequency bandwidths of these templates are often (Schlauch and Hafter, 1991; Hafter *et al.*, 1993), though not always (Macmillan and Schwartz, 1975), reported to be wider than those measured in conditions in which only a single frequency is expected. The secondary purpose of this experiment was to establish whether listeners

[a)]Portions of this work were presented in "Evidence for auditory signal-detection templates tuned to both the frequency and duration of an expected signal," 147th Meeting of the Acoustical Society of America, New York City, New York, May 2004.

[b)]Electronic mail: b-wright@northwestern.edu

also selectively monitor several different templates tuned to different *durations* in conditions involving duration uncertainty, and, if so, whether uncertainty affects the duration bandwidth of those templates.

To address these questions, we used a variant of the probe-signal method (Greenberg and Larkin, 1968) to measure the detectability of signals of expected or unexpected durations that were presented randomly at one of two expected frequencies. Each signal was detectable on ~90% of trials when it was the only possible signal. To the listener, the task appeared to be signal detection in the presence of uncertainty as to which of two signals was to be presented. These expected signals, presented on the majority of trials, were a brief, low-frequency tone and a long, high-frequency tone in one case, and a long, low-frequency tone and a brief, high-frequency tone in another case. Interspersed among these expected-signal trials were occasional unexpected-signal trials in which the signal had the same frequency as, but a different duration than, one or the other of the two expected signals. Thus, all of the signals were presented at one of two expected frequencies, but some signals had durations that were shorter, or longer, than expected.

Assuming that listeners would attend selectively to the two expected signal frequencies, as demonstrated previously (Macmillan and Schwartz, 1975; Howard *et al.*, 1986; Schlauch and Hafter, 1991; Hafter *et al.*, 1993), we reasoned that the detectability pattern for the signals with unexpected durations should indicate whether or not listeners (a) monitor both the frequency and duration of a given expected signal through a single template and (b) attend to two different duration-tuned templates in the present signal-uncertainty task. We report that detectability was high only for the brief signal at the frequency where it was expected and for the long signal at the frequency where it was expected. These results support the ideas that both frequency and duration are represented in a single template, and that listeners who are uncertain as to which of two signal durations is to be presented attend to two different duration templates, each of which is tuned to the duration of an expected signal.

II. METHOD

A. Listeners

Four normal-hearing adults (two female, two male) aged 24 to 41 years served as listeners. Three were paid for their participation. The fourth was the author (L4).

B. Stimuli and procedure

We used a modified version of the probe-signal procedure (Greenberg and Larkin, 1968; Dai *et al.*, 1991) to assess the detectability of tonal signals with expected and unexpected durations that were randomly presented at one of two expected frequencies in a continuous masking noise. In the *multiprobe condition*, we presented 12 different tonal signals in each block of 40 trials. Of these 12 signals, one of two expected target signals occurred on 30 trials (15 trials for each expected signal, in random order). Thus, as in a signal-uncertainty task, the listener expected one of two signals on each trial, but did not know which signal would be presented on any particular trial. For three listeners (L1–L3), the two target signals were a 4-ms, 1-kHz tone and a 299-ms, 4-kHz tone in one test, and a 299-ms, 1-kHz tone and a 4-ms, 4-kHz tone in another test. For the fourth listener (L4), a 0.25-kHz tone replaced the 1-kHz tone. On the remaining ten trials in each block, we presented ten unexpected probe signals on one trial each. These probe signals always had the same frequency as one or the other of the two target signals, but had one of five unexpected durations. The five unexpected durations were 7, 24, 86, 161, and 299 ms for the 4-ms target, and 4, 7, 24, 86, and 161 ms for the 299-ms target. Note that, within any given test, 4- and 299-ms signals occurred at both signal frequencies, but each served as the expected duration at one frequency and the unexpected duration at the other. A probe signal was randomly presented on one of every four successive trials, with the constraint that probe signals could not be presented on two consecutive trials. Only L4 knew of the presence of the probe durations in the multiprobe condition. Such awareness appears to have little to no effect on the results of probe-signal experiments (Scharf *et al.*, 1987; Dai *et al.*, 1991; Wright and Dai, 1994a). None of the other listeners reported hearing signals that had durations other than the expected one.

TABLE I. Mean signal level (dB SPL) for each signal required to yield a performance rate of approximately 90% correct detections. The number of listeners tested at each frequency (n) is listed below each frequency. The across-listener standard deviation for the 1 and 4 kHz signals averaged 1.2 dB.

	Signal duration (ms)					
	4	7	24	86	161	299
250 Hz ($n=1$)	67.0	63.5	54.0	47.0	45.0	44.0
1000 Hz ($n=3$)	67.8	64.0	51.7	49.0	44.7	44.7
4000 Hz ($n=4$)	65.4	62.6	56.9	52.0	50.4	49.5

For each listener, and each signal, we set the signal level to yield a performance rate of approximately 90% correct detections when only that signal was presented (Table I). These signal levels were determined by measuring the percentage of correct detections of a fixed-level signal in a 50-trial block, and adjusting the signal level between blocks until the listener detected the signal on ~90% of the trials. We checked performance at the established signal levels in the *probe-alone condition*.

The performance metric in both the multiprobe and probe-alone conditions was the percentage of correct responses for each signal. The task was two-interval forced-choice with feedback. The signal was presented randomly in either of two observation intervals, each of which was marked by a visual display on a computer screen. The listener indicated on a computer keyboard whether the signal occurred in the first or second interval, and was given visual feedback as to whether the response was correct or incorrect after every trial. The percent-correct detections in each of the two multiprobe conditions were based on 900 trials for each of the two expected target signals and on 60 trials for each of

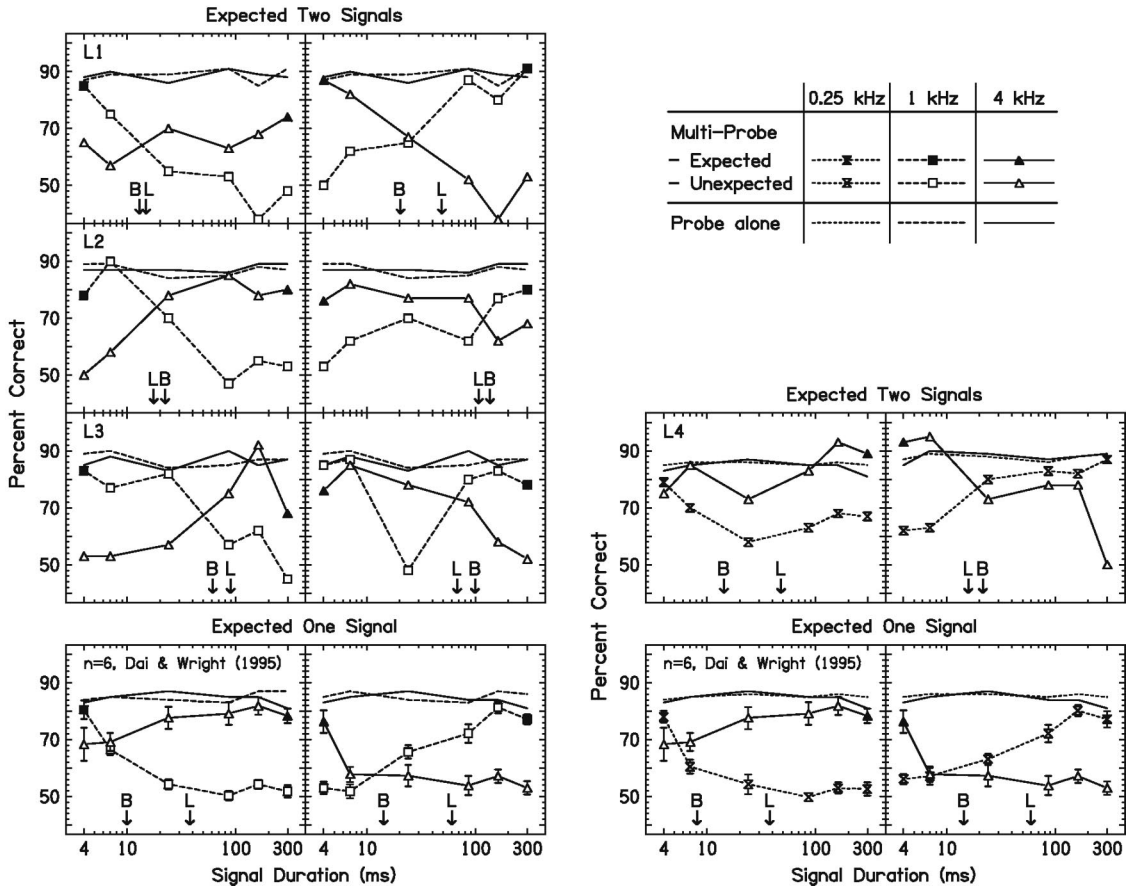


FIG. 1. Detectability of tonal signals of expected and unexpected durations, presented at one of two expected frequencies. Top panels (expected two signals) show the percentage of correct detections in the multiprobe conditions for signals of expected (filled symbols) and unexpected (open symbols) durations, measured with two different pairs of expected signals [brief, low frequency paired with long, high frequency (left column) and brief, high frequency paired with long, low frequency (right column)]. Also shown is performance in the probe-alone condition (lines without symbols). The signal frequencies were 1 kHz (squares, long-dashed lines) and 4 kHz (triangles, solid lines) for three listeners (L1–L3, left half), and 0.25 kHz (hourglasses, short-dashed lines) and 4 kHz (triangles, solid lines) for one listener (L4, right half). Bottom panels (expected one signal) show the mean results of six listeners who expected only a single signal, and were always presented signals at the expected frequency (from Dai and Wright, 1995). Error bars indicate plus and minus one standard error of the mean. Otherwise, as in the top panels. All panels indicate the individual (top panels) or mean (bottom panels) durations corresponding to $d'/2$ (arrows) for the brief (B) and long (L) expected signals.

the ten unexpected probe signals (data were pooled across trials per listener gathered from 60 blocks of 40 trials each). Performance in the probe-alone conditions was estimated from 150 trials (3 blocks of 50 trials each) for each of the 12 signals.

All signals were gated with 1-ms cosine-squared rise/fall envelopes. Reported durations were measured between the half-power points. The signals were generated digitally and played through a 16-bit digital-to-analog converter (Tucker Davis Technologies DD1) at a sampling frequency of 25 kHz, followed by an 8.5-kHz low-pass filter (TDT FT5). We obtained the continuous masker from a noise generator (TDT WG2) the output of which was passed through a low-pass filter with a cutoff frequency of 8.5 kHz (TDT FT5). The noise spectrum level was 25 dB SPL. The signals and noise were sent to separate programmable attenuators (TDT PA4), and then mixed (TDT SM3), before delivery to the left earpiece of Sennheiser HD450 earphones. All listeners were tested in a sound-attenuating room.

III. RESULTS

For each signal frequency, listeners detected signals of the expected duration on a greater percentage of trials than signals whose durations deviated markedly from the expected value (Fig. 1, top, expected two signals). Performance in the probe-alone condition (lines only) was quite uniform at about 87% correct detections, across the two signal frequencies and six signal durations used for each listener. In the multiprobe conditions (lines with symbols), the mean performance for the expected signals (filled symbols) was 5.6% lower than that in the probe-alone condition, consistent with other reports of the relatively small cost of signal uncertainty on detection (e.g., Green, 1961; Macmillan and Schwartz, 1975; Schlauch and Hafter, 1991; Dai and Wright, 1995). Most importantly, for each signal frequency, the detectability of the signals of unexpected duration (open symbols) typically decreased to near chance as the difference between the expected and unexpected duration, at that frequency, increased. In two exceptions out of 16 cases (L3, right panel; L4, left panel), the percent-correct detections for brief signals was equally high at both frequencies, but performance for the long signals was similar to that of other listeners.

To determine whether the present version of signal uncertainty affected the duration tuning revealed by perfor-

mance in the multiprobe condition, we compared the current results to those of a similar experiment in which listeners expected signals of only one frequency and duration (Dai and Wright, 1995; mean results shown in Fig. 1, bottom, expected one signal). In that experiment, signals of unexpected duration were also detected on progressively fewer trials as the disparity between the expected and unexpected durations grew. We based the comparison between the two experiments on half-detectability bandwidths for signal duration computed for each listener and expected signal (Greenberg and Larkin, 1968). We estimated these bandwidths by first converting the percentage of correct detections at each signal duration to d'. We next calculated $d'/2$ (equal to one-half of the maximum d' value) and then used the signal durations for the two d' values that flanked $d'/2$ to compute the linear interpolation of the signal duration for $d'/2$. On the assumption that d' and signal power are directly related, $d'/2$ corresponds to the half-power point (Greenberg and Larkin, 1968). The duration difference between the expected signal duration (4 or 299 ms) and the signal duration corresponding to $d'/2$ [Fig. 1, arrows, for brief (B) and long (L) expected signals] provided the half-detectability bandwidth. Note that, due to the distribution of unexpected signal durations, these half-detectability values represent the duration bandwidth only on the longer-duration side for the 4-ms expected signals, and only on the briefer-duration side for the 299-ms expected signals. The results would be influenced by this discrepancy if duration tuning were to prove to be asymmetrical around the expected duration. Asymmetries in template tuning have been reported (Dai and Buus, 1991; Wright and Dai, 1994b, 1998), but, to date, are the exception rather than the rule. We analyzed the logs of the half-detectability values with a 2 group (one versus two expected signals)×2 duration (4 versus 299 ms)×2 frequency (1 versus 4 kHz) analysis of variance (ANOVA). This analysis included data from the 3 two-signal listeners tested with signals at 1 kHz and 4 kHz (L1–L3) and from the 6 one-signal listeners tested with those same signals (Dai and Wright, 1995).[1]

The duration tuning was broader for the long than for the brief expected signals in both groups, and broader in the two-signal than the one-signal listeners for the brief, but not for the long, expected signals (Fig. 2). The ANOVA yielded significant main effects for group [$F(1,27)=15.65$, $p<0.0001$] and signal duration [$F(1,27)=148.38$, $p<0.0001$], and a significant group×duration interaction [$F(1,27)=17.31$, $p<0.0001$]. As confirmed by follow-up t-tests on the interaction, the half-detectability duration bandwidths were larger for the 299-ms (overall mean of 249 ms) than the 4-ms (overall mean of 31 ms) expected signals for both groups, indicating that the duration bandwidth is linked to the particular expected signal duration [one-signal: $t(10.8)=14.72$, $p<0.0001$; two-signal: $t(5.2)=4.29$, $p=0.007$]. This connection is evident in the relative bandwidths (duration bandwidth divided by the expected duration) as well, except that the relative bandwidths were consistently smaller for the long (overall mean of 0.84) than for the brief (overall mean of 6.11) expected signals. The absolute bandwidths were also significantly larger in the two-signal than the one-signal listeners, but only for the brief expected signals (average increase of 46 ms), showing a clear, though selective, influence of signal uncertainty on duration tuning [$t(15)=4.06$, $p<0.001$].

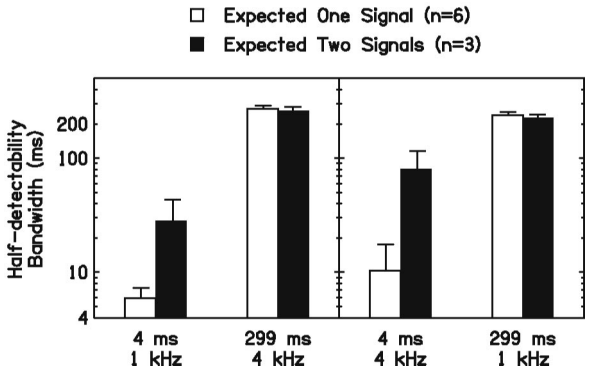


FIG. 2. Half-detectability bandwidths for duration tuning for listeners who were uncertain about both the frequency and duration of the signal to be presented, compared to listeners who had no uncertainty. Mean results are shown for the two different signal pairs [brief, low frequency paired with long, high frequency (left panel) and brief, high frequency paired with long, low frequency (right panel)] for six listeners who expected only one signal (white bars; data from Dai and Wright, 1995), and three listeners who expected two (black bars). Note that these values represent the duration bandwidth only on the longer-duration side for the 4-ms expected signals, and only on the briefer-duration side for the 299-ms expected signals.

IV. DISCUSSION

The present results are most consistent with the ideas that listeners monitor both the frequency and duration of a given expected signal through a single template, and that they attend to at least two templates, tuned to different durations, in duration-uncertainty conditions. Supporting this view, listeners who were uncertain as to which of two signals that differed in both frequency and duration was to be presented nearly always detected each of these expected signals on a greater percentage of trials than unexpected signals of markedly different duration (Fig. 1, top). Other outcomes, leading to different conclusions, were possible. For just two examples, detectability could have been high for both the brief and the long signals, but not for signals of intermediate duration, at both frequencies. That result would have indicated that the frequency and duration templates are separate, and that listeners attend to two different duration templates, each of which is tuned to the duration of an expected signal. Detectability also could have been high only for the brief signals or only for the long signals, at both frequencies. In that case, we would have concluded that there are separate templates for frequency and duration, and that listeners only attend to a single duration template under conditions of duration uncertainty.

The most important implication of the present results is that, in simple signal-detection tasks, listeners monitor a single template tuned to both the frequency and the duration of the expected signal, as opposed to listening through separate templates, one tuned to frequency and one tuned to duration. Thus, listeners appear to tap a representation in which the individual stimulus features have been combined, or bound together. This outcome need not have been the case,

because recent evidence from perceptual-learning experiments indicates that representations in which the frequency and duration of a stimulus are encoded separately can also influence behavior (Wright *et al.*, 1997; Wright, 1998; Delhommeau *et al.*, 2002). Interestingly, separate evidence for an attentional template tuned to both frequency and location (e.g., Mondor *et al.*, 1998) raises the possibility that the detection template observed here may be tuned to three stimulus attributes: frequency, duration, and location. It has yet to be determined whether the frequency-and-location or frequency-and-duration templates are also tuned to the expected presentation time of the signal, another stimulus attribute to which listeners are known to listen selectively based on expectations (Leis-Rossio, 1986; Chang, 1991; Wright and Fitzgerald, 2004).

The present results also indicate that listeners faced with the current version of signal-duration uncertainty attend selectively to the two expected stimulus durations, but do so through templates that are more broadly tuned for brief-duration signals than in the no-uncertainty condition. The evidence for duration-specific listening parallels that accepted for frequency-specific listening in frequency-uncertainty conditions (Macmillan and Schwartz, 1975; Howard *et al.*, 1986; Schlauch and Hafter, 1991; Hafter *et al.*, 1993). The broadening of the duration tuning of the brief-signal template is also consistent with several reports of a widening of the frequency tuning in cases of signal-frequency uncertainty (Schlauch and Hafter, 1991; Hafter *et al.*, 1993). However, the specificity of the broadening to a particular expected stimulus does not appear to have a counterpart in the frequency-uncertainty literature. There are at least three potential explanations for why the present uncertainty conditions only affected the duration tuning for the brief signal. One possibility is that uncertainty, as a rule, broadens tuning universally, but that this influence was obscured here by a ceiling effect in the bandwidth of the tuning for the long signal, which was already quite broad. Alternatively, uncertainty may specifically broaden the tuning on the longer-duration side of duration templates. Consistent with this account, broadening was observed when the unexpected signals were longer than (expected brief) but not when they were shorter than (expected long) the expected signal. It is also possible that template tuning under conditions of uncertainty is not affected uniformly for all expected signals, but rather is influenced by the particular combination of signals that is expected.

Finally, the current data extend the previous evidence that detection performance decreases only slightly when listeners are uncertain of the signal frequency or duration to a case in which there is uncertainty about both. The performance cost of listening for signals of a variety of frequencies (e.g., Macmillan and Schwartz, 1975; Schlauch and Hafter, 1991) or durations (Dai and Wright, 1995) typically reduces performance by only about 5% relative to when there is no uncertainty about the signal. The detectability of the expected signals in the multiprobe conditions here, which varied in both frequency and duration, also only decreased by an average of 6%. The lack of additivity of the influences of uncertainty about frequency with those about duration, which might have produced a ~10% decrease in performance, provides further evidence that both frequency and duration are encoded in the same signal-detection template.

ACKNOWLEDGMENTS

I collected these data at the University of California San Francisco and thank Dr. Michael Merzenich for his support at that time. I also thank Chris Stewart for assisting with data analysis as well as figure preparation, and Marshall Fong for technical support. Dr. Bertram Scharf and another, anonymous reviewer, as well as Matthew Fitzgerald, Julia Mossbridge, Jeanette Ortiz, and Yuxuan Zhang, provided helpful comments on earlier drafts of this paper. This work was supported by research grants from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health and the McDonnell-Pew Program in Cognitive Neuroscience.

[1]Across the three current two-signal listeners (total of four expected signals each) and six previous one-signal listeners (total of six expected signals each) included in the ANOVA, we were able to estimate the half-detectability bandwidth in all but three cases. All three involved the one-signal listeners. In two cases, the lowest d' never dropped below $d'/2$ when the expected signal was a 299-ms, 4-kHz tone, so the half-detectability bandwidth was set at 295 ms (the duration difference between the 299-ms expected signal duration and the shortest signal duration tested, 4 ms). In the remaining case (4-ms, 4-kHz), no value was assigned because d' was low at all durations.

Chang, P. (**1991**). "Temporal windows for signals presented at uncertain times," Unpublished master's thesis, University of Minnesota, Minneapolis.

Dai, H., and Buus, S. (**1991**). "Effect of gating the masker on frequency-selective listening," J. Acoust. Soc. Am. **89**, 1816–1818.

Dai, H., Scharf, B., and Buus, S. (**1991**). "Effective attenuation of signals in noise under focused attention," J. Acoust. Soc. Am. **89**, 2837–2842.

Dai, H., and Wright, B.A. (**1995**). "Detecting signals of unexpected or uncertain durations," J. Acoust. Soc. Am. **98**, 798–806.

Delhommeau, K., Micheyl, C., Jouvent, R., and Collet, L. (**2002**). "Transfer of learning across durations and ears in auditory frequency discrimination," Percept. Psychophys. **64**, 426–436.

Green, D.M. (**1961**). "Detection of auditory sinusoids of uncertain frequency," J. Acoust. Soc. Am. **33**, 897–903.

Greenberg, G.S., and Larkin, W.D. (**1968**). "Frequency-response characteristics of auditory observers detecting signals of a single frequency in noise: The probe-signal method," J. Acoust. Soc. Am. **44**, 1513–1523.

Hafter, E.R., Schlauch, R.S., and Tang, J. (**1993**). "Attending to auditory filters that were not stimulated directly," J. Acoust. Soc. Am. **94**, 743–747.

Howard, Jr., J.H., O'Toole, A.J., and Rice, S.E. (**1986**). "The role of frequency versus informational cues in uncertain frequency detection," J. Acoust. Soc. Am. **79**, 788–791.

Leis-Rossio, B.L. (**1986**). "Temporal specificity: Signal detection as a function of temporal position," Unpublished doctoral dissertation, University of Iowa, Iowa City.

Macmillan, N.A., and Schwartz, M. (**1975**). "A probe-signal investigation of uncertain-frequency detection," J. Acoust. Soc. Am. **58**, 1051–1058.

Mondor, T.A., Zatorre, R.J., and Terrio, N.A. (**1998**). "Constraints on the selection of auditory information," J. Exp. Psychol. Hum. Percept. Perform. **24**, 66–79.

Scharf, B., Quigley, S., Aoki, C., Peachey, N., and Reeves, A. (**1987**). "Focused auditory attention and frequency selectivity," Percept. Psychophys. **42**, 215–223.

Schlauch, R.S., and Hafter, E.R. (**1991**). "Listening bandwidths and frequency uncertainty in pure-tone signal detection," J. Acoust. Soc. Am. **90**, 1332–1339.

Wright, B.A. (**1998**). "Generalization of auditory-discrimination learning," Assoc. for Res. in Otolaryngol. Abstr., Abstract 413, 104 (A).

Wright, B.A., and Dai, H. (**1994a**). "Detection of unexpected tones with short and long duration," J. Acoust. Soc. Am. **95**, 931–938.

Wright, B.A., and Dai, H. (**1994b**). ''Detection of unexpected tones in gated and continuous maskers,'' J. Acoust. Soc. Am. **95**, 939–948.
Wright, B.A., and Dai, H. (**1998**). ''Detection of sinusoidal amplitude modulation at unexpected rates,'' J. Acoust. Soc. Am. **104**, 2991–2996.
Wright, B.A., and Fitzgerald, M.B. (**2004**). ''The time course of attention in a simple auditory detection task,'' Percept. Psychophys. **66**, 508–516.
Wright, B.A., Buonomano, D.V., Mahncke, H.W., and Merzenich, M.M. (**1997**). ''Learning and generalization of auditory temporal-interval discrimination in humans.'' J. Neurosci. **17**, 3956–3963.
Yama, M., and Robinson, D. (**1982**). ''Comparison of frequency selectivity for the monaural and binaural hearing systems: Evidence from a probe-frequency procedure,'' J. Acoust. Soc. Am. **71**, 694–700.

Detection-theoretic analysis of same–different judgments for the amplitude discrimination of acoustic sinusoids[a)]

Miao-Fen Wang,[b)] R. John Irwin,[c)] and Michael J. Hautus[d)]
Department of Psychology, University of Auckland, Private Bag 92019, Auckland, New Zealand

(Received 1 February 2004; revised 2 December 2004; accepted 3 December 2004)

Three methods were applied to determine which decision model best accounted for same–different judgments about the amplitude of acoustic sinusoids. The methods were (1) analysis of the shape of the receiver operating characteristic; (2) analysis of an observer's decision space; and (3) a correlation method based on the conditional-on-single-stimulus procedure. In one experiment, observers rated their confidence that a pair of 1-kHz sinusoids was drawn from the same or from different sets. The two sets had identical Gaussian distributions of decibel amplitudes, but their mean amplitudes differed by 3 dB. The differencing model, in which observers base their decision on the absolute difference between the two observations, accounted for the data better than the independent-observation model, in which observers compute independent likelihood ratios for each observation. A second experiment added to every trial an interval that contained a 1-kHz sinusoid whose amplitude equaled the grand mean of both stimulus sets. Despite this additional information, which is needed to adopt the independent-observation model, the differencing model again better accounted for the data. [DOI: 10.1121/1.1853272]

PACS numbers: 43.66.Fe [NFV] Pages: 1305–1313

I. INTRODUCTION

Two detection-theoretic models, the independent-observation and the differencing models, have been developed to account for an observer's decision process for same–different judgements (see Noreen, 1981, or Macmillan and Creelman, 1991, for a review). A complication is that the two models yield discrepant estimates of an observer's ability to distinguish between two similar stimuli, measured by the accuracy index d'. The objective of this research is to determine, by means of three methods, which model best accounts for an observer's judgments about the amplitude of acoustic sinusoids. The application of the three methods in this research is also evaluated.

A. Elements of the same–different task and its decision models

For a standard version of the same–different experiment, every trial contains two independent observation intervals, separated either in time or space. Each of two stimuli, e.g., *A* or *B*, normally has an equal probability of occurring in each observation interval. Therefore, on each trial, there is an equal probability of four possible stimulus arrangements; i.e., $\langle AA\rangle$, $\langle BB\rangle$, $\langle AB\rangle$, or $\langle BA\rangle$. Observers are required to decide whether the two stimuli are the same or different. To simplify detection-theoretic analysis, we make the standard detection-theory assumptions: (1) the two observations on each trial are independent of each other; and (2) the underlying distributions associated with each stimulus class, *A* and *B*, are Gaussian and are of equal variance.

To elucidate the two decision models, let $x=\langle x_1,x_2\rangle$ denote the standardized observations (i.e., measured in units of standard deviation) in the first and second intervals, respectively, of each trial. The independent-observation model assumes the observer compares the likelihood of a pair of observations given the "same" type of trial (i.e., $\langle AA\rangle$ or $\langle BB\rangle$) against the likelihood of the same pair of observations given the "different" type of trial (i.e., $\langle AB\rangle$ or $\langle BA\rangle$). The likelihood ratio of x on which the observer bases a decision, is usually denoted by $L(x)$ and is given as (see Noreen, 1981; Dai, Versfeld, and Green, 1996)

$$L(x)=L(x_1,x_2)=\frac{e^{d'x_1}+e^{d'x_2}}{1+e^{d'(x_1+x_2)}}.$$

The criterial value of $L(x)$, usually denoted β, equals 1 for unbiased responses (i.e., not favoring one response over the other). This unbiased likelihood-ratio decision rule has proved equivalent to a categorical decision rule (Noreen, 1981), by which the observer independently categorizes each observation and then compares their categories. The categorization can be based on likelihood ratios for each observation arising from *A* against *B*, or it can be based on sensory strength with the criterial value equal to the grand mean of all possible observed values (Noreen, 1981; Dai *et al.* 1996). Noreen (1981) claimed that this equivalence holds only for the unbiased case. However, the independent-observation rule based on the likelihood ratio of x with a criterial value, β, can be proved to be equivalent to the categorical rule with a decision criterial value,[1] $\pm[\ln(\beta)]/d'$ (see Ross, 1998, pp. 257 and 303; cf. Irwin and Hautus, 1997)

[a)]Portions of this work were presented in "Decision processes used by human observers in an auditory same-different task," Abstracts of the 29th Australasian Experimental Psychology Conference, Adelaide, Australia, April, 2002, and "Three methods for determining decision strategies in an auditory same–different task," Proceedings of the 18th Annual Meeting of the International Society for Psychophysics, Rio de Janeiro, Brazil, October, 2002.

[b)]Electronic mail: m.wang@auckland.ac.nz
[c)]Electronic mail: j.irwin@auckland.ac.nz
[d)]Electronic mail: m.hautus@auckland.ac.nz

Respond R_1 (or "1"), if and only if both x_1 and x_2 are greater than
$\ln(\beta)/d'$ *or both are less than* $-\ln(\beta)/d'$;
otherwise respond R_0 (or "0").

The general form of the categorical rule introduced here and its equivalence to the independent-observation rule explain how the categorical rule can be coupled with the confidence-rating method (see later) to trace out ROC (receiver operating characteristic) curves, using the same parametric equations as those for the independent-observation rule.

On the other hand, the differencing model assumes the observer computes the difference between the two observed values, x_1 and x_2, and compares the difference to a criterion, k. Note that x_1, x_2, and k are all standardized and measured in units of standard deviation.

B. Optimal decision rules

For uncorrelated observations, as assumed in this paper, the independent-observation rule is the optimal rule to maximize correct responses and minimize errors (Dai *et al.*, 1996; Noreen, 1981), while for highly correlated observations, the differencing rule is optimal (Dai *et al.*, 1996). Irwin and his colleagues (1993, 1995, and 1997) have shown that independent-observation judgments are difficult, or even impossible, to make for simple sensory stimuli, like the amplitudes of 1-kHz sinusoids (Hautus, Irwin, and Sutherland, 1994) or the concentrations of orange-flavored drinks (Irwin *et al.*, 1993). However, the differencing model, which is not the optimal model for the case, accounted well for the observers' judgments about these simple sensory stimuli (cf. Irwin and Francis, 1995; Francis and Irwin, 1995). Three methods are used in experiment 1 to examine this issue more fully.

Macmillan and Creelman (1991) have indicated that knowledge of the distributions of stimuli is necessary for an observer to adopt the independent-observation rule; and Dai *et al.* (1996) have shown that knowledge of the grand mean value of the distributions is necessary for adopting the unbiased categorical rule—a simple form of the independent-observation rule. Therefore, in addition to investigating the decision model for same–different judgments about the amplitudes of individual acoustic sinusoids, this research also investigates whether an observer can be oriented to make independent-observation judgments (i.e., optimal judgments) about simple sensory stimuli of this sort. In experiment 2 we endeavor to do this by familiarizing the observer with the distributions of stimuli and with their grand mean value.

We next describe three methods for determining the decision process adopted, which is a prerequisite for a valid estimate of d'.

C. Shape analysis of the ROC

The ROCs obtained from the differencing and independent-observation models have fundamentally different shapes. The differencing model predicts asymmetric ROCs about the minor diagonal of the unit square (Sorkin, 1962), while the independent-observation decision model predicts symmetric ROCs about this diagonal (Noreen, 1981). The shape of an empirical ROC can therefore be used to infer the decision strategy adopted by an observer (e.g., Hautus, Irwin, and Sutherland 1994; Irwin and Francis, 1995). The parametric equations of the ROC for the differencing model have been provided by Macmillan *et al.* (1977), and those for the independent-observation models have been provided by Irwin and Hautus (1997).

D. Two-dimensional decision-space analysis

The observed values of the two stimuli on each trial can be represented by one point in a two-dimensional space, in which one dimension represents the observation in the first interval, and the other dimension represents the observation in the second interval. The differencing rule partitions the decision space by two straight lines with slope=1 and intercept=±criterion difference, k, measured in units of standard deviation. The independent-observation rule partitions the decision space by two lines (or curves) which represent a constant likelihood ratio of a pair of observations given the "same" type of trial relative to the same pair of observations given the "different" type of trial. For a particular criterion β, the criterion boundary can be described by the equation

$$x_2 = h(x_1) = \frac{1}{d'} \ln\left(\frac{\beta - e^{d'_{x_1}}}{1 - \beta \cdot e^{d'_{x_1}}} \right).$$

For the unbiased case (i.e., $\beta=1$), the criterion boundary is defined by the axes (Dai, Versfeld, and Green, 1996; Irwin and Hautus, 1997).

E. The correlation method

Dai *et al.* (1996) suggested a correlation method to contrast the independent-observation and the differencing models[2] (cf. Thorndike, 1978). This method originates from the conditional-on-a-single-stimulus (COSS) procedure (Berg, 1989), integrated with correlational analysis suggested by Richards and Zhu (1994) and Lutfi (1995). To implement the correlation method, let each observation variable take on the value of a physical element of the presented stimulus in the ith observation interval on each trial, and be standardized and denoted x_i. The observer's behavioral response, R, can then be related to multiple observation variables by a linear combination of these observation variables, e.g., $\alpha_1 x_1 + \alpha_2 x_2$, where α_1 and α_2 are the combination weights. The relation between a physical stimulus variable, e.g., intensity or frequency, and the corresponding psychological observation variables, x_i, should be assumed linear, at least locally[3] (Johnson, 1980; Wang, Irwin, and Hautus, 1998; Macmillan and Creelman, 1991, pp. 212–216), to comply with the correlation method (Richards and Zhu, 1994).

When applying the correlation method to contrast the differencing and the independent-observation models, we assume that the weights an observer gives to x_1 and x_2 are ± 1 when adopting the differencing rule. On the other hand, we assume that the weights for x_1 and x_2 are both 1 for the categorical rule (i.e., another form of the independent-

observation rule). Richards and Zhu's (1994) theorem states that the correlation between the binary response variable, R, and each observation variable, x_i, is proportional to the linear combination weight, α_i. Thus, it is predicted that the two decision models give different parametric distributions of correlations, r_{Rx_i}, where r_{Rx_i} denotes the correlation coefficient between R and x_i. For each type of trial (i.e., $\langle AA\rangle$, $\langle AB\rangle$, $\langle BA\rangle$, or $\langle BB\rangle$) an observer's judgments can be simulated by a Monte Carlo method based on the differencing and the independent-observation models. Then, for each type of trial, two correlation distributions can be obtained from the Monte Carlo simulation for each decision model, respectively (see Dai *et al.*, 1996 for a simulation example). The two distributions are separated by a boundary line with slope -1 for $\langle AA\rangle$ and $\langle BB\rangle$, and slope 1 for $\langle AB\rangle$ and $\langle BA\rangle$ (Dai *et al.*, 1996; Thorndike, 1978). Both two-dimensional distributions are at first projected onto the line $y=x$ for $\langle AA\rangle$ and $\langle BB\rangle$ and onto the line $y=-x$ for $\langle AB\rangle$ and $\langle BA\rangle$, to form new one-dimensional Gaussian distributions. To optimize the decision whether the pair of empirical correlations is from the differencing or the independent-observation distribution, the boundary line is located where the likelihood ratio of the two newly formed one-dimensional distributions equals 1. The separation between the two distributions is indicated by d_a, which equals the difference of the means of the two distributions divided by the root-mean-square of their variances.

Note that this application of COSS analysis differs from Berg's (1989) COSS analysis in the way that the internal noise and criterion variability can somewhat reduce the accuracy of estimates of correlation coefficients used to obtain estimates of weights[4] (Richards and Zhu, 1994). Berg's (1989) COSS analysis estimates weight ratios, α_1/α_2, by regressing response probabilities on the perturbations added to x_i. The estimated weight ratios have been shown to be independent of all additive sources of internal noise.

We now investigate the application of these three methods for the case of discriminating the amplitudes of acoustic sinusoids with the same–different task.

II. EXPERIMENT 1

A. Methods

1. Observers

Three paid observers (GS, NB, and FG) with normal hearing participated. Normal hearing was assessed by no hearing loss in excess of 15 dB (re: I.S.O standard, 1975) at any audiometric frequency over 500 Hz to 6 kHz on a Békésy-tracking task with a Brüel & Kjær audiometer (type 1800). GS, a graduate student, was experienced in psychoacoustic tasks, but NB and FG, both undergraduate students, were inexperienced observers.

2. Stimuli, apparatus, and procedure

The stimuli were 1-kHz sinusoids of 100-ms duration, generated by a programmable synthesizer (Hewlett Packard 8904A) and shaped by a Hanning ($\cos^2$) window with 10-ms rise and fall times to reduce transients. The sound level of each sinusoid was controlled by a programmable attenuator (TDT, PA4) with a resolution of 0.1 dB. Two Gaussian distributions of sound levels were selected to form two stimulus sets. One distribution, denoted by *A*, had an average sound level at the earphone of 50 dB SPL and the other, denoted by *B*, of 53 dB SPL. Both distributions had a standard deviation of 1.5 dB, so $d'=2$ for an ideal observer. It was expected that the external Gaussian perturbations were relatively large and would therefore swamp the observer's internal noise (assumed independent and additive), so that observers would make judgments according to the physical values of stimuli but not according to internal fluctuation (see Lee, 1963; Kubovy, Rapoport, and Tversy, 1971; Ward, 1973; and Ashby and Gott, 1988, for this technique).

On each trial, the level of the stimulus in each interval was drawn at random from either distribution *A* or *B*. Therefore, the two levels in each interval were not actually the same when drawn from the same distribution, and could be nearly equal when drawn from different distributions. Prior to the experiment, all observers were informed about the stimuli and their distributions, as just described. They were also informed about the possible decision strategies. One strategy is to first categorize each tone of each trial and then decide they were from the same or different distributions; the other is to make a decision based on the difference in magnitude between the two tones.

In an experimental session, the observer was seated in a sound-attenuating booth (Amplaid, model E) in front of a computer terminal. The stimuli were presented monaurally by means of earphones (Telephonics, TDH-39P). The terminal's screen presented the instructions and its keyboard served as a response panel.

An observer's task was to make a same–different judgment on each trial about two 1-kHz sinusoidal stimuli separated by a 500-ms silent interval. The judgments were made using a six-point confidence-rating scale. Rating 1 corresponded to "very confident that the two stimuli were from different stimulus classes." Rating 6 corresponded to "very confident that the two stimuli were from the same class." Ratings were entered using the numeric keys from 1 to 6 of the keyboard.

Each observer received at least 2000 practice trials with trial-by-trial feedback. Data analysis was based on 1000 trials without trial-by-trial feedback (see Robinson and Watson, 1972, for a review about the advantages and disadvantages of providing trial-by-trial feedback).

B. Results

1. Shape analysis of the ROC

The parametric equations of the ROC provided by Macmillan *et al.* (1977) and Irwin and Hautus (1997) for each decision model were used to fit the rating data. We used a maximum-likelihood method, similar to those reported by Dorfman and Alf (1969) for the normal–normal model of the yes–no task, to estimate the best-fitting accuracy index d' and decision criteria (k or β); and the goodness-of-fit statistic χ^2 was also provided. Table I presents the maximum-likelihood estimates of d', based on each decision model, together with the goodness-of-fit statistic, χ^2 (df=4), and its

TABLE I. Maximum-likelihood estimates of d', based on the differencing and the independent-observation models, for observers GS, NB, and FG in experiment 1. The associated obtained values of the goodness-of-fit statistic, χ^2 (df=4), and probability, p, are also reported.

	Differencing			Independent observation		
Observer	d'	χ^2	p	d'	χ^2	p
GS	1.8	1.05	0.903	1.4	20.14[a]	0.000
NB	1.6	0.38	0.984	1.3	8.71	0.069
FG	1.6	8.93	0.063	1.3	4.53	0.340

[a]$p<0.01$.

probability, p. For observer GS, the differencing model provides a good fit to the rating judgments at $p<0.01$, while the independent-observation model does not. For the other two observers, NB and FG, the ROC analysis does not distinguish between the two decision models.

2. Decision space analysis

Two approaches to decision-space analysis are adopted in this research. One approach constructs an observer's internal decision space using the observer's accuracy index, d'. The observed value of a stimulus represented in the decision space is a linear z-score transformation of the physical magnitude of the stimulus in terms of the maximum-likelihood estimates of d' and the decision criteria (k or β) from ROC analysis. This approach assumes the observer's internal noise and unstable criteria interfere with the discriminability of the two stimulus classes. For experiment 1, the value of d' for an ideal observer is 2. However, as ROC analysis has shown, for all observers the estimated values of d' are less than 2. For a human observer, an additional source of variability could be introduced to the underlying distributions, because of internal noise and unstable criteria. This reduces the estimate of the observer's accuracy (Green and Swets, 1966; Wickelgren, 1968). In these cases, given the expected mean physical magnitude of each stimulus class, denoted by M_A and M_B, respectively, and given the maximum-likelihood estimate of d', the variability of the underlying distributions was estimated by dividing $M_A-M_B=3$ dB by d'. Therefore, a linear transformed value equals the distance in standard deviation units from the physical magnitude of, say, a stimulus A to the expected mean physical magnitude of that stimulus, plus $-d'/2$ (or $+d'/2$ for stimulus B).

Figure 1 shows an example of the partitioning pattern of the decision space for observer GS's same–different judgments, which provides a preliminary indication of the decision strategy the observer might have adopted. Figure 1(a) shows the partitioning pattern of the observer's decision space, assuming the differencing model, and Fig. 1(b) shows the partitioning pattern assuming the independent-observation model. Table II(a) shows the proportion of the same–different judgments made by each observer that is correctly accounted for by the decision criteria estimated from ROC analysis, based on each decision model. The ROC analysis is the same as that discussed in Sec. II B 1. As shown in this table, for each observer the differencing model appears to better account for individual judgments than the other model.

An alternative approach to decision-space analysis, which can be implemented independently of ROC analysis, assumes that the observer's true accuracy, d', equals 2 and that reduced estimates of sensitivity are due to criterion variation. This approach investigates how the decision space, defined by the physical values of the stimuli, is segregated into different response regions. Using this approach, an ob-

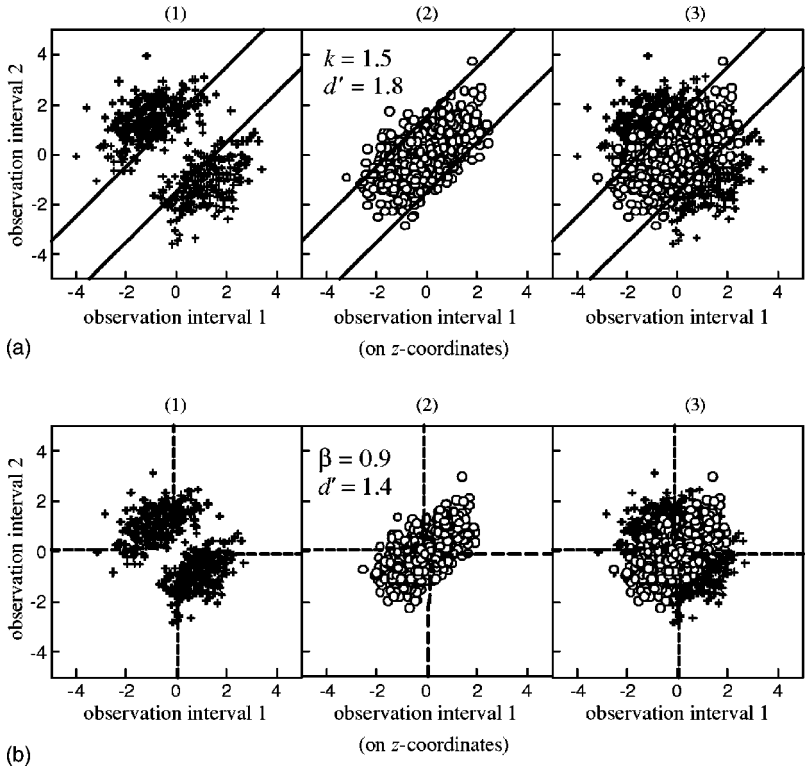


FIG. 1. (a) and (b) show how observer GS's judgments, ''same'' or ''different,'' partition the two-dimensional decision space. Two sets of lines illustrate the maximum-likelihood estimates of the decision criteria based on the differencing (solid lines) and the independent-observation (broken lines) models. GS's judgments of ratings 1, 2, and 3 have been defined as ''different'' (crosses) and ratings 4, 5, and 6 as ''same'' (circles). Panel (1) shows ''different'' judgments, (2) shows ''same'' judgments, and (3) shows both types of judgment combined. The symbols represent the pairs of standardized observation values, based on $d'=1.8$ in (a) and $d'=1.4$ in (b).

TABLE II. (a) Proportion of judgments made by each observer in experiment 1 that is correctly accounted for by the best-fitting decision criteria estimated by ROC analysis, based on the differencing or the independent decision model. (b) Maximum proportion of judgments made by each observer in experiment 1 that is correctly accounted for by the decision criteria determined by the iterative method based on the differencing or the independent-observation decision model, assuming the observer's accuracy, d', equals 2.

		Differencing			Independent observation		
	Observer	d'	k	$P(c)$	d'	β	$P(c)$
(a)	GS	1.8	1.5	0.844	1.4	0.9	0.786
	NB	1.6	1.2	0.715	1.3	1.0	0.687
	FG	1.6	1.4	0.738	1.3	0.9	0.719
(b)	GS	2.0	1.9	0.855	2.0	0.5	0.789
	NB	2.0	1.8–1.9	0.735	2.0	0.6	0.715
	FG	2.0	1.5–1.6, or 1.8–1.9	0.740	2.0	0.6	0.721

server's criteria were determined by an iterative method in which the value of k or β started from a low value around zero, and then was gradually increased by 0.01 on each iteration.

The proportion of an observer's judgments that can be accounted for by each value of k or β was recorded and the criterial value that can best account for the observer's judgments was then found. Table II(b) shows the maximum proportion of the same–different judgments made by each observer that is correctly accounted for by the decision criteria determined by the iterative method, based on the differencing and the independent-observation models, respectively. Note that using the iterative method, the criterial value that can best account for each observer's judgments is not well defined. Again, for each observer the differencing model appears to better account for the same–different judgments than the other model. These results, like those of the first approach that uses the parametric estimates from ROC analysis to construct the decision space, favor the differencing model for all observers.

3. The COSS method

Figures 2 and 3 show empirical correlations and the patterns of the simulated correlation distributions for observers GS and NB, based on each decision model. The two simulated correlation distributions for each type of trial were based on the observer's best-fitting d' and same–different decision criteria from ROC analysis for each decision rule. The index d_a indicates the separation of the two simulated distributions and equals the difference of the means of the two distributions divided by the root-mean-square of their variances.

As shown in Fig. 2, the three pairs of correlations calculated from observer GS's empirical data for $\langle AA\rangle$, $\langle BB\rangle$, and $\langle BA\rangle$ trials fall in the regions of the differencing distribution. The pair of empirical correlations for $\langle AB\rangle$ trials falls on the other side of the optimal boundary separating the two distributions, but very close to the boundary. The COSS pattern indicates that the differencing model tends to prevail for GS. For NB, the differencing model appears to slightly prevail overall, by a narrow margin (see Fig. 3). This is because the individual COSS patterns for $\langle AA\rangle$ and $\langle BB\rangle$ favor the differencing model and that for $\langle BA\rangle$ favors the other model, but which model the COSS pattern for $\langle AB\rangle$ favors is uncertain. For FG, no decisive conclusion can be drawn from the COSS method, because the individual COSS patterns for $\langle AA\rangle$ and $\langle BB\rangle$ favor the differencing model but those for $\langle AB\rangle$ and $\langle BA\rangle$ favor the independent-observation model.

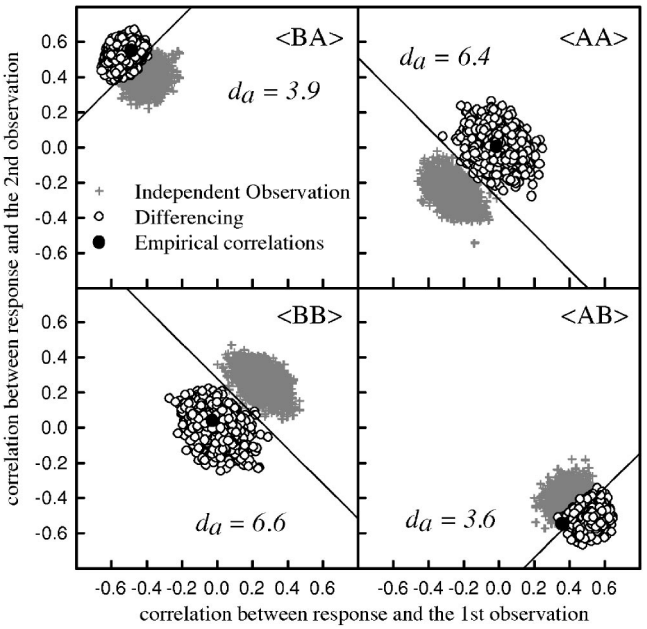


FIG. 2. Empirical correlations for observer GS and the simulated correlation distributions based on the independent-observation and the differencing decision models (d_a: the separation of the two simulated correlation distributions).

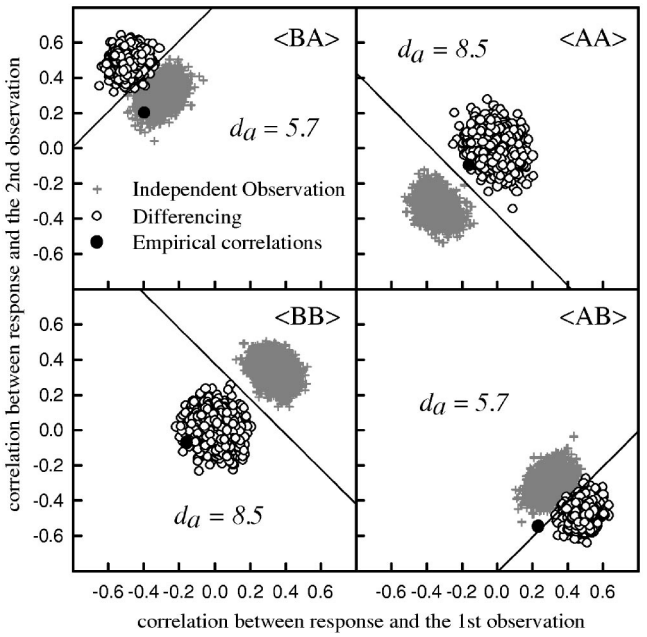


FIG. 3. Empirical correlations for observer NB and the simulated correlation distributions based on the independent-observation and the differencing decision models (d_a: the separation of the two simulated correlation distributions).

C. Discussion

For observer GS, the results from all three methods for determining the decision models suggest that the observer adopted the differencing decision strategy for the amplitude discrimination of single sinusoids. For observers NB and FG, shape analysis of the ROC and COSS analysis does not clearly discriminate between the two decision models. The results from decision-space analysis, however, appear to favor the differencing model over the other. For all three observers, the results from decision-space analysis tend to favor the differencing model over the other.

Note that all three methods assume that the observer consistently adopts one of the two decision strategies (see Hautus, Irwin, and Sutherland, 1994; Irwin and Francis, 1995; Francis and Irwin, 1995 for a justification of this assumption). However, human observers may intermingle the two decision rules or adopt an even more complicated rule (Dai *et al.*, 1996). A modest conclusion is that the observers generally tended to adopt the differencing rule for the amplitude discrimination of acoustic single sinusoids. This conclusion accords with the results reported by Hautus *et al.* (1994), who conducted a similar experiment. Hautus *et al.*'s experiment, however, had no perturbation added to each type of stimulus, so the decision strategy cannot be disclosed from decision-space analysis and also cannot be investigated with the COSS method. The results of experiment 1 show a consistent trend with the results reported by Hautus *et al.* (1994), and thus show, to some extent, the validity of the three methods.

It is worth noting that shape analysis of the ROC coupled with the rating method can be complicated by the unknown effects of unstable criteria on each category rating and thus affect the estimated values of d'. This in turn affects the reliability of our application of COSS analysis, because we use these parametric estimates from ROC analysis to simulate the correlation distributions. In addition, the correlation distributions were simulated assuming stable decision criteria, but the empirical correlation points could be volatile, because of unstable criteria. Unstable criteria reduce the accuracy of the estimated correlation coefficients (Richards and Zhu, 1994).

In the present experiment, the imprecise information the observers had about the distributions of the stimuli and their grand mean might have contributed to the observers' adoption of the differencing rule instead of the independent-observation rule, which is the optimal rule for uncorrelated observations. The next experiment investigates whether an observer can be oriented to make optimal judgments by providing the observer with knowledge of the distributions of stimuli and with the reference value (i.e., the grand mean value of the distributions). Although it is difficult to be sure that an observer has sufficient knowledge of the distributions of stimuli, many practice trials were provided to familiarize the observer with the distributions, as in experiment 1.

TABLE III. Maximum-likelihood estimates of d' for observers LG, TG, and ZL in experiment 2, based on the differencing and the independent-observation models. The associated obtained values of the goodness-of-fit statistic, χ^2 (df=4), and probability, p, are also reported.

	Differencing			Independent observation		
Observer	d'	χ^2	p	d'	χ^2	p
LG	1.4	6.61	0.158	1.1	20.08[a]	0.000
TG	1.8	10.44	0.033	1.5	1.04	0.903
ZL	1.6	12.84	0.012	1.2	21.59[a]	0.000

[a] $p<0.01$.

III. EXPERIMENT 2

A. Methods

1. Observers

Three paid graduate students (LG, TG, and ZL) with normal hearing participated. None had previous experience in psychoacoustic tasks.

2. Stimuli, apparatus, and procedure

The stimuli, apparatus, and procedure were the same as for experiment 1, except that there were three observation intervals on each trial instead of two. Each trial contained one information interval that preceded two stimulus-presentation intervals. The information interval of each trial contained a 1-kHz sinusoid of 100-ms duration with a stimulus level equal to the grand mean of distributions *A* and *B*. Each observer received at least 2000 practice trials with trial-by-trial feedback. Data analysis was based on 800 trials without trial-by-trial feedback.

B. Results

1. Shape analysis of the ROC

Table III shows the maximum-likelihood estimates of the accuracy index, d', and their obtained χ^2 (df=4) and probabilities, p, for each observer for each model. For both LG and ZL, the ROCs based on the differencing model provide good fits at $p<0.01$, while those based on the independent-observation model do not. For TG, neither of the obtained χ^2 (df=4) for the differencing or the independent-observation models is significant at $p<0.01$.

2. Decision space analysis

As in experiment 1, two approaches to decision-space analysis were undertaken: one constructed the decision space using the parametric estimates from the ROC analysis, and the other was undertaken independently of ROC analysis. The results from both approaches show that, for all observers, the differencing model accounts for their same–different judgments consistently better than the other model (see Table IV).

3. The COSS method

The results from COSS analysis generally favor the differencing model over the independent-observation model. As shown in Fig. 4, for LG, the empirical correlations for $\langle AA \rangle$ and $\langle BB \rangle$ trials clearly fall in the region of the differencing

TABLE IV. (a) Proportion of judgments made by each observer in experiment 2 that is correctly accounted for by the best-fitting decision criteria estimated by ROC analysis, based on the differencing or the independent-observation decision model. (b) Maximum proportion of judgments made by each observer in experiment 2 that is correctly accounted for by the decision criteria determined by the iterative method based on the differencing or the independent-observation decision model, assuming the observer's accuracy, d', equals 2.

		Differencing			Independent observation		
	Observer	d'	k	$P(c)$	d'	β	$P(c)$
(a)	LG	1.4	1.5	0.788	1.1	0.9	0.743
	TG	1.8	1.2	0.761	1.5	1.1	0.650
	ZL	1.6	1.2	0.749	1.2	1.0	0.708
(b)	LG	2.0	2.1–2.2	0.790	2.0	0.3	0.741
	TG	2.0	1.3–1.4	0.761	2.0	0.7	0.723
	ZL	2.0	1.3–1.4, or 1.8–1.9	0.751	2.0	0.9	0.716

correlation cluster, whereas the other empirical correlations for $\langle AB \rangle$ and $\langle BA \rangle$ trials fall almost on the boundaries of the optimal region. The whole COSS pattern for LG appears to favor the differencing model more than the other model. For TG, all but the pair of correlations for $\langle BA \rangle$ trials fall on the differencing side of the optimal boundary (see Fig. 5). The whole COSS pattern, again, appears to favor the differencing model more than the other model. For ZL, the pair of empirical correlations for $\langle AA \rangle$ and $\langle BB \rangle$ trials clearly falls in the region of the differencing correlation cluster, whereas the pair of empirical correlations for $\langle AB \rangle$ trials falls on the other side of the optimal boundary and the pair of empirical correlations for $\langle BA \rangle$ trials falls almost on the optimal boundary (see Fig. 6).

C. Discussion

As shown above, although the results from ROC analysis generally favor the differencing model, the ROC analysis for TG can reject neither model at the 0.01 level. Interestingly, the ROC analysis for TG's judgments shows a large discrepancy between the two probability levels of the obtained χ^2 (df=4) for each decision model. The probability level of the obtained χ^2 (df=4) is 0.03 for the differencing model and 0.90 for the independent-observation model. Decision-space analysis provides some insight into this.

An advantage of decision-space analysis is that it can reveal whether the symmetric assumption holds. The symmetric assumption made to simplify ROC analysis for both the independent-observation model and the differencing

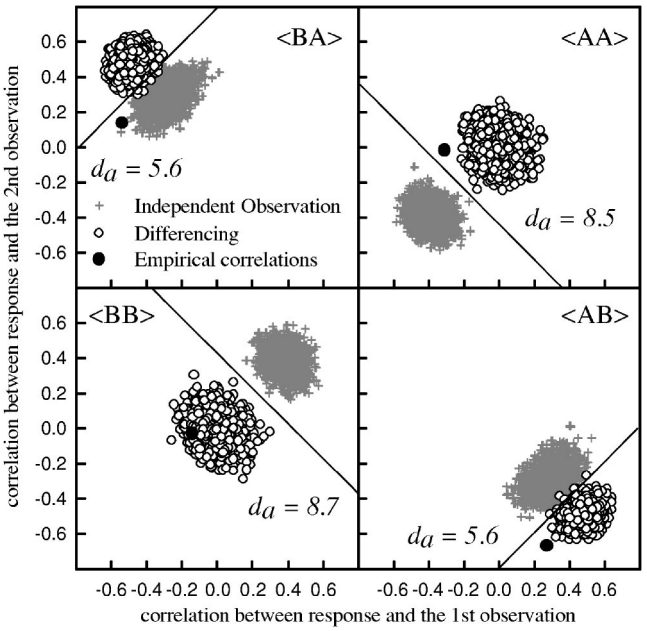


FIG. 5. Empirical correlations for observer TG and the simulated correlation distributions based on the independent-observation and the differencing decision models.

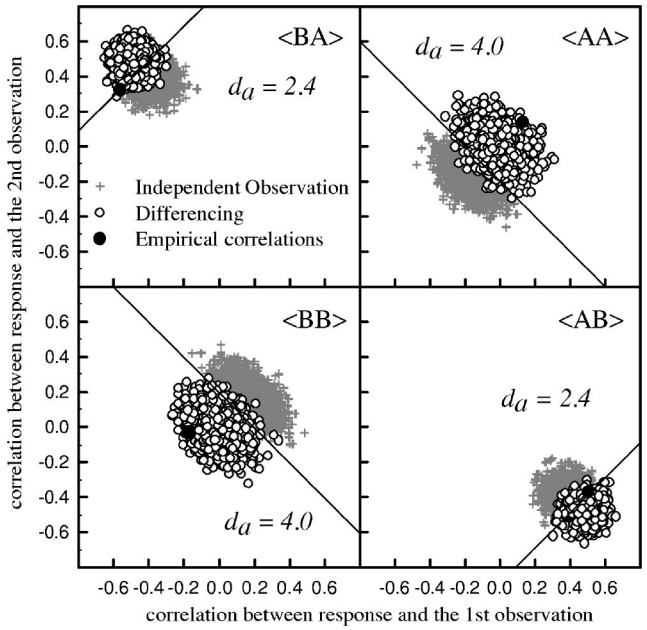


FIG. 4. Empirical correlations for observer LG and the simulated correlation distributions based on the independent-observation and the differencing decision models.

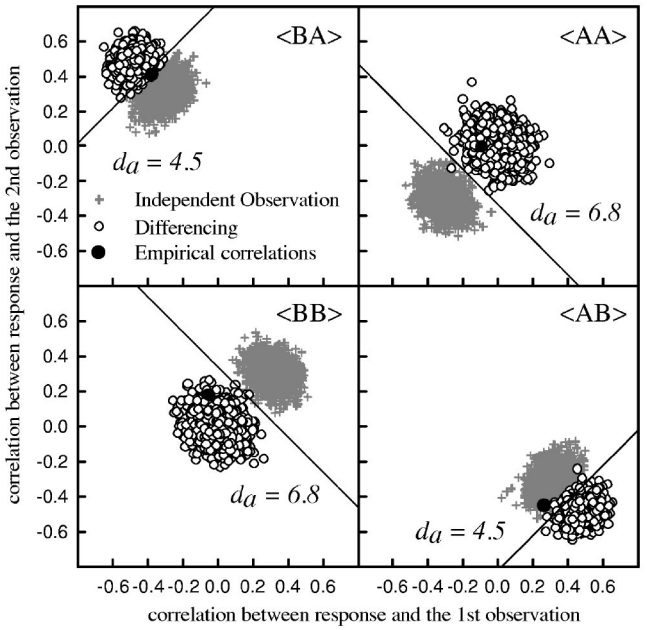


FIG. 6. Empirical correlations for observer ZL and the simulated correlation distributions based on the independent-observation and the differencing decision models.

TABLE V. The asymmetry of each observer's same–different judgments in experiment 2 [see (a)] and in experiment 1 [see (b)] is indicated by the ratio of $P(\text{``same''}|\langle AA\rangle)$ over $P(\text{``same''}|\langle BB\rangle)$ and that of $P(\text{``different''}|\langle AB\rangle)$ over $P(\text{``different''}|\langle BA\rangle)$.

		P("same" $\|\langle AA\rangle$)	P("same" $\|\langle BB\rangle$)	$P(\text{``same''}\|\langle AA\rangle)$ / $P(\text{``same''}\|\langle BB\rangle)$	P("different" $\|\langle AB\rangle$)	P("different" $\|\langle BA\rangle$)	P("different" $\|\langle AB\rangle$)/ P("different" $\|\langle BA\rangle$)
(a)	LG	0.725	0.705	1.028	0.495	0.525	0.943
	TG	0.770	0.420	1.833	0.665	0.700	0.950
	ZL	0.720	0.555	1.297	0.660	0.610	1.082
(b)	GS	0.792	0.656	1.207	0.588	0.584	1.007
	NB	0.748	0.492	1.520	0.656	0.620	1.058
	FG	0.705	0.655	1.076	0.595	0.560	1.062

model is that $P(\text{``different''}|\langle AB\rangle)$ equals $P(\text{``different''}|\langle BA\rangle)$ and that $P(\text{``same''}|\langle AA\rangle)$ equals $P(\text{``same''}|\langle BB\rangle)$. That is, on the two-dimensional decision space, the two halves of the decision contour for the independent-observation model are symmetric, and so are the two halves of that for the differencing model. This symmetric assumption automatically comes from the assumptions made in this research about the two observations on each trial: (1) x_1 and x_2 are equal-variance Gaussian given A or B ($\sigma_A^2=\sigma_B^2$), and (2) x_1 and x_2 are independent of each other. Decision-space analysis shows that the symmetric assumption for ROC analysis is violated in TG's case. The asymmetry of TG's judgments can be not only visually inspected across the decision space but also quantified by the proportion of TG's judgments correctly accounted for by a certain decision criterion for each type of trial [see Table V(a)]. If the symmetric assumption is met, the probability of responding "same" given $\langle AA\rangle$ should be close to that given $\langle BB\rangle$, and the probability of responding "different" given $\langle AB\rangle$ should be close to that given $\langle BA\rangle$. The asymmetry of TG's same–different judgments is indicated by the ratio of $P(\text{"same"}|\langle AA\rangle)$ over $P(\text{"same"}|\langle BB\rangle)$ and that of $P(\text{"different"}|\langle AB\rangle)$ over $P(\text{"different"}|\langle BA\rangle)$. The asymmetry of each observer's same–different judgments in experiment 1 is also shown in Table V(b). From Table V, it is apparent that TG's same–different judgments for the "same" type of trial are the least symmetric among the observers in both experiments 1 and 2.

Our application of COSS analysis in this research is also subject to the symmetric assumption and can be complicated by asymmetric judgments, because the correlation distributions were simulated based on the assumption of independent observations (note that COSS analysis itself need not have the symmetric assumption). This symmetric assumption implies, in terms of the COSS technique, that equal weights or importance are given to each observation on each trial of the same–different task, although the directions or signs are opposite.

IV. GENERAL DISCUSSION AND CONCLUSIONS

We found that the three methods (i.e., ROC analysis, decision-space analysis, and the COSS method) provided converging evidence, tending to favor the differencing model for the amplitude discrimination of acoustic sinusoids. Although ROC and COSS analyses could be complicated by asymmetric judgments and unstable criteria, they did more or less provide evidence for the adoption of the differencing strategy. Intriguingly, by visually inspecting the decision space, the partitioning patterns of the decision space for the six observers all resemble the partitioning pattern of the differencing strategy rather than that of the independent-observation strategy. Direct visual inspection of the decision space might incur the criticism that this approach is rather subjective. For experiments 1 and 2, the proportion of each observer's same–different judgments correctly described by the differencing model appears higher than that by the independent-observation model. These results mitigate, to some extent, the subjectivity entailed in making inferences from visual inspection of the decision space.

As shown in the Results sections of experiments 1 and 2, fitting the differencing or the independent-observation decision models to the same set of judgments yields discrepant estimated values of d'. The maximum-likelihood estimates of d' for the differencing model were found to be 20%–33% greater than those for the independent-observations model. This is in accord with theoretical calculations for derived d's for each model based on the same set of unbiased judgments (Wang, 2002). This also accords with the fact that the independent-observation strategy achieves a greater unbiased percentage correct than the differencing strategy does, given an underlying d' (Noreen, 1981; Wang, 2002). In this sense, the independent-observation strategy is an optimal decision strategy compared to the differencing decision strategy. The first experiment in this paper has shown that when an observer has limited information about the distributions of stimuli, the observer adopts the differencing rule. The second experiment has shown that observers still do not adopt the independent-observation rule, an optimal rule, even if equipped with sufficient information about the distributions of stimuli. The reference information intended to orient the observers to adopt the categorical rule—the simple form of the independent-observation rule—does not facilitate its adoption for this type of simple sensory stimuli. This might suggest that these simple sensory stimuli place a constraint on human observers to adopt an optimal decision rule for the same–different task.

[1]By statistical theorems (see Ross, 1998, pp. 257 and 303), the independent-observation rule can be proved to be equivalent to the categorical rule.

Irwin and Hautus's (1997) work is a precursor to the development of the general form of the categorical rule and its equivalence to the independent-observation rule.

[2]The correlation method is analogous to discriminant analysis, which distinguishes among two or more groups on two variables and is a special application of multiple regression analysis (Thorndike, 1978).

[3]The assumption of linearity is often made about the relation between similar stimuli and their discriminability, i.e., for fine discrimination (e.g., Johnson, 1980). A justification for this assumption comes from studies in fine discrimination of stimuli, such as studies in cumulative d' (e.g., Wang, Irwin, and Hautus, 1998; Macmillan and Creelman, 1991, pp. 212–216).

[4]In Berg's (1989) COSS analysis, the internal noise is added to x_i and multiplied by α_i; while in our application of COSS, like Richards and Zhu's (1994) and Lutfi's (1995) modified version of COSS analysis, the internal noise term is appended to the weighted sum of observation variables.

Ashby, F. G., and Gott, R. E. (**1988**). ''Decision rules in the perception and categorization of multidimensional stimuli,'' J. Exp. Psychol. Learn. Mem. Cogn. **14**, 33–53.

Berg, B. G. (**1989**). ''Analysis of weights in multiple observation tasks,'' J. Acoust. Soc. Am. **86**, 1743–1746.

Dai, H., Versfeld, N., and Green, D. (**1996**). ''The optimum decision rules in the same–different paradigm,'' Percept. Psychophys. **58**, 1–9.

Dorfman, D. D., and Alf, E. Jr., (**1969**). ''Maximum-likelihood estimation of parameters of signal-detection theory and determination of confidence intervals—Rating-method data,'' J. Math. Psychol. **6**, 487–496.

Francis, M. A., and Irwin, R. J. (**1995**). ''Decision strategies and visual-field asymmetries in same–different judgments of word meaning,'' Mem. Cognit. **23**, 301–312.

Green, D. M., and Swets, J. A. (**1966**). *Signal Detection Theory and Psychophysics* (Wiley, New York).

Hautus, M. J., Irwin, R. J., and Sutherland, S. (**1994**). ''Relativity of judgments about sound amplitude and the asymmetry of the same–different ROC,'' Q. J. Exp. Psychol. A **47A**, 1035–1045.

Irwin, R. J., and Francis, M. A. (**1995**). ''Perception of simple and complex visual stimuli: Decision strategies and hemispheric differences in same–different judgments,'' Perception **24**, 787–809.

Irwin, R. J., and Hautus, M. J. (**1997**). ''Likelihood-ratio decision strategy for independent observations in the same–different task: An approximation to the detection-theoretic model,'' Percept. Psychophys. **59**, 313–316.

Irwin, R. J., Stillman, J. A., Hautus, M. J., and Huddleston, L. M. (**1993**). ''Measurement of taste discrimination with the same–different task: A detection-theory analysis,'' J. Sens. Stud. **8**, 229–239.

Johnson, K. O. (**1980**). ''Sensory discrimination: Decision process,'' J. Neurophysiol. **43**, 1771–1792.

Kubovy, M., Rapoport, A., and Tversy, A. (**1971**). ''Deterministic vs probabilistic strategies in detection,'' Percept. Psychophys. **9**, 427–429.

Lee, W. (**1963**). ''Choosing among confusably distributed stimuli with specific likelihood ratios,'' Percept. Mot. Skills **16**, 445–467.

Lutfi, R. A. (**1995**). ''Correlation coefficients and correlation ratios as estimates of observer weights in multiple-observation tasks,'' J. Acoust. Soc. Am. **97**, 1333–1334.

Macmillan, N. A., and Creelman, C. D. (**1991**). *Detection Theory: A User's Guide* (Cambridge University Press, New York).

Macmillan, N. A., Kaplan, H. L., and Creelman, C. D. (**1977**). ''The psychophysics of categorical perception,'' Psychol. Rev. **84**, 452–471.

Noreen, D. L. (**1981**). ''Optimal decision rules for some common psychophysical paradigms,'' in *Proceedings of the Symposium in Applied Mathematics of the American Mathematical Society and the Society for Industrial and Applied Mathematics*, edited by S. Grossberg (American Mathematical Society, Providence, RI), Vol. 13, pp. 237–279.

Robinson, D. E., and Watson, C. S. (**1972**). ''Psychophysical methods in modern psychoacoustics,'' in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, New York), Vol. 2, pp. 101–131.

Ross, S. M. (**1998**). *A First Course in Probability*, 5th ed. (Prentice-Hall, Englewood Cliffs, N.J.).

Richards, V. M., and Zhu, S. (**1994**). ''Relative estimates of combination weights, decision criteria, and internal noise based on correlation coefficients,'' J. Acoust. Soc. Am. **95**, 423–434.

Sorkin, R. D. (**1962**). ''Extension of the theory of signal detectability to matching procedures in psychoacoustics,'' J. Acoust. Soc. Am. **34**, 1745–1751.

Thorndike, R. M. (**1978**). *Correlational Procedure for Research* (Gardner, New York).

Wang, M. (**2002**). ''Decision Processes of Same–different Judgments for the Amplitude Discrimination of Single Sinusoids and Auditory Profile Complexes,'' Unpublished doctoral dissertation, the University of Auckland, Auckland, New Zealand.

Wang, M., Irwin, R. J., and Hautus, M. J. (**1998**). ''Discriminability in length of lines in the Müller–Lyer figure,'' Percept. Psychophys. **60**, 511–517.

Ward, L. M. (**1973**). ''Use of Markov-encoded sequential information in numerical signal detection,'' Percept. Psychophys. **14**, 337–342.

Wickelgren, W. A. (**1968**). ''Unidimensional strength theory and component analysis of noise in absolute and comparative judgments,'' J. Math. Psychol. **5**, 102–122.

Effect of frequency-modulation coherence for inharmonic stimuli: Frequency-modulation phase discrimination and identification of artificial double vowels

Johannes Lyzenga and Brian C. J. Moore
Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, United Kingdom

(Received 29 July 2004; revised 3 November 2004; accepted 15 December 2004)

The ability to compare patterns of frequency modulation (FM) in separate frequency regions was explored. In experiment 1, listeners had to distinguish whether the FM applied to two nonharmonically related sinusoidal carriers was in phase or out of phase. The FM rate was the same for each carrier. The starting phase of the modulation was randomized for each stimulus in a three alternative, forced-choice (3AFC) trial. Subjects were sensitive to relative FM phase for modulation rates of 2 and 4 Hz, but not for higher rates. In experiment 2, vowel identification was compared for artificial single and double vowels. The vowels were constructed from complex tones with components spaced at 2-ERB_N (equivalent rectangular bandwidth) intervals, by increasing the levels of three components by 15 dB, to create three ''formants.'' In the double vowels, the components of the two vowels were interleaved, to give 1-ERB_N spacing. The three ''formant'' components were frequency modulated at 2, 4, or 8 Hz, with either the same or different rates for the two vowels. The identification of double vowels was not improved by a difference in FM rate across vowels, suggesting that differences in FM rate do not support perceptual segregation of inharmonic stimuli. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1856251]

PACS numbers: 43.66.Fe, 43.66.Nm, 43.71.Mk [RAL] Pages: 1314–1325

I. INTRODUCTION

When sounds are presented in two nonoverlapping spectral regions, a variety of interaction effects can be observed, depending on the circumstances. For example, in modulation detection interference (MDI) (Yost *et al.*, 1989; Yost and Sheft, 1989; Moore *et al.*, 1991) across-frequency processing has a detrimental effect on perception, whereby amplitude modulation (AM) or frequency modulation (FM) of one carrier frequency interferes with the perception of AM or FM on another carrier frequency. On the other hand, across-frequency processes can improve the detection or discrimination of signals in some circumstances, as in the phenomena of comodulation masking release (CMR) (Hall *et al.*, 1984; Buus, 1985; Schooneveldt and Moore, 1987) and profile analysis (Spiegel *et al.*, 1981; Green, 1988; Gockel, 1998). Also, across-frequency processing is clearly useful in the perception of timbre (Plomp, 1970) and in vowel recognition (Klein *et al.*, 1970; Assmann *et al.*, 1982).

This paper is concerned with the ability to compare patterns of FM in separate frequency regions. Two aspects of this ability are explored here. The first aspect is whether listeners can actually detect differences in the pattern of FM applied to different carriers under conditions where within-channel interactions are minimized and where cues related to harmonicity (see below) are not available. This was explored by requiring subjects to discriminate in-phase (coherent) FM from out-of-phase (incoherent) FM applied to two widely separated carriers. The second aspect is whether differences in the pattern of FM applied to different carriers can be used as a cue for perceptual segregation of complex sounds.

Evidence for sensitivity to FM coherence has been sought using paradigms such as MDI (Wilson *et al.*, 1990; Moore *et al.*, 1991; Carlyon, 1991, 1994), CMR (Schooneveldt and Moore, 1988; Grose and Hall, 1990), CMR interference (Grose *et al.*, 1995), discrimination between coherent and incoherent FM (McAdams, 1984; Demany and Semal, 1988; Grose and Hall, 1990; Carlyon, 1991), segregation of simultaneous vowels (McAdams, 1989; Marin and McAdams, 1991; Chalikia and Bregman, 1993; Summerfield and Culling, 1992; Culling and Summerfield, 1995), effects of FM coherence on signal detectability (Cohen and Chen, 1992; Carlyon, 1994), and the shift of the first formant phoneme boundary (Gardner and Darwin, 1986). While some studies have been interpreted as indicating a weak sensitivity to FM coherence (Chalikia and Bregman, 1993; Cohen and Chen, 1992), the majority of studies has failed to indicate such sensitivity.

Carlyon (1991, 1994, 2000) has presented evidence that apparent sensitivity to FM coherence can often be explained in terms of perceptual grouping by harmonicity. Sounds with harmonically related components tend to fuse and to be heard as a single sound, whereas when a single harmonic is mistuned it tends to ''pop out'' and be heard as a separate sound (Moore *et al.*, 1985a, 1985b, 1986). When two carriers are modulated incoherently, this may result in them moving alternately in and out of a simple harmonic relationship, and this can provide a cue for discriminating coherent from incoherent FM. When potential harmonicity cues are avoided, sensitivity to FM coherence has not usually been observed. In addition, Carlyon (1991) has suggested that within-channel cues related to beating between adjacent components may have been used in some cases.

Not all cases of sensitivity to FM coherence can be ex-

plained in terms of the use of harmonicity cues or beating cues. Wilson *et al.* (1990) measured the ability to detect FM of a sinusoidal carrier in the presence of a second nonharmonically related carrier. When the second carrier contained FM at the same rate, it interfered with the detection of the FM of the target carrier. The effect was greater when the two carriers were modulated in phase than when they were modulated out of phase, which was taken as indicating a sensitivity to FM coherence. However, Carlyon (1994) demonstrated that the effect was caused by detection of combination tones rather than by a sensitivity to FM coherence *per se*.

Furukawa and Moore (1996, 1997a) used an indirect way of assessing whether there was a mechanism sensitive to FM coherence. They investigated the detectability of FM imposed on two nonharmonically related carriers. The modulation on each carrier was equally detectable, as determined in preliminary experiments, and a pink-noise background was used to mask distortion products and to mask the outputs of auditory filters tuned between the two carriers. They found that detectability was generally better when the FM was in phase across carriers than when the FM was in antiphase.

Furukawa and Moore (1996, 1997a) discussed several possible mechanisms that might underlie the effect of FM coherence. One is that the auditory system is sensitive to the coherence or relative phase of amplitude modulation (AM) at the outputs of the auditory filters where the changes in excitation level evoked by the FM are greatest. This mechanism was referred to as the ''FM-induced-AM mechanism.'' However, Furukawa and Moore (1997b) later showed that the detectability of AM on two widely spaced inharmonically related carriers was not influenced by the coherence of the AM across carriers, suggesting that the sensitivity to FM coherence cannot be explained by FM-induced-AM. A second mechanism suggested by Furukawa and Moore (1996, 1997a, 1997b) to explain the effect of FM coherence was that judgments were partly based on fluctuations in overall pitch or residue pitch evoked by the two carriers; these fluctuations would be bigger when the modulation was coherent across carriers than when it was incoherent. The results of Carlyon (2000) support this interpretation. Whatever the mechanism, the results of Furukawa and Moore do show a perceptual effect of across-frequency FM coherence under conditions where within-channel cues and cues related to harmonicity were not available. The effect decreased with increasing FM rate from 2.5 to 10 Hz, suggesting that sensitivity to FM coherence is greatest for very low FM rates.

In the present study, we used a *direct* measure of sensitivity to FM coherence, by requiring subjects to discriminate in-phase FM from out-of-phase FM applied to two sinusoidal carriers. Most previous studies of FM coherence used FM rates of 5 Hz and above. Given the evidence from the studies of Furukawa and Moore that the effect FM coherence on the detectability of FM is greatest for very low FM rates, we used rates as low as 2 Hz. To make the stimuli relevant to the perceptual grouping of formants in vowel perception, the carrier frequencies were chosen to fall in the regions of the first and the second formants in natural speech.

In the second experiment, we assessed whether differences in the pattern of FM applied to the formants of two simultaneous synthetic vowels would enhance the identification of the vowels. If so, this might be taken as evidence that across-frequency comparison of FM patterns can affect perceptual grouping, as might be expected from the Gestalt principle of ''common fate'' (Bregman, 1990). Recognition of simultaneous vowels has been studied mainly for vowels with harmonic spectra and with either identical or different fundamental frequencies (Scheffers, 1983; Assmann and Summerfield, 1989, 1990; Summerfield and Assmann, 1991; Culling and Darwin, 1993, 1994; McKeown and Patterson, 1995; de Cheveigné *et al.*, 1997b, 1997a; de Cheveigné, 1999). The results from those studies have indicated that, for a given vowel pair, the identification of one vowel is often robust, while the identification of the other vowel is much poorer. This has led to the notion of a dominant and a nondominant vowel in simultaneously presented vowel pairs. When listening to two competing speakers, it is not clear *a priori* which one of two simultaneous vowels will emerge as the dominant one; it could belong either to the target or to the competing voice. A second outcome of work on the identification of competing vowels is an effect of their fundamental frequencies (F_0's). A difference between their F_0's usually leads to higher identification scores than when the F_0's are the same (Scheffers, 1983; Zwicker, 1984; Assmann and Summerfield, 1990; Culling and Darwin, 1994). The effect of an F_0 difference is usually greater for the nondominant vowel in a pair than for the dominant vowel.

In the present study we wanted to avoid using harmonic spectra for several reasons. First, the resolvability of harmonics changes with harmonic number (Plomp, 1964). Low harmonics are well resolved and higher ones are less well resolved or not resolved at all. This may lead to a strong emphasis on the lower harmonics of the vowels in the perceptual processing of harmonic vowel pairs (Culling and Darwin, 1994; Culling and Summerfield, 1995). Second, harmonicity of the component spectra provides a very strong perceptual grouping cue (Darwin, 1984; Darwin and Gardner, 1986; Culling and Darwin, 1993) that helps to perceptually separate the two vowels when they have different F_0's and tends to make them fuse when they have the same F_0. The powerful effects of harmonicity might make it difficult to observe the more subtle effects of cues arising from across-frequency comparison of FM patterns, and this could account for previous failures to find effects of FM coherence on the identification of double vowels with harmonic components (McAdams, 1989; Marin and McAdams, 1991; Chalikia and Bregman, 1993; Culling and Summerfield, 1995).

The effects of FM coherence on the identification of double vowels with inharmonic components was studied by Chalikia and Bregman (1993). They used stimuli in which harmonics were randomly shifted in frequency from their nominal values. The amplitudes of the shifted harmonics were adjusted so as to restore the original spectral envelope for each vowel. Each vowel had components that glided upwards in frequency by six semitones, or glided downwards in frequency by the same amount. They found better identification performance when the two vowels in a pair glided in opposite directions than when they glided in parallel, main-

taining a constant nominal difference in fundamental frequency of six semitones. This was taken as evidence for a role of FM coherence in perceptual grouping by the principle of ''common fate.''

The experiment of Chalikia and Bregman was criticized by Summerfield and Culling (1992). They pointed out that the stimuli differed in the range of nominal fundamental frequency across conditions. This meant that the phonetic distinctiveness of the vowels might have differed across conditions, depending on the precision with which the components defined the positions of the formant peaks. Summerfield and Culling (1992) carried out an experiment using inharmonic vowels, but they measured the threshold for detecting one vowel in the presence of another vowel. Both vowels were frequency modulated at the same rate (2.5 Hz), and the modulation could be either in phase (coherent) or could differ in phase across the two vowels (incoherent). No effect of FM phase was found.

The difference between the results of Chalikia and Bregman (1993) and of Summerfield and Culling (1992) might have occurred because of confounding effects in the stimuli of Chalikia and Bregman, namely differences in the distinctiveness of vowels across conditions. Alternatively, it might have occurred because the differences in the pattern of FM used by Chalikia and Bregman (parallel glides versus crossing glides) produced larger effects on perceptual segregation than the differences in FM phase (at the same rate) used by Summerfield and Culling.

In our experiment 2, we used vowel pairs whose FM *rate* differed across vowels. This contrasts with most previous studies of the role of FM coherence in vowel perception, which used FM of the same rate, but different phase applied to the different vowels (McAdams, 1989; Marin and McAdams, 1991; Chalikia and Bregman, 1993; Summerfield and Culling, 1992; Culling and Summerfield, 1995; de Cheveigné *et al.*, 1997a). We reasoned that differences in FM rate across vowels might provide a more potent cue for perceptual segregation than differences in FM phase at the same rate. This reasoning was partly based on the phenomenon of MDI, which is thought to be caused at least partly by perceptual grouping of the two carriers (Yost *et al.*, 1989; Hall and Grose, 1990, 1991; Moore and Shailer, 1992). Interference with the detection or discrimination of FM on a given carrier produced by FM of a remote carrier is not affected by the relative phase of the modulators when the FM rates are the same (Carlyon, 1991, 1994; Moore *et al.*, 1991), but it does decrease when the FM rate of the interferer differs from that of the target (Moore *et al.*, 1991). This may indicate that the perceptual segregation of the two carriers is enhanced by an FM rate difference across the carriers, but not by a phase difference when they have the same rate.

In our experiment, to avoid the influence of effects of harmonicity, we used artificial vowels, each of which had components spaced at 2-ERB_N intervals, where ERB_N stands for the average value of the equivalent rectangular bandwidth of the auditory filter as determined using young, normally hearing listeners at moderate sound levels (Glasberg and Moore, 1990; Moore, 2003). Hence, all the partials of the vowels were spaced at approximately equal distances along the length of the basilar membrane (Moore, 1986) and all components were approximately equally resolvable (Moore and Ohgushi, 1993). We measured vowel recognition for isolated vowels and for simultaneously presented vowel pairs. For the latter, the partials of the vowels were interleaved, so that there was a fixed separation of 1 ERB_N between adjacent partials, thus minimizing spectral overlap while using a maximal number of partials.

TABLE I. Ratios between the frequencies of the second and the first carriers of the stimuli used in experiment 1. All $Fc2/Fc1$ ratios are well away from integer numbers.

	$Fc1 = 208.1$ Hz	$Fc1 = 374.9$ Hz
$Fc2 = 923.9$ Hz	4.44	2.46
$Fc2 = 1364.0$ Hz	6.55	3.64

II. EXPERIMENT 1: FM COHERENCE DISCRIMINATION

A. Stimuli

Each stimulus contained two carriers that were sinusoidally frequency modulated. The carriers were modulated either in phase (coherent modulation) or 180 deg out of phase (incoherent modulation), at the same rate and with the same modulation depth. In this paper, FM depths are expressed as peak frequency deviation from the carrier frequency divided by the carrier frequency, multiplied by 100, i.e., as percent deviation from the carrier frequency. The FM depths were 0.625%, 2.5%, and 10%. Modulation rates of 2, 4, and 8 Hz were used in the main experiment, and supplementary measurements were made using a 16-Hz rate. The first carrier frequency ($Fc1$) was chosen to fall in the region of the first formant frequency of natural vowels and was either 208.1 Hz (ERB_N number=6) or 374.9 Hz (ERB_N number=9). The second carrier frequency ($Fc2$) fell in the region of the second formant and was 923.9 Hz (ERB_N number=15) or 1364.0 Hz (ERB_N number=18). These four frequencies were chosen because there are not any simple harmonic relations between $Fc1$ and $Fc2$. Table I shows the ratios between the possible values of $Fc1$ and $Fc2$; all these ratios are well away from integer numbers. The duration of each stimulus was 500 ms including 10-ms raised-cosine ramps, and the interstimulus interval was 600 ms.

The stimuli were presented at an overall level of 64 dB SPL (61 dB per carrier) in a pink-noise background with a wideband level of 46 dB SPL (100 Hz to 8 kHz) and a spectral level of 10 dB (re: 20 μPa) at 1 kHz. The excitation levels of the individual carriers were roughly 30 dB above the excitation level evoked by the pink-noise background, as calculated using the excitation pattern model of Glasberg and Moore (1990).

The stimuli were generated offline using a PC, and were stored on the PC disk. During the measurements, they were retrieved from the disk and converted to analog form using a Turtle Beach Montego II soundcard (16-bit resolution, 44.1-kHz sample rate). The pink noise was generated using an Ivie Electronics IE20B noise generator. The signals and pink noise were mixed using a Mackie 1202 VLZ Pro 12-channel mixing desk. The mixing desk drove the right headphone of

Sennheiser HD 580 headphones through an impedance-matched Hatfield 2125 attenuator. Subjects were tested in a double-walled sound-attenuating booth.

B. Method

Psychometric functions for the discrimination of FM coherence were measured using a three-interval, three-alternative forced-choice task (3I,3AFC) with feedback. In each trial, two stimuli had coherent FM and one stimulus, selected randomly, had incoherent FM. The task of the subject was to indicate which of the three stimuli was modulated incoherently. To prevent subjects from using the starting phase of the modulation applied to a single carrier as a cue, the starting modulator phase for the lower carrier frequency was selected randomly for each stimulus from the values 0 and 180 deg. Feedback was supplied through a window in the sound-attenuating booth, via a computer screen.

Stimuli were presented in blocks of 30 (20 for subject 1) trials with fixed carrier frequencies and a fixed modulation depth. For a given pair of carrier frequencies, the first block used the largest FM depth (10%) and subsequent blocks used the 2.5% and 0.625% depths. A new pair of carrier frequencies was then selected for testing. These blocks were interspersed with runs for the vowel identification task of experiment 2. Two blocks were obtained for each subject for each condition, with reversed order of modulation depths in the second block, resulting in 60 responses per condition for each subject (40 for subject 1).

C. Subjects

Six subjects were used, five females and one male, with ages ranging from 29 to 43 with a mean of 35. All subjects had normal hearing in their right ear and were tested using that ear. All subjects were highly experienced in psychoacoustic experiments, including experiments involving FM detection and discrimination. Subjects were given two blocks of practice trials for each pair of carrier frequencies, using the largest FM depth, before data collection started (240 practice trials in total).

D. Results

Overall performance varied somewhat across subjects, but the pattern of results was similar for all subjects. To prevent possible ceiling effects for high percent-correct scores from distorting the averages and their variances, the percent-correct scores were converted to d' values using standard tables (Hacker and Ratcliff, 1979). The average results across subjects are shown in Fig. 1 as d' versus modulation depth. The error bars indicate the across-subject 95% confidence intervals. The relatively large error bars partly reflect the fact that overall performance differed across subjects, one subject performing markedly better than the others. The four panels show the results for the different combinations of carrier frequency. ''Low'' means that both the first and second carriers were at the lower of the two possible frequencies (see Table I for exact values). ''High'' means that both carriers were at the higher of the two possible frequencies. ''Narrow'' means that the first carrier was at the higher of the two possible frequency values and the second carrier was at the lower of the two values. For the case ''wide,'' the reverse was true. The squares, circles, and triangles show data for the 2-, 4-, and 8-Hz modulation rates, respectively.

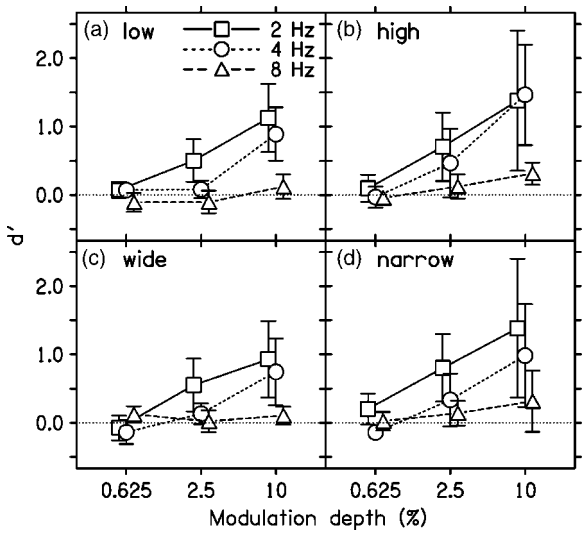


FIG. 1. Scores for the FM phase discrimination task, expressed as d'. Panels (a) to (d) show the results for the conditions ''low,'' ''high,'' ''wide,'' and ''narrow'' (see the text), respectively. The squares, circles, and triangles show the results for the 2-, 4-, and 8-Hz modulation rates, respectively. The error bars show 95% confidence intervals.

Scores were at or close to chance for the smallest of the three modulation depths and for the 8-Hz modulation rate. However, for the 2- and the 4-Hz modulation rates, the d' values approached or exceeded 1 for the two larger modulation depths. In other words, discrimination of relative modulator phase was possible for the two lower modulation rates. To assess the significance of these effects, a three-way within-subjects analysis of variance (ANOVA) was conducted on the mean d' values for each subject, with factors modulation rate, modulation depth, and combination of carrier frequencies. This showed significant effects of modulation rate ($F_{(1,7)}=7.46$, $p<0.03$) and depth ($F_{(1,6)}=12.2$, $p<0.02$) and a significant interaction between these two factors ($F_{(4,20)}=8.50$, $p<0.0005$). The main effect of combination of carrier frequencies was not significant ($F_{(1,7)}=4.26$, $p=0.07$), and no interactions involving this factor reached significance. As the effect of combination of carrier frequencies was not significant, the data were pooled across the four pairs of carriers. A two-way ANOVA based on the pooled data, with factors modulation rate and modulation depth showed significant effects of both main factors ($F_{(2,36)}=23.9$, $p<0.0005$; $F_{(1,29)}=39.8$, $p<0.0005$) and their interaction ($F_{(4,92)}=20.0$, $p<0.0005$). Bonferroni-corrected *post-hoc* pairwise comparisons showed significant differences between the scores for the 2- and the 8-Hz rates ($p<0.0005$), between the 2- and the 4-Hz rates ($p<0.01$), and between the 4- and the 8-Hz rates ($p<0.0005$). Similar pairwise comparisons for the three modulation depths showed the d' values for all three depths to be significantly different from each other ($p<0.0005$ in all cases).

Figure 2 shows the individual data plotted as a function of modulation depth for the three modulation rates. Because

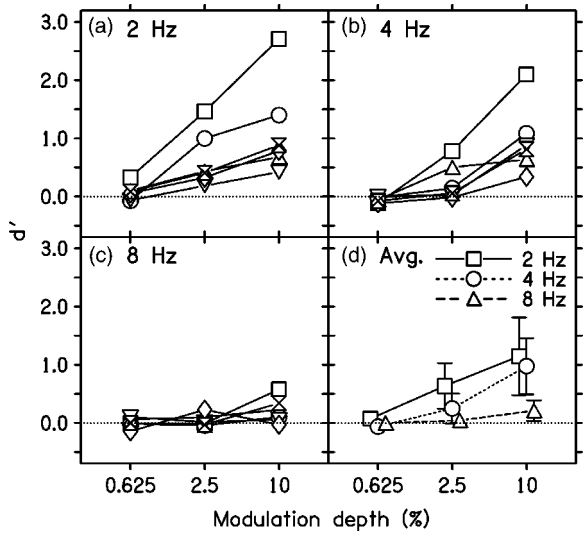


FIG. 2. Individual results for the three modulation rates plotted as a function of modulation depth. Panels (a) to (c) show the individual scores for the 2-, the 4-, and 8-Hz rates, respectively. Panel (d) shows the averages and 95% confidence intervals of the scores for each modulation rate.

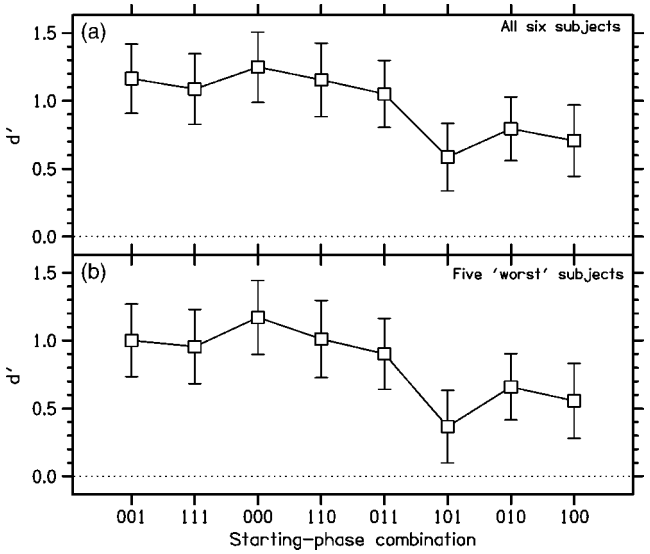


FIG. 3. Average scores for the 2-Hz modulation rate and 10% modulation depth for the eight possible combinations of modulator starting phase. The scores shown in the top panel are averages across all six subjects and those in the bottom panel are averages across the five subjects with the ''worst'' scores. The error bars show the 95% confidence intervals. The different starting phases are coded sequentially with ''0'' or ''1'' for the two reference stimuli and the target, where 0 means that lower component started in the ''up'' direction and 1 means that it started in the ''down'' direction.

of the lack of an effect of combination of carrier frequencies, the d' values have been averaged across the various pairs of carrier frequencies. Figure 2 shows that one subject (subject 1) performed much better than the others for the 2-Hz modulation rate and slightly better than the others for the 4-Hz rate. Nevertheless, for the 2- and the 4-Hz rates, the d' values increase with increasing modulation depth for all subjects.

We consider next the results for the different starting modulator phases used in the experiment. There were two possible starting phases for each of the three stimuli within a trial, leading to a total of eight combinations. Two of these combinations had identical starting phases for the lower ($Fc1$) carrier of the target stimulus (incoherently modulated) and the two ''standard'' (coherently modulated) stimuli, and two had identical starting phases of the higher carrier ($Fc2$) for all three stimuli. For the remaining four combinations, the two standard stimuli had opposite starting phases, creating a larger uncertainty in the stimuli, and possibly leading to lower scores. Figure 3 shows the average results for the modulation rate and depth that led to the highest scores (2-Hz rate, 10% depth), separately for each of the eight starting phase combinations. Panel (a) shows results for all six subjects and panel (b) shows results for the five ''worst'' subjects. The error bars indicate 95% confidence intervals. Scores were indeed higher for the combinations with identical starting phases for the two standard stimuli (the four conditions on the left), but they still reached levels well above chance for the four combinations with different starting phases for the two standard stimuli (the four conditions on the right).

A potential problem in interpreting these results occurs for the largest modulation depth used. With this depth (10%), the carrier frequencies for the incoherently modulated stimuli would have fallen momentarily close to harmonic relationships. Consider, for example, the case when $Fc1$ was 374.9 Hz and $Fc2$ was 923.9 Hz. The maximum instantaneous frequency of $Fc1$ (412.4 Hz) coincided with the minimum instantaneous frequency of $Fc2$ (831.5 Hz), and these two frequencies had almost an octave relationship. Thus, subjects might have identified the interval with incoherent modulation by detecting the momentary harmonicity in the target interval. This problem does not apply for the 2.5% modulation depth, as the instantaneous frequencies of the two carriers always stayed well away from a simple harmonic relationship. For this modulation depth, the mean value of d' for the 2-Hz modulation rate was 0.64, with a standard error of 0.106. This is significantly greater than 0 ($p<0.001$). For the 4-Hz rate, the mean value of d' was 0.254, with a standard error of 0.085. Again, this is significantly greater than 0 ($p<0.05$).

We conclude that the results show a sensitivity to across-frequency FM coherence that cannot be explained in terms of the use of harmonicity cues. Sensitivity to FM coherence was moderately high for the 2-Hz modulation rate, poor for the 4-Hz rate, and absent for the 8 Hz rate.

III. EXPERIMENT 2: VOWEL IDENTIFICATION

A. Stimuli

The stimuli were complex tones with components uniformly spaced on an ERB_N scale, for the reasons described in the introduction. Each ''vowel'' had components spaced at 2-ERB_N intervals. One set of vowels (set A) had components with ERB_N numbers from 2 to 28, while the other set (B) had components with ERB_N numbers from 3 to 29. Double vowels always contained one vowel from set A and one from set B, so that the combined stimuli had components spaced at 1-ERB_N intervals. In this way low-frequency beating cues, which are of use in the identification of simultaneously presented harmonic vowels (Culling and Darwin, 1994; Culling

TABLE II. Component numbers and corresponding formant frequencies for the three-formant artificial vowels of experiment 2. Six vowel types were constructed for each of the two interleaved 2-ERB_N spaced spectra, making a total of 12 vowels. The component numbers refer to these 2-ERB_N spaced spectra. The corresponding component numbers for the combined single-ERB_N spaced spectra are two times the component number for the "set A" vowels and two times the component number plus 1 for the "set B" vowels.

Vowel:	Formant component numbers	Frequency 1st formant (Hz)	Frequency 2nd formant (Hz)	Frequency 3rd formant (Hz)
/a/ set A	8 10 12	1201.2	1972.0	3158.1
/a/ set B	8 10 12	1055.1	1747.1	2812.1
/ɛ/ set A	6 10 12	700.4	1972.0	3158.1
/ɛ/ set B	6 10 12	605.5	1747.1	2812.1
/i/ set A	4 11 13	375.0	2501.4	3972.9
/i/ set B	4 11 13	313.3	2222.5	3543.6
/ɔ/ set A	6 8 10	700.4	1201.2	1972.0
/ɔ/ set B	6 8 10	605.5	1055.1	1747.1
/o/ set A	5 7 11	520.2	923.9	2501.4
/o/ set B	5 7 11	443.7	806.2	2222.5
/u/ set A	4 8 10	375.0	1201.2	1972.0
/u/ set B	4 9 11	313.3	1364.0	2222.5

and Summerfield, 1995), could be minimized. Also, there were no harmonicity cues or spectral regularity cues (Roberts and Bregman, 1991; Roberts and Brunstrom, 2001) that could be used to segregate the two vowels.

The stimuli contained "background" components, all of equal level. Six types of artificial vowels were created by increasing the levels of three "formant" components by 15 dB. Harmonic stimuli similar to these have been used in previous studies of vowel-like sounds, and have been found to be readily identified as vowels, even by untrained listeners (Leek *et al.*, 1987; Alcántara and Moore, 1995). For each vowel type, a version was made for both set A and set B; these two versions were made as perceptually similar as possible, given the limitation imposed by the different component frequencies for the two sets. The adjustments involved picking the component to be incremented from the components closest in frequency to the "target" value for each formant. The naturalness of each vowel, and the similarity of each vowel in set A and B were judged subjectively by the authors.

The six vowels were /i/, /ɛ/, /a/, /ɔ/, /o/, and /u/. The component numbers of the formant components and their corresponding frequencies are given in Table II. Examples of spectra of the stimuli are shown in Fig. 4. All components of the complexes had a starting phase of 0 deg. To make the vowels sound somewhat more natural and to allow us to investigate the effect of modulation rate, the three formant components of each vowel were sinusoidally frequency modulated at a rate of 2, 4, or 8 Hz, with a modulation depth of 5%; for a given vowel, all components were modulated at the same rate. The modulation of each formant component had a starting phase of either 0 or 180 deg. The three formant components either all had the same starting phase or the middle formant had the opposite phase from the other two (later analyses indicated that there was no influence of this phase difference on the results). During the measurements, stimuli were chosen at random from the four resulting phase configurations.

For the isolated vowels, identification scores were measured for all three FM rates. For the double vowels, one vowel was modulated at an 8-Hz rate while the FM rate of the other vowel was either 2, 4, or 8 Hz. The goal was to determine whether a modulation-rate difference between the two simultaneously presented vowels would aid their perceptual segregation and hence lead to improved identification.

The overall level of each vowel was 67 dB SPL; the level of each formant component was 61 dB SPL. To prevent the detection of combination tones, and to limit peripheral interactions, all stimuli were presented in a pink-noise background with a spectral level of 10 dB (re: 20 μPa) at 1 kHz. These levels resulted in excitation levels (Glasberg and Moore, 1990) for the vowel components that were roughly 15 dB above the excitation level evoked by the pink-noise background for the background components and 30 dB above that level for the formant components. Stimulus generation and presentation methods were identical to those for experiment 1.

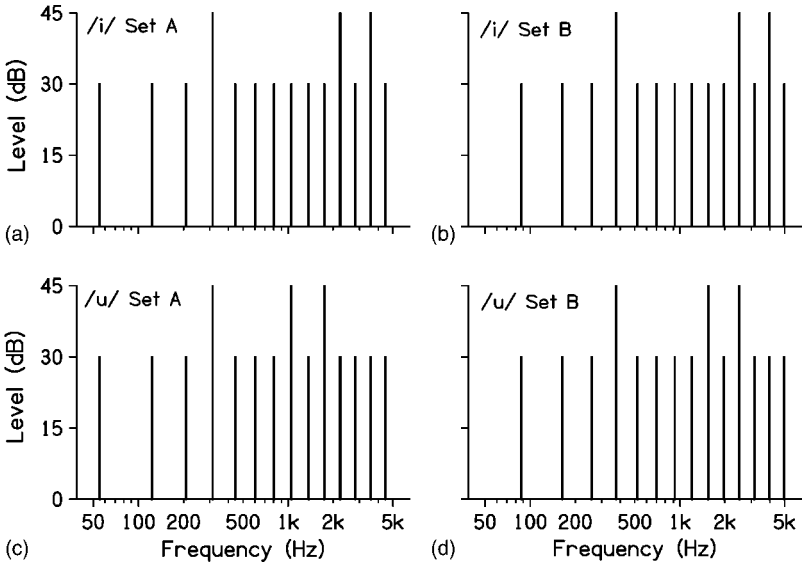


FIG. 4. Schematic spectra of some of the vowel stimuli. The top row shows the spectrum of /i/ when in set A and set B; the nominal formant component numbers were the same for the two sets. The bottom row shows the spectrum of /u/ when in set A and set B; the nominal formant component numbers were different for the two sets.

B. Method

For the vowel identification task with the isolated vowels, subjects were asked to use a mouse to select one of six ''boxes'' on a computer screen, indicating the six vowel types. The vowel type was indicated in an ''h–vowel–d'' context. Feedback was given immediately after the response, via the computer screen, which indicated the correct answer. For the vowel pairs, subjects had to select two different boxes indicating the two perceived vowels, after which feedback was given for both vowels. Before each run, subjects could play a sample of each vowel (from set A and set B) or vowel pair as many times as they liked, until they felt that they were familiar with the characteristics of the stimuli. Each run contained 40 presentations of either an isolated vowel or a vowel pair. For each condition, subjects performed 10 or 12 runs. This resulted in 70 to 80 responses per isolated vowel and 140 to 160 responses per vowel for the vowel pairs. Runs for the vowel identification task were interspersed with runs for the FM-phase discrimination task of experiment 1.

C. Subjects

The six subjects of experiment 1 were all tested using the 8-Hz modulation rate, both for the isolated vowels and for the double vowels. Five subjects out of those six were also tested using the 4-Hz modulation rate for the isolated vowels, and using the 8-Hz rate combined with the 4-Hz rate for the double vowels. For convenience, these subjects are referred to as the ''4-Hz'' group. A different group of five subjects out of the original six was tested using the 2-Hz modulation rate for the isolated vowels, and using the 8-Hz rate combined with the 2-Hz rate for the double vowels. These subjects are referred to as the ''2-Hz'' group.

D. Results

All subjects reported that the stimuli sounded like vowels, especially when presented in isolation. The pattern of results was broadly similar across subjects, so we focus on the mean results. Figure 5(a) shows mean scores for the isolated vowels for the five subjects tested using 2- and 8-Hz formant modulation rates (the 2-Hz group). Figure 5(b) shows the corresponding scores for the 4-Hz group. For all six vowel types, the scores for the 4-Hz rate are very close to those for the 8-Hz rate, but the scores for the 2-Hz rate are consistently somewhat smaller. For the 2 Hz group the overall average scores were 70% and 75% for the 2- and the 8-Hz modulation rates, respectively. For the 4-Hz group the average scores were 77% and 76% for the 4- and the 8-Hz modulation rates, respectively. Possibly, the slightly reduced scores for the 2-Hz rate reflect a reduced ability to estimate the mean formant frequencies when the FM rate is very low.

As the confusion matrices for all FM rates were found to be very similar (the rms differences were smaller than 3%), they have been combined into a single confusion matrix, shown in Table III. That the confusion matrices were very similar for the three FM rates indicates that the modulation rate had very little effect on the perceptual distances between the isolated vowels. The most frequent confusion was that /ɔ/ was identified as /o/ and vice versa.

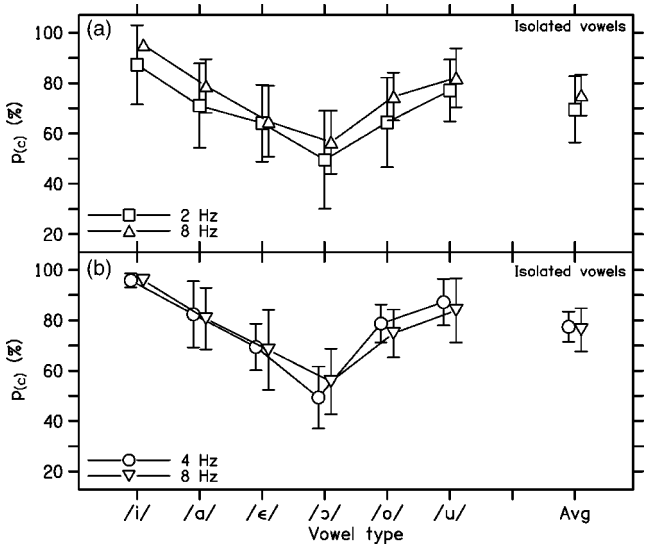


FIG. 5. Percent-correct identification of the artificial vowels presented in isolation. The top panel contains the data for the 2-Hz group, showing the results for the 2-Hz (squares) and 8-Hz (triangles) modulation rates. The bottom panel shows the data for the 4-Hz group for the 4-Hz (circles) and 8-Hz (inverted triangles) FM rates. The error bars show 95% confidence intervals.

Figure 6(a) shows scores for the double vowels when both were modulated at an 8-Hz rate (triangles) and when one was modulated at 2 Hz and one at 8 Hz (squares). The average scores across the five subjects (2-Hz group) and six vowel types were 54% for the former and 56% for the latter. Figure 6(b) shows scores for the double vowels when both were modulated at an 8-Hz rate (inverted triangles) and when one was modulated at 4 Hz and one at 8 Hz (circles). The average score across the five subjects (4-Hz group) and six vowel types was 55% for the former and 54% for the latter. There is little or no effect of a difference in FM rate on the identification of double vowels, at least in these mean scores. However, it should be remembered that scores for isolated vowels modulated at a 2-Hz rate were lower than scores for isolated vowels modulated at an 8-Hz rate. Therefore, the comparison shown in Fig. 6(a) may be somewhat unfair.

TABLE III. Confusion matrix for the vowels presented in isolation. The data for the 2-, 4-, and 8-Hz FM rates have been combined, giving a total of 7920 vowel presentations. The vowel stimuli are shown in the left column and the responses are shown in the top row. All scores are expressed as a proportion of the total presentations of a given vowel. The order of the vowels has been chosen to minimize the number of responses more than one position away from the diagonal of the matrix.

	Response					
Stimulus	/i/	/ɛ/	/a/	/ɔ/	/o/	/u/
/i/	0.896	0.052	0.018	0.005	0.003	0.025
/ɛ/	0.054	0.676	0.076	0.106	0.035	0.054
/a/	0.018	0.075	0.806	0.087	0.008	0.006
/ɔ/	0.005	0.114	0.095	0.541	0.220	0.025
/o/	0.003	0.033	0.007	0.193	0.710	0.054
/u/	0.026	0.054	0.006	0.023	0.057	0.834

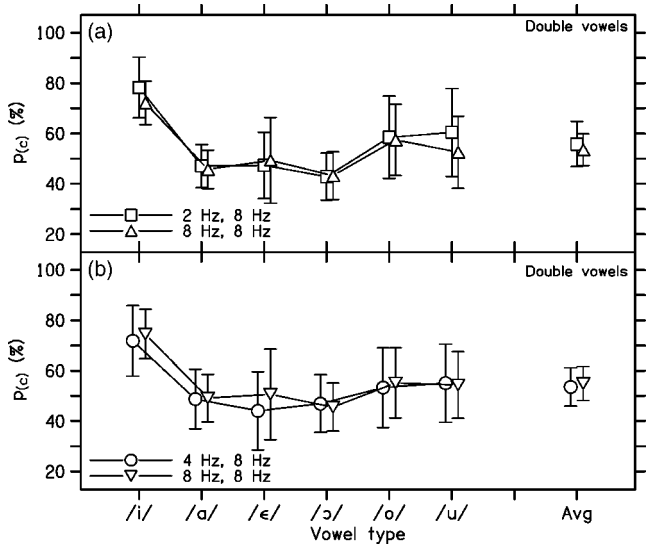


FIG. 6. Percent-correct identification of the simultaneously presented vowel pairs. The format is similar to that for Fig. 5. The symbols indicate the combination of modulation rates used.

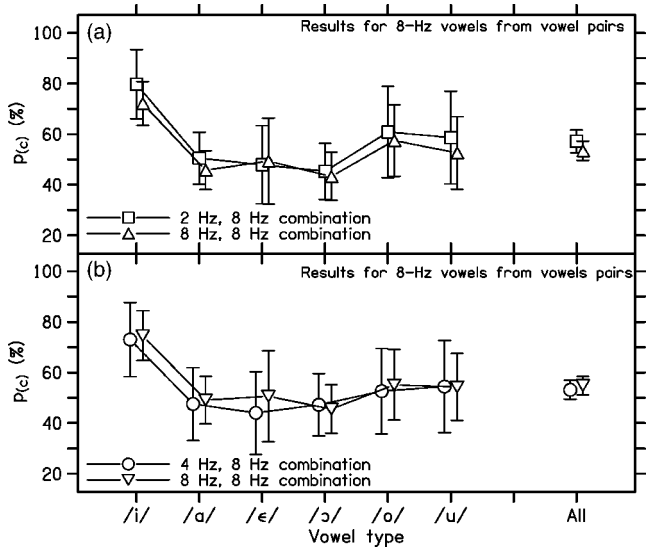


FIG. 7. Part (a) shows mean scores for each vowel type for a vowel modulated at 8 Hz, when that vowel was combined with another vowel modulated at 2 Hz (squares) and when it was combined with another vowel modulated at 8 Hz (triangles) (data from the 2-Hz group). Part (b) shows corresponding scores for a vowel modulated at 8 Hz, when that vowel was combined with another vowel modulated at 4 Hz (circles) and when it was combined with another vowel modulated at 8 Hz (inverted triangles) (data from the 4-Hz group). The error bars show 95% confidence intervals.

Another way of assessing whether there is a benefit in having a difference in modulation rate across vowels is to compare performance in identifying a vowel modulated at an 8-Hz rate when that vowel is presented with a second vowel modulated at either 8 Hz or another rate (2 or 4 Hz). For that comparison, the only difference between the two cases is whether the second vowel is modulated at the same or at a different rate from the first vowel. To make this comparison, the overall scores for the case where one vowel was modulated at 2 Hz (or 4 Hz) and one at 8 Hz were partitioned into the score obtained for the vowel that was modulated at 2 Hz (or 4 Hz) and the score for the vowel that was modulated at 8 Hz. The latter could then be compared with the overall score when both vowels were modulated at 8 Hz. Figure 7(a) shows mean scores for each vowel type for a vowel modulated at 8 Hz, when that vowel was combined with another vowel modulated at 2 Hz (squares) and when it was combined with another vowel modulated at 8 Hz (triangles)(data from the 2-Hz group). The scores tend to be slightly higher for the former than for the latter, but the differences are not large. The overall mean was 57% when the two vowels were modulated at different rates and 53% when they were modulated at the same rate. A within-subjects ANOVA based on the scores shown in Fig. 7(a), with factors modulation-rate combination and vowel type, did not give a significant effect of modulation-rate combination ($F_{(1,4)}=0.76$, $p=0.43$), although the effect of vowel type was significant ($F_{(5,20)}=4.46$, $p<0.01$).

Figure 7(b) shows corresponding mean scores for the vowel modulated at 8 Hz when that vowel was combined with another vowel modulated at 4 Hz (circles) and when it was combined with another vowel modulated at 8 Hz (inverted triangles)(data from the 4-Hz group). There is no trend for scores to be higher for the former than for the latter. The overall mean score was 53% when the two vowels were modulated at different rates and 55% when they were modulated at the same rate. A within-subjects ANOVA based on the scores shown in Fig. 7(b) showed no significant effect of modulation-rate combination ($F_{(1,4)}=3.16$, $p=0.15$) or of vowel type ($F_{(5,20)}=2.40$, $p=0.07$).

We conclude that the data do not provide any clear support for the hypothesis that differences in FM rate across vowels can promote perceptual segregation and thereby enhance the identification of simultaneously presented vowels.

As the confusion matrices for all three FM rate combinations were found to be very similar (the rms differences were smaller than 5%), they have been combined into a single confusion matrix shown in Table IV. The confusion matrix shows a different pattern from that in Table III, with an increase in the proportion of errors, and a change in the distribution of those errors. These effects will be examined in more detail in the Discussion section.

TABLE IV. As Table III, but for vowels presented in pairs. The data for all three formant-modulation rate combinations have been combined, making a total number of 7800 presented vowel pairs (i.e., 15 600 vowels). When neither of the vowels in a pair was identified correctly, it was not possible to determine which reported vowel was confused with which presented one. The entries in these cases have been assigned randomly to the possible categories of confusions.

	Response					
Stimulus	/i/	/ɛ/	/a/	/ɔ/	/o/	/u/
/i/	0.710	0.089	0.063	0.041	0.032	0.065
/ɛ/	0.089	0.452	0.129	0.145	0.088	0.097
/a/	0.068	0.139	0.510	0.129	0.071	0.083
/ɔ/	0.042	0.151	0.125	0.450	0.133	0.098
/o/	0.034	0.094	0.071	0.137	0.586	0.078
/u/	0.068	0.101	0.081	0.098	0.076	0.575

IV. DISCUSSION

A. Experiment 1: FM phase coherence discrimination

The average d' values for the across-frequency FM phase discrimination task, shown in Fig. 1, indicate that this task can be performed, but only for very low modulation rates. To provide further confirmation that the task could not be performed for high modulation rates, the two subjects who had given the highest scores for the 2-Hz rate participated in a pilot study of the discrimination of FM coherence using a modulation rate of 16 Hz. Their scores for that relatively high modulation rate did not deviate significantly from chance for any of the FM depths tested.

It is worth considering whether some cue other than across-frequency FM coherence might have been used to perform the task. As discussed earlier, a harmonicity cue might have been used for the 10% modulation depth, but this cue was not available for the 2.5% depth, as the two carrier frequencies did not approach a simple harmonic relationship at any point during the modulation cycle for the stimuli with incoherent modulation. Performance remained above chance for the 2.5% FM depth for the 2- and 4-Hz modulation rates. A second possible cue might arise from interaction of the carriers at the outputs of auditory filters tuned between them (Carlyon, 1994). However, it seems unlikely that this cue would have been usable. Consider, for example, the ''wide'' condition, in which the two carriers had frequencies of 208 and 1364 Hz. Calculation of the excitation pattern for these two tones presented simultaneously (Glasberg and Moore, 1990) showed that the excitation level at the center frequency where they interacted most strongly was -16.5 dB, while the excitation level evoked by the pink-noise background at this center frequency was 31 dB. Peripheral interaction seems extremely unlikely in this case. Even for the ''narrow'' condition, where the carrier frequencies were 375 and 924 Hz, the excitation level at the center frequency where the interaction was greatest (642 Hz) was only 25.5 dB, while the pink noise again evoked an excitation level of 31 dB.

We can conclude that neither harmonicity nor within-channel cues can account for the observed sensitivity to FM coherence. Overall, the results indicate that the auditory system does have at least some sensitivity to across-frequency FM coherence for very low FM rates. Previous reports of a complete absence of such sensitivity (Carlyon, 1991, 1994) can be attributed to the use of FM rates of 5 Hz and above, and perhaps to individual differences in sensitivity. Our results are consistent with the indirect measures of Furukawa and Moore (1996, 1997a) in showing greater effects of FM coherence for very low modulation rates.

It is noteworthy that the average d' values did not rise far above unity, even for the largest FM depth used (10%), for which the FM was highly audible. The task never became really easy, even for the lowest FM rate and for highly audible modulation depths. Subjects reported that they performed the task by ''following'' the modulation, i.e., by listening for the pattern of upward and downward changes in the frequencies of the two carriers. They reported that they did not perform the task by judging the degree of perceptual fusion of the two carriers. This is consistent with earlier reports (McAdams, 1989; Marin and McAdams, 1991; Chalikia and Bregman, 1993; Summerfield and Culling, 1992; Culling and Summerfield, 1995; Gardner and Darwin, 1986), and with the results of our experiment 2, indicating that FM coherence provides at most a very weak cue for perceptual grouping.

It is possible that sensitivity to FM coherence is found only for very low FM rates because the ability to ''follow'' the pattern of frequency changes depends on the use of information extracted from phase locking. It has been proposed that the mechanism that decodes phase-locking information is ''sluggish'' and cannot track rapid FM (Sek and Moore, 1995; Moore and Sek, 1996). If this explanation is correct, the ability to discriminate FM coherence should disappear if one or both of the carriers lies above 5 kHz, a frequency above which phase locking is almost absent (Johnson, 1980). This prediction remains to be tested.

B. Experiment 2: Vowel identification

The results of experiment 2 failed to provide evidence that a difference in FM rate across two simultaneously presented synthetic vowels could enhance the perceptual separation of those vowels and hence lead to improved identification. This happened even though each vowel had components that were equally spaced on an $\mathrm{ERB_N}$ scale, which minimized cues for segregation based on harmonicity. It might be argued that our failure to find an effect of a difference in FM rate occurred because our stimuli were not sufficiently vowel-like. The lack of harmonic structure might have led to the perceptual segregation of the first and second ''formants'' within a given ''vowel''. Also, the values of the ''formant'' frequencies were restricted to values falling at equal spacing on the $\mathrm{ERB_N}$ scale, which meant that the choice was not ideal in all cases.

To assess the extent to which our vowels were perceived in a similar way to natural vowels, the confusion matrices in Tables III and IV were used to calculate two-dimensional perceptual maps for the single and double vowels (Klein *et al.*, 1970; Kruskal, 1964). Figure 8 shows the results of this for the isolated vowels, after rotation and scaling of the results to align them with the $F1-F2$ plane, which is often used to represent vowel space. The bold symbols connected by a dotted line show the results of the multidimensional scaling, while the fainter symbols indicate the positions of the two versions of each vowel (one from set A and one from set B) in the $F1-F2$ plane. There is a good correspondence between the perceptual space and the physical representation in the $F1-F2$ plane. Furthermore, the pattern of the perceptual results is similar to that found for more natural vowels (Klein *et al.*, 1970).

Figure 9 shows the corresponding perceptual map for the simultaneous vowel pairs. The map is similar to that for the isolated vowels with the exception of the vowel /a/. For the isolated vowels, there was a large perceptual distance between /a/ and the two vowels /i/ and the /u/, indicating large dissimilarity. For the simultaneous vowels, these perceptual distances were reduced. In the data, this effect was revealed by more confusions between /a/ and /i/ and between

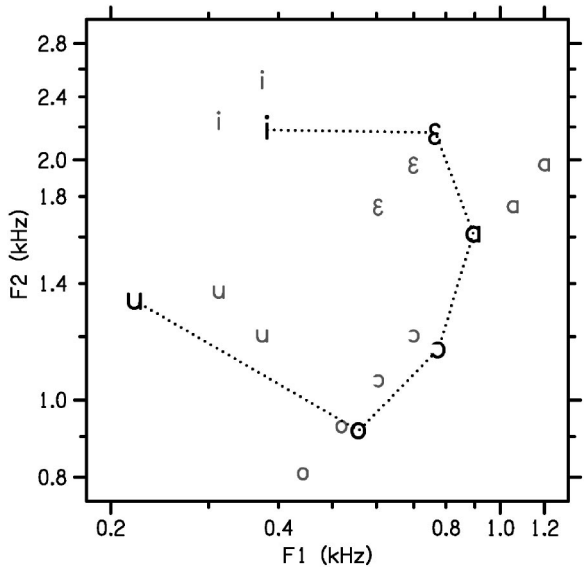


FIG. 8. Large bold symbols connected by dotted lines show the results of multidimensional scaling for the artificial vowels presented in isolation, based on combined data for the 2-, 4-, and 8-Hz rates. After rotation and scaling of the perceptual space, the dimensions 1 and 2 corresponded very well with the frequencies of the first and second formants, $F1$ and $F2$. Hence, the results are displayed in the $F1-F2$ space. The small isolated symbols indicate the positions of the stimulus vowels in $F1-F2$ space.

/a/ and /u/. Overall, though, the perceptual space for the double vowels was broadly similar to that for the isolated vowels and for more natural vowel stimuli. Hence, it seems unlikely that our failure to find an effect of differences in FM rate on the identification of double vowels was caused by our stimuli being insufficiently vowel-like.

As noted in the Introduction, when listening to two simultaneous vowels, one of the two vowels often stands out as dominant and more easily identifiable, while the other, nondominant, vowel is harder to identify. Scores for the nondominant vowel tend to be affected more than scores for the dominant vowel by manipulations such as differences in fundamental frequency between vowels. It seemed possible that

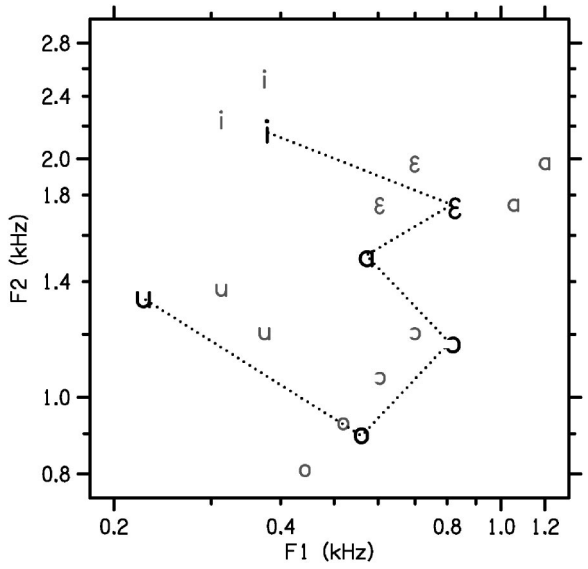


FIG. 9. As Fig. 8, but for the simultaneous vowel pairs.

TABLE V. The probability of zero (p_0), one (p_1), and two (p_2) correct responses for simultaneously presented vowels for a vowel with a 2-, 4-, or 8-Hz FM rate combined with a vowel with an 8-Hz FM rate. Columns 4 to 6 show the derived values of the overall average probability correct (p_{avg}), and the average probability of identifying the dominant vowel (p_d) and the nondominant vowel (p_{nd}). For comparison, the last column shows the average probability of correct identification for vowels presented in isolation (p_{iso}).

	p_0	p_1	p_2	p_{avg}	p_d	p_{nd}	p_{iso}
2 Hz, 8 Hz	0.15	0.58	0.27	0.56	0.63	0.43	0.72
4 Hz, 8 Hz	0.16	0.60	0.24	0.54	0.60	0.40	0.77
8 Hz, 8 Hz	0.14	0.63	0.23	0.55	0.60	0.39	0.76

effects of differences in FM rate across vowels might show up for the nondominant vowel, but not for the dominant vowel. To explore this possibility, we determined the proportion of trials with no (p_0), a single (p_1), and two vowels correct (p_2). From these scores it is possible to calculate the probability of correct identification of the dominant (p_d) and nondominant (p_{nd}) vowels as well as the average probability correct (p_{avg}). From the relations

$$p_2 = p_d p_{nd}, \tag{1}$$

$$p_0 = (1-p_d)(1-p_{nd}), \tag{2}$$

$$p_1 = 1 - p_2 - p_0, \tag{3}$$

p_d can be derived as

$$p_d = p_1 + 2p_2 - 0.5\sqrt{(p_1^2 + 4p_1p_2)}, \tag{4}$$

after which p_{nd} can be computed using Eq. (1). The average probability of correct identification is related to p_1 and p_2 by

$$p_{avg} = 0.5p_1 + p_2. \tag{5}$$

The results of this exercise are summarized in Table V. The results are consistent with the notion of a dominant and a nondominant vowel. The values of p_d are only about 0.09 to 0.17 lower than the mean proportions correct for the vowels presented in isolation (right-most column of Table V), while the values of p_{nd} are 0.29 to 0.37 lower. The values of p_d and p_{nd} are slightly greater for the 8- and 2-Hz FM rate combination than for the 4- and 8-Hz or 8- and 8-Hz combinations. However, the differences are small, and are not greater for p_{nd} than for p_d. Thus, the data provide no evidence to support the idea that effects of FM rate differences on the identification of double vowels are greater for the nondominant than for the dominant vowel.

V. CONCLUSIONS

For complex sounds with two nonharmonically spaced carrier frequencies, one in the first-formant region of vowels and one in the second-formant region, subjects were able to score above chance in a task requiring discrimination of coherent (in phase) and incoherent (out of phase) FM applied to the two carriers, but only for modulation rates of 2 and 4 Hz, and not for a rate of 8 Hz. We conclude that the auditory system is sensitive to across-frequency FM coherence, but only for very low modulation rates.

For vowel-like sounds with components uniformly spaced on an ERB_N scale (to eliminate effects of harmonicity), subjects were not able to use differences in FM rate across two simultaneously presented vowels to improve vowel identification. These results suggest that differences in the FM rate applied to the two vowels did not promote perceptual segregation of the vowels. Subjective reports are consistent with this interpretation.

ACKNOWLEDGMENTS

This work was supported by the MRC (UK). We thank two anonymous reviewers for helpful comments on an earlier version of this paper.

Alcántara, J. I., and Moore, B. C. J. (**1995**). ''The identification of vowel-like harmonic complexes: Effects of component phase, level, and fundamental frequency,'' J. Acoust. Soc. Am. **97**, 3813–3824.

Assmann, P. F., and Summerfield, Q. (**1989**). ''Modeling the perception of concurrent vowels: Vowels with the same fundamental frequency,'' J. Acoust. Soc. Am. **85**, 327–338.

Assmann, P. F., and Summerfield, A. Q. (**1990**). ''Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies,'' J. Acoust. Soc. Am. **88**, 680–697.

Assmann, P. F., Nearey, T. M., and Hogan, J. T. (**1982**). ''Vowel identification: Orthographic, perceptual, and acoustic aspects,'' J. Acoust. Soc. Am. **71**, 975–989.

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (Bradford Books, MIT Press, Cambridge, MA).

Buus, S. (**1985**). ''Release from masking caused by envelope fluctuations,'' J. Acoust. Soc. Am. **78**, 1958–1965.

Carlyon, R. P. (**1991**). ''Discriminating between coherent and incoherent frequency modulation of complex tones,'' J. Acoust. Soc. Am. **89**, 329–340.

Carlyon, R. P. (**1994**). ''Further evidence against an across-frequency mechanism specific to the detection of frequency modulation (FM) incoherence between resolved frequency components,'' J. Acoust. Soc. Am. **95**, 949–961.

Carlyon, R. P. (**2000**). ''Detecting coherent and incoherent frequency modulation,'' Hear. Res. **140**, 173–188.

Chalikia, M. H., and Bregman, A. S. (**1993**). ''The perceptual segregation of simultaneous vowels with harmonic, shifted, and random components,'' Percept. Psychophys. **53**, 125–133.

Cohen, M. F., and Chen, X. (**1992**). ''Dynamic frequency change among stimulus components: Effects of coherence on detectability,'' J. Acoust. Soc. Am. **92**, 766–772.

Culling, J. F., and Darwin, C. J. (**1993**). ''Perceptual separation of simultaneous vowels: Within and across-formant grouping by $F0$,'' J. Acoust. Soc. Am. **93**, 3454–3467.

Culling, J. F., and Darwin, C. J. (**1994**). ''Perceptual and computational separation of simultaneous vowels: Cues arising from low-frequency beating,'' J. Acoust. Soc. Am. **95**, 1559–1569.

Culling, J. F., and Summerfield, Q. (**1995**). ''The role of frequency modulation in the perceptual segregation of concurrent vowels,'' J. Acoust. Soc. Am. **98**, 837–846.

Darwin, C. J. (**1984**). ''Perceiving vowels in the presence of another sound: Constraints on formant perception,'' J. Acoust. Soc. Am. **76**, 1636–1647.

Darwin, C. J., and Gardner, R. B. (**1986**). ''Mistuning a harmonic of a vowel: Grouping and phase effects on vowel quality,'' J. Acoust. Soc. Am. **79**, 838–845.

de Cheveigné, A. (**1999**). ''Vowel-specific effects in concurrent vowel identification,'' J. Acoust. Soc. Am. **106**, 327–340.

de Cheveigné, A., McAdams, S., and Marin, C. M. H. (**1997a**). ''Concurrent vowel identification. II. Effects of phase, harmonicity and task,'' J. Acoust. Soc. Am. **101**, 2848–2856.

de Cheveigné, A., Kawahara, H., Tsuzaki, M., and Aikawa, K. (**1997b**). ''Concurrent vowel identification. I. Effects of relative amplitude and F_0 difference,'' J. Acoust. Soc. Am. **101**, 2839–2847.

Demany, L., and Semal, C. (**1988**). ''Dichotic fusion of two tones one octave apart: Evidence for internal octave templates,'' J. Acoust. Soc. Am. **83**, 687–695.

Furukawa, S., and Moore, B. C. J. (**1996**). ''Across-frequency processes in frequency modulation detection,'' J. Acoust. Soc. Am. **100**, 2299–2312.

Furukawa, S., and Moore, B. C. J. (**1997a**). ''Dependence of frequency modulation detection on frequency modulation coherence across carriers: Effects of modulation rate, harmonicity, and roving of the carrier frequencies,'' J. Acoust. Soc. Am. **101**, 1632–1643.

Furukawa, S., and Moore, B. C. J. (**1997b**). ''Effect of the relative phase of amplitude modulation on the detection of modulation on two carriers,'' J. Acoust. Soc. Am. **102**, 3657–3664.

Gardner, R. B., and Darwin, C. J. (**1986**). ''Grouping of vowel harmonics by frequency modulation: Absence of effects on phonemic categorization,'' Percept. Psychophys. **40**, 183–187.

Glasberg, B. R., and Moore, B. C. J. (**1990**). ''Derivation of auditory filter shapes from notched-noise data,'' Hear. Res. **47**, 103–138.

Gockel, H. (**1998**). ''On possible cues in profile analysis: Identification of the incremented component,'' J. Acoust. Soc. Am. **103**, 542–552.

Green, D. M. (**1988**). *Profile Analysis* (Oxford University Press, Oxford).

Grose, J. H., and Hall, J. W. (**1990**). ''The effect of modulation coherence on signal threshold in frequency-modulated noise bands,'' J. Acoust. Soc. Am. **88**, 703–710.

Grose, J. H., Hall, J. W., and Mendoza, L. (**1995**). ''Perceptual organization in a comodulation masking release interference paradigm: Exploring the role of amplitude modulation, frequency modulation, and harmonicity,'' J. Acoust. Soc. Am. **97**, 3064–3071.

Hacker, M. J., and Ratcliff, R. (**1979**). ''A revised table of d' for M-alternative forced choice,'' Percept. Psychophys. **26**, 168–170.

Hall, J. W., and Grose, J. H. (**1990**). ''Consequences of onset asynchrony for modulation detection interference,'' J. Acoust. Soc. Am. **88**, S145.

Hall, J. W., and Grose, J. H. (**1991**). ''Some effects of auditory grouping factors on modulation detection interference (MDI),'' J. Acoust. Soc. Am. **90**, 3028–3035.

Hall, J. W., Haggard, M. P., and Fernandes, M. A. (**1984**). ''Detection in noise by spectro-temporal pattern analysis,'' J. Acoust. Soc. Am. **76**, 50–56.

Johnson, D. H. (**1980**). ''The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones,'' J. Acoust. Soc. Am. **68**, 1115–1122.

Klein, W., Plomp, R., and Pols, L. C. (**1970**). ''Vowel spectra, vowel spaces, and vowel identification,'' J. Acoust. Soc. Am. **48**, 999–1009.

Kruskal, J. B. (**1964**). ''Nonmetric multidimentional scaling: A numerical method,'' Psychometrika **29**, 115–129.

Leek, M. R., Dorman, M. F., and Summerfield, Q. (**1987**). ''Minimum spectral contrast for vowel identification by normal-hearing and hearing-impaired listeners,'' J. Acoust. Soc. Am. **81**, 148–154.

Marin, C. M. H., and McAdams, S. (**1991**). ''Segregation of concurrent sounds. II. Effects of spectral envelope tracing, frequency modulation coherence, and frequency modulation width,'' J. Acoust. Soc. Am. **89**, 341–351.

McAdams, S. (**1984**). ''Spectral fusion, spectral parsing and the formation of the auditory image,'' Ph.D. thesis, Stanford.

McAdams, S. (**1989**). ''Segregation of concurrent sounds. I. Effects of frequency modulation coherence,'' J. Acoust. Soc. Am. **86**, 2148–2159.

McKeown, J. D., and Patterson, R. D. (**1995**). ''The time course of auditory segregation: Concurrent vowels that vary in duration,'' J. Acoust. Soc. Am. **98**, 1866–1877.

Moore, B. C. J. (**1986**). ''Parallels between frequency selectivity measured psychophysically and in cochlear mechanics,'' Scand. Audiol. Suppl. **25**, 139–152.

Moore, B. C. J. (**2003**). *An Introduction to the Psychology of Hearing*, 5th ed. (Academic, San Diego).

Moore, B. C. J., and Ohgushi, K. (**1993**). ''Audibility of partials in inharmonic complex tones,'' J. Acoust. Soc. Am. **93**, 452–461.

Moore, B. C. J., and Sek, A. (**1996**). ''Detection of frequency modulation at low modulation rates: Evidence for a mechanism based on phase locking,'' J. Acoust. Soc. Am. **100**, 2320–2331.

Moore, B. C. J., and Shailer, M. J. (**1992**). ''Modulation discrimination interference and auditory grouping,'' Philos. Trans. R. Soc. London, Ser. B **336**, 339–346.

Moore, B. C. J., Glasberg, B. R., and Peters, R. W. (**1985a**). ''Relative dominance of individual partials in determining the pitch of complex tones,'' J. Acoust. Soc. Am. **77**, 1853–1860.

Moore, B. C. J., Peters, R. W., and Glasberg, B. R. (**1985b**). ''Thresholds for the detection of inharmonicity in complex tones,'' J. Acoust. Soc. Am. **77**, 1861–1867.

Moore, B. C. J., Glasberg, B. R., and Peters, R. W. (**1986**). ''Thresholds for hearing mistuned partials as separate tones in harmonic complexes,'' J. Acoust. Soc. Am. **80**, 479–483.
Moore, B. C. J., Glasberg, B. R., Gaunt, T., and Child, T. (**1991**). ''Across-channel masking of changes in modulation depth for amplitude- and frequency-modulated signals,'' Q. J. Exp. Psychol. A **43A**, 327–347.
Plomp, R. (**1964**). ''The ear as a frequency analyzer,'' J. Acoust. Soc. Am. **36**, 1628–1636.
Plomp, R. (**1970**). ''Timbre as a multidimensional attribute of complex tones,'' in *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G. F. Smoorenburg (Sijthoff, Leiden).
Roberts, B., and Bregman, A. S. (**1991**). ''Effects of the pattern of spectral spacing on the perceptual fusion of harmonics,'' J. Acoust. Soc. Am. **90**, 3050–3060.
Roberts, B., and Brunstrom, J. M. (**2001**). ''Perceptual fusion and fragmentation of complex tones made inharmonic by applying different degrees of frequency shift and spectral stretch,'' J. Acoust. Soc. Am. **110**, 2479–2490.
Scheffers, M. T. M. (**1983**). ''Sifting vowels: Auditory pitch analysis and sound segregation,'' Ph.D. thesis, Groningen University, The Netherlands.
Schooneveldt, G. P., and Moore, B. C. J. (**1987**). ''Comodulation masking release (CMR): Effects of signal frequency, flanking-band frequency, masker bandwidth, flanking-band level, and monotic versus dichotic presentation of the flanking band,'' J. Acoust. Soc. Am. **82**, 1944–1956.
Schooneveldt, G. P., and Moore, B. C. J. (**1988**). ''Failure to obtain comodulation masking release with frequency-modulated maskers,'' J. Acoust. Soc. Am. **83**, 2290–2292.
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.
Spiegel, M. F., Picardi, M. C., and Green, D. M. (**1981**). ''Signal and masker uncertainty in intensity discrimination,'' J. Acoust. Soc. Am. **70**, 1015–1019.
Summerfield, A. Q., and Assmann, P. F. (**1991**). ''Perception of concurrent vowels: Effects of pitch-pulse asynchrony and harmonic misalignment,'' J. Acoust. Soc. Am. **89**, 1364–1377.
Summerfield, Q., and Culling, J. F. (**1992**). ''Auditory segregation of competing voices: Absence of effects of FM or AM coherence,'' Philos. Trans. R. Soc. London, Ser. B **336**, 357–366.
Wilson, A. S., Hall, J. W., and Grose, J. H. (**1990**). ''Detection of frequency modulation (FM) in the presence of a second FM tone,'' J. Acoust. Soc. Am. **88**, 1333–1338.
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.
Zwicker, U. T. (**1984**). ''Auditory recognition of diotic and dichotic vowel pairs,'' Speech Commun. **3**, 265–277.

Dominance region for pitch: Effects of duration and dichotic presentation[a)]

Hedwig Gockel[b)] and Robert P. Carlyon
MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 2EF, United Kingdom

Christopher J. Plack
Department of Psychology, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, United Kingdom

(Received 19 July 2004; revised 2 December 2004; accepted 3 December 2004)

The dominance region (DR) for pitch was determined for 16- and 200-ms complex tones containing the first seven harmonics of a fundamental frequency (F0) of 250 Hz. A tone was presented with one of the harmonics mistuned upwards or downwards by 3%, followed 500 ms later by a perfectly harmonic tone of the same duration. Listeners adjusted the F0 of the harmonic tone so that its pitch matched that of the mistuned complex. In experiment 1, stimuli were presented monaurally. The DR was significantly higher in harmonic number for the short than for the long duration. The overall sum of the pitch shifts produced by all harmonics was significantly larger for the short than for the long duration, presumably due to stronger perceptual fusion for the former. In experiment 2, the mistuned harmonic was presented only contralaterally to the remainder of the complex. A similar shift in the DR with duration was observed, although the pitch shifts were smaller than for monaural presentation. There was no significant effect of duration on the overall pitch shifts. The results are discussed in terms of pattern recognition and autocorrelation models of pitch perception, and a role of attention in pitch matching is suggested.
[DOI: 10.1121/1.1853111]

PACS numbers: 43.66.Hg, 43.66.Rq, 43.66.Ba [AJO] Pages: 1326–1336

I. INTRODUCTION

The question of which harmonics are the most important in determining the pitch of a complex tone (the residue pitch) has been the subject of a considerable amount of research (Bilsen and Ritsma, 1967; Plomp, 1967; Ritsma, 1967, 1970; Patterson and Wightman, 1976; Moore *et al.*, 1985; Dai, 2000). For example, Plomp (1967) presented his subjects with a perfectly harmonic complex tone of a given fundamental frequency (F0) followed by a complex whose components were split into two groups, the lower group of components having a slightly lower F0 and the higher group having a slightly higher F0. Subjects had to indicate which of the two tones was higher in pitch. He found that harmonic numbers 3–5 were dominant in determining the pitch. Ritsma (1967) used very similar methods and came to a similar conclusion. Plomp (1967) also found that the region of dominant harmonics depended somewhat on the F0; the harmonic numbers of the dominance region tended to decrease with increasing F0. Generally, the lower resolved harmonics have been found to be dominant in determining the residue pitch. In agreement with this, pattern recognition models of pitch perception assume that residue pitch is derived on the basis of information from harmonics that are resolved in the peripheral auditory system, and therefore these models predict the lower resolved components to be dominant (Terhardt, 1972; Goldstein, 1973; Wightman, 1973; Terhardt, 1974).

While the experiments of Ritsma and Plomp assessed a *region* of dominant harmonics, Moore *et al.* (1985) devised a method which enabled assessment of the relative dominance of *individual* partials in determining the pitch of complex tones. The method involves mistuning a single partial in the complex tone up or down by a certain percentage of its nominal (harmonic) frequency, and measuring the corresponding shift in the residue pitch. Using this method with a complex tone duration of 420 ms, Moore *et al.* (1985) found that, for mistunings up to ±2%–3%, the shift in the residue pitch was approximately a linear function of the shift in the harmonic. For greater mistunings, the shift in the residue was reduced. For all three subjects and for the three F0s tested (100, 200, 400 Hz), the dominant harmonics were contained within the six lowest partials. However, the exact distribution of dominance varied markedly between individuals. The distribution of dominance for each subject reflected the relative frequency difference limens (FDLs) for individual partials when presented within the complex tone, which had been measured for the same subjects by Moore *et al.* (1984); low FDLs were associated with high dominance. In other words, components that made a high contribution to the residue pitch tended to be those that had low FDLs, with the FDL for each component being determined for presentation within the complex. Moore *et al.* (1985) concluded that these results are compatible with Goldstein's (1973) pattern recognition model of pitch perception, according to which the contribution of a given component to the residue pitch depends upon the precision with which the frequency of that component is represented at the input to a ''central processor.'' However, they suggested that their results are also compatible with a slightly modified version of Terhardt's model (Terhardt,

[a)]Parts of this work were presented at the 147th meeting of the Acoustical Society of America, New York, New York, 24–28 May 2004 [J. Acoust. Soc. Am. **115**, 2389 (2004)].
[b)]Electronic mail: hedwig.gockel@mrc-cbu.cam.ac.uk

1979; Terhardt *et al.*, 1982a, b), in which the residue pitch is derived as a weighted average of the virtual pitches produced by the individual components instead of being determined solely by the virtual pitch of the single partial with the greatest weight in the ''virtual pitch pattern'' (as assumed by Terhardt and co-workers).

Up to now, research on the dominance of harmonics for residue pitch has focused on stimuli with relatively long durations. However, many natural sounds are quite brief and, thus, it is of interest to know whether the relative importance of harmonics for pitch depends on tone duration. The first objective of the present study was to assess whether there was an effect of duration. The second, more theoretical objective was to use the method devised by Moore *et al.* (1985) to test a prediction of pattern recognition models of pitch perception for the pattern of the dominance region. These models predict an effect of the duration of the stimulus on the distribution of dominance across individual partials. The reasoning behind this prediction is described below.

It is well established that the increase in the FDL with decreasing stimulus duration is more marked at low than at high frequencies (Liang and Chistovich, 1961; Sekey, 1963; Henning, 1970; Moore, 1973; Freyman and Nelson, 1986; Micheyl *et al.*, 1998). As mentioned above, according to pattern recognition models of pitch perception, the contribution of a given component to residue pitch depends upon the precision with which the frequency of that component is represented at the input to the central processor. Due to the frequency-dependent effect of duration on FDLs, which are generally considered to provide a measure of the precision of frequency encoding, one would expect that the contribution to the residue pitch from lower partials would be relatively smaller for short than for long stimulus durations. In contrast, higher partials should contribute relatively more for short than for long durations. More detailed predictions of pattern recognition models for the expected change in the dominance region will be presented below (Sec. III D).

The results of the first experiment showed that the predicted effect of duration on the relative contribution from lower and higher harmonics to the residue pitch does indeed occur. Naturally, the fact that a pattern recognition model does predict such an effect does not imply that other models of pitch perception are unable to account for it. Below (Sec. III E) we show that an autocorrelogram model (Meddis and Hewitt, 1991a, b; Meddis and O'Mard, 1997), which does not derive residue pitch from estimates of the frequencies of individual partials, is in principle also able to qualitatively predict the observed effect. The experimental results also showed that, overall, the contribution of single mistuned harmonics to the residue pitch was higher for the short than for the long stimulus duration. The second experiment assessed whether the same pattern of results occurred for contralateral presentation of the mistuned partial.

II. GENERAL METHOD AND STIMULI

A complex tone with a nominal F0 of 250 Hz, containing the first seven harmonics added in sine phase, with equal amplitude, was employed as the test stimulus in both experiments. The duration of the tones was either 16 or 200 ms, including 8-ms raised-cosine onset and offset ramps. The choice of F0 and durations was motivated by the desire to optimize the chance of observing an effect of duration on the dominance region (if it exists), and was based on Moore's (1973) data. Moore (1973) measured FDLs for pure tones of various durations with frequencies of 125, 250, 500, 1000, 2000, 4000 Hz and higher. For 125 Hz, performance at durations smaller than 40 ms was so poor for some subjects that a FDL could not be measured. For 250 Hz, a threshold could be measured for all subjects down to durations as small as 12.5 or 10 ms, and the increase in FDL relative to that observed for a duration of 200 ms was substantial; for one subject the FDL was increased by a factor of 15. Generally, the increase in FDL with decreasing duration was largest for frequencies at and below about 500 Hz. Thus, to increase the chance of observing an effect of duration on the dominance region, the F0 should be low enough for the tone to have at least one dominant harmonic below about 500 Hz. Also, the short duration complex tone should be as short as possible, while still allowing the subjects to be able to perceive a pitch. When choosing the parameters of the present stimuli according to these requirements, it was assumed that, as a first approximation, the effect of duration on the FDL as a function of frequency does not depend greatly on whether the tone is presented on its own or as part of a harmonic complex.

A two-interval, two-alternative adaptive procedure was used to obtain pitch matches between a complex that had one of its partials mistuned by $\pm 3\%$ and a perfectly harmonic complex tone; the F0 of the latter was adjusted to obtain the match. A mistuning of $\pm 3\%$ was chosen in order to maximize the expected pitch shifts and because, as was mentioned above, Moore *et al.* (1985) found that, for mistunings up to $\pm 2\%$–3%, the shift in the residue pitch was approximately a linear function of the shift in the harmonic. The complex tone with the mistuned partial was always presented in the first interval. After an interstimulus interval of 500 ms, the harmonic complex was presented. Each interval was marked by a light. Subjects were required to indicate the interval containing the complex with the higher pitch. A two-down, two-up rule was used to adjust the F0 of the harmonic complex. After two consecutive answers of the same type, the F0 of the harmonic complex was increased if the first interval was judged higher and decreased if the second interval was judged higher. This procedure tracks the point corresponding to 50% ''higher'' (or ''lower'') judgments of the adjustable harmonic complex. The starting step size was 5 Hz. Following each turnpoint (a change in the direction of the F0 change), the step size was halved until the minimum step size of 0.125 Hz was reached.

The starting F0 of the harmonic complex was varied quasirandomly in the range 250 ± 25 Hz. For each mistuned partial there were always four runs in a sequence; in two the partial was mistuned upwards and in two it was mistuned downwards. One of the two runs for each mistuning direction had a starting F0 for the harmonic complex higher than 250 Hz, and one had an F0 lower than 250 Hz. The different partials were tested in a controlled order, from low harmonic number to high harmonic number to low harmonic number,

etc. Two buttons on a response box, marked 1 and 2, respectively, were used to indicate the interval with the higher pitch. By pressing a third button (underneath a red LED), subjects initiated the presentation of the same stimulus pair again. By pressing a fourth button (underneath a green LED) subjects indicated when they were satisfied with the pitch match. Subjects were encouraged to ''bracket'' the overall pitch of the complex containing a mistuned partial with the pitch of the harmonic complex, that is, to go from higher to lower to higher, etc. They were also encouraged to listen a few times to the same stimulus pair before indicating the interval with the higher pitch. Only after they reached the minimum step size was their response ''Pitch of the two tones is the same'' accepted. If a subject pressed the fourth button before the minimum step size was reached, the response was ignored and the same stimulus pair was presented again. Subjects were informed about the procedure, but no feedback was provided.

The matching pitch was defined as the F0 of the harmonic complex presented immediately before the subject indicated that the pitch in the two intervals was equal. At least ten pitch matches were obtained for each condition and subject. The data reported are the means of these ten (or more) pitch matches. The total duration of a single session was about 2 h, including rest times. To familiarize subjects with the procedure and equipment, they were given about 2 h of practice.

The complex tones were presented at a level of 65 dB SPL per component in a continuous pink background noise that had a spectrum level of 9 dB at 1 kHz (re 20 μPa). The pink noise was presented so that the individual components would have approximately equal sensation levels. The tones were generated digitally, and were played out using a 16-bit digital-to-analog converter (CED 1401 plus), with a sampling rate of 10 kHz. Stimuli were passed through an antialiasing filter (Kemo 21C30) with a cutoff frequency of 4.3 kHz (slope of 96 dB/oct), and presented using Sennheiser HD250 headphones. Subjects were seated individually in an IAC double-walled sound-attenuating booth.

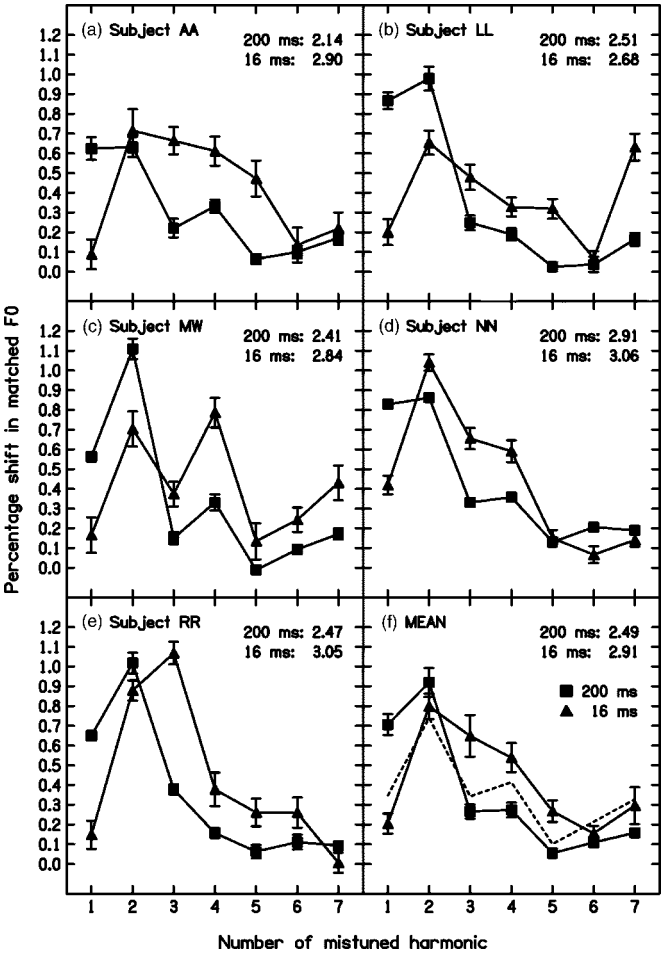


FIG. 1. Results of experiment 1 (monaural stimulus presentation) for the five individual subjects [panels (a)–(e)], and the mean averaged over all subjects [panel (f)], together with the corresponding standard errors. For the mean results [panel (f)] the standard error is the error of the mean of the average pitch shifts across subjects. The mean *absolute* pitch shift was calculated as half the difference in pitch matches between corresponding positive and negative mistunings (3%) and expressed as a percentage of the F0 (250 Hz). The squares and the triangles show pitch shifts as a function of the number of the mistuned harmonic for stimulus durations of 200 and 16 ms, respectively. The sum of the pitch shifts for each harmonic, expressed as a percentage of the F0, is indicated in the upper-right corner in each panel, separately for each duration. The dashed line in panel (f) shows the predictions of a pattern recognition model for the 16-ms duration, based on the observed pitch shifts for the 200-ms duration. See text for details.

III. EXPERIMENT 1: MONAURAL PRESENTATION

A. Stimuli and listeners

To investigate the basic effect of stimulus duration on the pattern of the dominance region, all stimuli were presented monaurally. Five subjects participated in all conditions. They ranged in age from 22 to 30 years, and their quiet thresholds at octave frequencies between 250 and 4000 Hz were within 15 dB of the 1969 ANSI standard. All of them had considerable musical experience and were screened for their ability to give reliable pitch matches for *both* durations. Especially for the short duration, the task was not easy, and, initially, two more subjects were screened and then excluded as they were unable to give reliable pitch matches at the short duration. Pitch matches were considered to be unreliable when matches to the same stimuli differed by more than about 20 Hz.

B. Results

The results are presented (following Moore *et al.*, 1985) as the mean *absolute* pitch shift, which is calculated as half the difference in pitch matches between corresponding positive and negative mistunings, expressed as a percentage of the F0. Figure 1 shows the results of the individual subjects [panels (a)–(e)] and the mean results averaged over all subjects [panel (f)], together with the corresponding standard errors; for the mean results [panel (f)] the standard error is the error of the mean of the average pitch shifts across subjects. The squares and the triangles show pitch shifts as a function of the number of the mistuned harmonic for stimulus durations of 200 and 16 ms, respectively. Also indicated in each panel is the sum of the pitch shifts over all harmonics, expressed as a percentage of the F0, separately for each of the two stimulus durations.

For the 200-ms duration, mistuning of the second and the first harmonics generally produced the largest pitch shifts. Mistuning of one of the higher harmonics had a reduced effect on the pitch. The smallest pitch shifts were observed when the fifth harmonic was mistuned. For the 16-ms duration, typically the standard errors were larger than for the 200-ms duration, and the pattern of results varied more between subjects. For most subjects, mistuning of the second, the third, or the fourth harmonic produced the largest pitch shifts. For two of the subjects, the seventh harmonic was among the three with the largest contribution to pitch. Most importantly, for all subjects, mistuning of the first harmonic produced a markedly smaller shift in pitch for the 16-ms duration than for the 200-ms duration. In contrast, for all subjects, mistuning of the third or fourth harmonic produced a larger pitch shift for the 16-ms than for the 200-ms duration. Also, the sum of the pitch shifts produced by all harmonics was larger, and closer to 3%, for the 16-ms duration (2.91%) than for the 200-ms duration (2.49%). The sum of the pitch shifts would be expected to be 3%, if the pitch is determined as a weighted average of pitch cues derived from the individual partials, as proposed by Moore *et al.* (1985). This is based on the assumption that a harmonic, when mistuned by 3%, makes the same contribution as it would if it were not mistuned.

To determine the statistical significance of these effects, a repeated measures two-way ANOVA (with factors duration and harmonic number) was calculated, using the mean pitch shift in percent for each subject and condition as input. The ANOVA showed significant main effects of harmonic number [$F(6,24)=19.8$, $p<0.001$][1] and duration [$F(1,4)=12.9$, $p<0.05$], the latter reflecting the overall larger pitch shift for the shorter duration. Most importantly here, the interaction between duration and harmonic number was highly significant [$F(6,24)=13.0$, $p<0.001$]. *Post hoc* contrasts based on Fisher's least significant difference procedure showed that pitch shifts were significantly smaller for the 16-ms than for the 200-ms duration when the first harmonic was mistuned ($p<0.001$), but were significantly larger for the 16-ms than for the 200-ms duration when the third, fourth, or fifth harmonic was mistuned ($p<0.05$ for all hamonics). For all other harmonics (second, sixth, and seventh) there was no significant difference between pitch shifts for the short and long durations. Thus, an effect of duration on the dominance region for pitch was indeed observed, namely a relatively smaller contribution from low harmonics and relatively greater contribution from higher harmonics for the short duration than for the long duration.

C. Discussion

As was also reported by Moore *et al.* (1985), individual differences occurred with respect to which harmonics had the greatest influence on residue pitch. However, the differences were smaller for the 200-ms than for the 16-ms duration. For the 200-ms duration, individual differences were smaller than those observed by Moore *et al.* (1985) for an F0 of 200 Hz and a duration of 420 ms (the condition most similar to the present one); in the present data, the contribution to the residue pitch was highest for the second and the first harmonics for all five subjects, while the pattern was more variable in Moore *et al.*'s data.

While Moore *et al.* found no contribution from harmonics higher than the sixth, three out of our five subjects showed some contribution to pitch from the seventh component. This was probably due to the fact that, in the present experiment, the seventh component was the highest component present and therefore would have been masked less by neighboring higher components than in Moore *et al.*'s study, where a 12-component complex was used (see also Moore *et al.*, 1984).

For the 200-ms duration, the sum of the pitch shifts produced by each of the mistuned harmonics was 2.49%. This is slightly smaller than the 2.8% found by Moore *et al.* (1985) for the same amount of mistuning (3%), averaged across their three F0s (100, 200, and 400 Hz), and a stimulus duration of 420 ms [read from Fig. 1(b) in Moore *et al.*]. For the 16-ms duration, the sum of the pitch shifts was significantly larger (2.91%) than for the long duration. This can probably be explained by the fact that for the short duration the mistuned components were less likely to be perceptually segregated (Moore *et al.*, 1986) and by assuming that harmonics that are fused more contribute more to the residue pitch. Compatible with the latter assumption, Moore *et al.* (1985) found reduced pitch shifts for mistunings larger than 3%, when there would have been increased segregation of the mistuned harmonic from the remainder of the complex (Moore *et al.*, 1986). In fact, Darwin (1992) assumed that the pitch shift caused by a mistuned harmonic is a joint function of the degree to which it is mistuned and the degree to which it is perceptually segregated (due to mistuning). Based on this assumption, he successfully modeled the nonmonotonic variation in pitch shifts with increasing mistuning by assuming that the contribution of a harmonic to pitch is reduced monotonically (according to a Gaussian curve) as the mistuning increases, i.e., as the perceptual segregation increases.

We are not aware of any other study that has investigated the effect of stimulus duration on the *sum* of the pitch shifts from all partials. There are, however, two studies which investigated the effect of duration on pitch shifts produced by mistuning a single partial only. Moore (1987) compared the pitch shift obtained by mistuning a harmonic, chosen individually for each subject to be one of the most dominant, for stimulus durations of 50 and 410 ms and for various degrees of mistunings. He found that the pitch shift was significantly greater for the 50-ms than for the 410-ms duration only for mistunings of 6% and larger, but not for mistunings of 4% and smaller. His conclusion was that the tolerance of the pitch mechanism for including mistuned partials increases with decreasing stimulus duration. Based on this and the results of his earlier study (Moore *et al.*, 1985), one would not necessarily have expected the effect of stimulus duration seen in the present data.

The second study that investigated the effect of duration on the contribution of one mistuned partial to the residue pitch was presented by Darwin and Ciocca (1992). They found that, when the fourth harmonic of a complex tone with an F0 of 155 Hz was mistuned by 1.6%, 3.2%, or 8%, the

pitch shift was *always* slightly but significantly larger for a duration of 90 ms than for a duration of 410 ms. Darwin and Ciocca attributed the discrepancy between their own and Moore's (1987) findings to the fact that they did not choose a harmonic individually for each subject to be a dominant one, and possibly to the differences in F0 and duration between the two studies. The results of our first experiment, together with Darwin and Ciocca's findings, support the existence of an effect of duration on the contribution of mistuned harmonics to residue pitch even for a relatively small amount of mistuning.

The main finding of experiment 1 was the effect of duration on the pattern of the dominance region, i.e., a relatively smaller contribution from low harmonics and relatively greater contribution from higher harmonics for the short duration than for the long duration. Because FDLs for pure tones increase with decreasing duration and do so more at low than at high frequencies (Liang and Chistovich, 1961; Sekey, 1963; Henning, 1970; Moore, 1973; Freyman and Nelson, 1986; Micheyl *et al.*, 1998), this finding is qualitatively consistent with predictions derived from pattern recognition models of pitch perception which state that the contribution of a given component to the residue pitch depends upon the precision with which the frequency of that component is represented at the input to the central processor (Goldstein, 1973).

So far we have not considered why FDL increases more with decreasing duration at low frequencies than at high frequencies. However, previous data have shown that this pattern of results is compatible with frequency discrimination being mainly based upon temporal information for frequencies below about 5 kHz and upon ''place'' information for frequencies above 5 kHz (Moore, 1973). Mounting evidence suggests that the increase in the FDL with decreasing duration for frequencies below about 5 kHz is due to the decreasing number of periods available for sampling by a temporal mechanism and not due to spectral splatter (Moore, 1973; Goldstein and Srulovicz, 1977; Micheyl *et al.*, 1998). However, note that Goldstein (1973) pointed out that, in his optimum processor theory, the frequencies of individual harmonics might be derived from either place or temporal information. Thus, pattern recognition models of pitch perception, and consequently the prediction derived from pattern recognition models that has been tested here, do not depend upon a specific assumption about the origin of the information about component frequencies.

D. Predictions of a pattern recognition model

We next present more quantitative predictions of a pattern recognition model of the effect of shortening the stimulus duration from 200 to 16 ms on the dominance region. It was assumed that, as a first approximation, the effect of duration on the FDL as a function of frequency is the same when the tone is presented on its own and when it is presented as harmonic within a complex tone. This might not be exactly true, but it seems very unlikely that FDLs as a function of duration when measured in the presence of a complex tone would show a qualitatively different pattern from that observed for pure tones. Thus, in the absence of any published data on the effect of duration on FDLs for individual harmonics within complex tones, the following predictions were based on FDLs for pure tones at various durations as measured by Moore (1973). The increases in FDL for tones with a duration of 16 ms relative to those with a duration of 200 ms were taken to be 8.96, 5.48, 3.42, 2.82, 2.39, 2.26, and 2.12 (expressed as a factor) for tones with frequencies of 250, 500, 750, 1000, 1250, 1500, and 1750 Hz, respectively. These values were obtained by linear interpolation between the geometric means of the values determined by Moore (1973) for three subjects for frequencies of 250, 500, 1000, and 2000 Hz, when comparing FDLs for a duration of either 12.5 and 25 ms (one subject) or 10 and 40 ms (2 subjects) to FDLs for a duration of 200 ms.

As mentioned in the Introduction, Moore *et al.* (1985) showed that the distribution of dominance across partials for each subject reflected the relative FDLs for individual partials when presented within the complex tone (measured for the same subjects by Moore *et al.*, 1984); low FDLs were associated with high dominance, and high FDLs were associated with low domininance. These authors reported a correlation of 0.84 between the dominance of a partial and the reciprocal of the normalized FDL, where normalization means expressing each FDL as a percentage and dividing it by the lowest FDL obtained for that subject. Therefore, for each subject in the present study, the FDLs for individual partials within the 200-ms hamonic complex should be inversely related to the pitch shifts measured for a duration of 200 ms.

In order to obtain predictions of the dominance pattern for a duration of 16 ms, for each of the present subjects, the pitch shifts obtained for the 200-ms duration were divided by the above frequency-dependent factors describing the increase in FDLs for a duration of 16 ms relative to those for a duration of 200 ms. Following this, all (pitch shift) values were rescaled by a constant factor (4.41) chosen such that the overall sum of the predicted pitch shifts for the 16-ms duration was equal to the overall sum of the observed pitch shifts for the 200-ms duration. Hence, this simple model makes *predictions* of the effect of a shortened duration on the *relative* contribution of individual partials. It does not make an *a posteriori* fit to the data in the 16-ms conditions. Thus, it does not try to model other possible effects such as changes in the fusion of partials due to the shortening of duration, which in turn might affect (overall) pitch shifts.

The predictions of the mean data at the 16-ms duration are shown by the dashed line in Fig. 1(f). Relative to the 200-ms duration, pitch shifts for the 16-ms duration are predicted to decrease for the first and second harmonics and increase for all harmonics above the second. This is generally the case (except for the sixth harmonic). For harmonics one, two, six, and seven the predicted pitch shifts for the 16-ms duration were close to the observed values for the short duration. For harmonics three, four, and five, predicted shifts were, however, somewhat below the measured pitch shifts. Thus, the observed changes in the contribution of individual harmonics to the residue pitch with shortening of the stimulus duration go in the predicted direction. However, quantitatively the predictions are far from perfect. There are

several possible reasons for this. Firstly, and most importantly, the effect of duration on the precision of frequency encoding was estimated from pure tone FDLs. Predictions might improve if they were based on measurements of the effect of duration on FDLs for components within a complex tone. Secondly, the effect of duration on the precision of frequency encoding was estimated from data of different subjects. Thirdly, the prediction was based on pattern recognition models and therefore does not include possible effects of an overall (or differential) increase in pitch shifts due to an increase in fusion of the components for the shorter duration.

Overall, the observed effect of duration on the dominance region was broadly consistent with predictions derived from a pattern recognition model. Note that the predicted effect of duration on the dominance region depends on the combination of F0 and the shortest duration chosen for the stimuli. If a higher F0 had been chosen, e.g., 500 Hz, and if a longer duration than the present value of 16 ms had been used, then smaller effects (if any) would be expected.

E. Predictions of an autocorrelogram model of pitch perception

Here we consider whether the observed effect of duration on the dominance region can be predicted by a model which does not derive the residue pitch from estimates of the frequencies of the individual partials. The model considered is the autocorrelogram model of pitch perception (Meddis and Hewitt, 1991a, b; Meddis and O'Mard, 1997). Predictions of the pitch matches to our test stimuli were derived using this model and the method suggested by Meddis and colleagues (Meddis and Hewitt, 1991a, b; Meddis and O'Mard, 1997). The model includes the following stages: (1) A second-order bandpass filter with lower and upper cutoff frequencies of 450 Hz and 8.5 kHz, respectively, to simulate the operation of the outer and middle ear; (2) A gammatone filterbank, comprising 60 channels evenly spaced on an ERB_N scale (Glasberg and Moore, 1990; Moore, 2003) with center frequencies between 100 Hz and 8 kHz, simulating the mechanical filtering on the basilar membrane; (3) Simulation of the mechanical to neural transduction at the hair cell (Meddis, 1986; 1988; Meddis *et al.*, 1990), which gave the probability of occurrence of a spike in the auditory nerve fibers for each of the 60 channels as a function of time; (4) A running autocorrelation function (ACF) with a given time constant (τ), calculated separately within each channel, which provides an estimate of the distribution of intervals between all spikes originating from fibers within a given channel, measured at a certain offset time (the time constant τ and the offset were parameters whose effect on the predictions was investigated; see below); (5) A summary ACF, derived by averaging the ACFs across all channels. The summary ACF was calculated for a given test stimulus, e.g., the inharmonic complex with the third component mistuned upwards, and for many perfectly harmonic complexes with the same duration but varying in F0; F0s varying from 246.25 to 253.75 Hz (250 Hz$\pm$1.5%) in steps of 0.125 Hz were used. The F0 giving the smallest Euclidean distance (D^2) between the summary ACFs of the inharmonic test stimulus and the harmonic ''matching'' stimulus was taken as the predicted match to the inharmonic stimulus. D^2 was computed over a range of lags restricted around the ''F0 region,'' containing the major peak in the summary ACF (lags corresponding to between 225 and 275 Hz, except where explicitly stated). The shifts in F0 determined for the complex with the positively mistuned partial and the complex with the negatively mistuned partial were averaged, as in the experiment.

The input stimuli were identical to the ones used in the experiment, with one exception. The sample rate was increased to 100 kHz, in order to have reasonable resolution in the summary ACF (about 0.25% of the F0 of 250 Hz). The time constant τ of the running ACF calculations was set either to 2.5 ms, as in Meddis and Hewitt (1991a, b), or to 3.5 ms. Meddis and O'Mard (1997) used a time constant of 10 ms for their predictions of the dominance of individual harmonics, as described by Moore *et al.* (1985). We were unable to use such a large time constant, because the short stimulus duration of 16 ms allowed a maximum value of 3.5 ms. This restriction to a maximum time constant of 3.5 ms follows from (*i*) the assumption that the calculation of the ACF over time should be restricted to the time where the stimulus is actually present, i.e., it should not include time before[2] or after stimulus presentation, and (*ii*) the fact that the ACF is calculated over a duration *T*, with *T* equal to three times τ (Meddis and Hewitt, 1991a), which means that a stimulus with duration *T* plus the maximum lag must be available. The time point at which to measure the ACF (the offset) was fixed in two ways: (*i*) At the end of the stimulus (the autocorrelation is calculated backwards in time) and (*ii*) for the stimuli of 16-ms duration, it was set to 12 ms for the inharmonic stimuli, and it was slightly adjusted around 12 ms, to include a full number of cycles, for the harmonic stimuli. For the stimuli of 200-ms duration, it was set to 100 ms for the inharmonic stimuli, and was slightly adjusted around 100 ms, to include a full number of cycles, for the harmonic stimuli. The latter way of slightly adjusting the offset for harmonic stimuli follows the method suggested by Meddis and Hewitt (1991a, b) and Meddis and O'Mard (1997).

Figure 2 shows the predicted pitch shifts (as a percentage of the F0) as a function of the number of the mistuned harmonic. The squares and the triangles show predicted pitch shifts as a function of the number of the mistuned harmonic for stimulus durations of 200 and 16 ms, respectively. Panel (a) shows the predictions for an offset at the end of the stimulus with τ equal to 3.5 ms. With this choice of parameters, the predicted pitch shift patterns are qualitatively similar to the observed ones. For panel (b) the offset is unchanged (at the end of the stimulus), but τ equals 2.5 ms. The predictions with this choice of parameters do *not* show less contribution from the first harmonic for the 16-ms duration than for the 200-ms duration stimulus. In panel (c), the offset is set to the end of a full cycle for the harmonic stimuli, with τ equal to 2.5 ms. For the predictions shown in panel (c), D^2 was computed over a range of lags corresponding to F0s between 231.5 and 275 Hz, instead of the ''normal'' range of 225 to 275 Hz. This was necessary because, for a 12-ms offset, the maximum lag for which ACFs could be calculated was 4.32 ms (231.5 Hz). Panel (c) shows that there is rea-

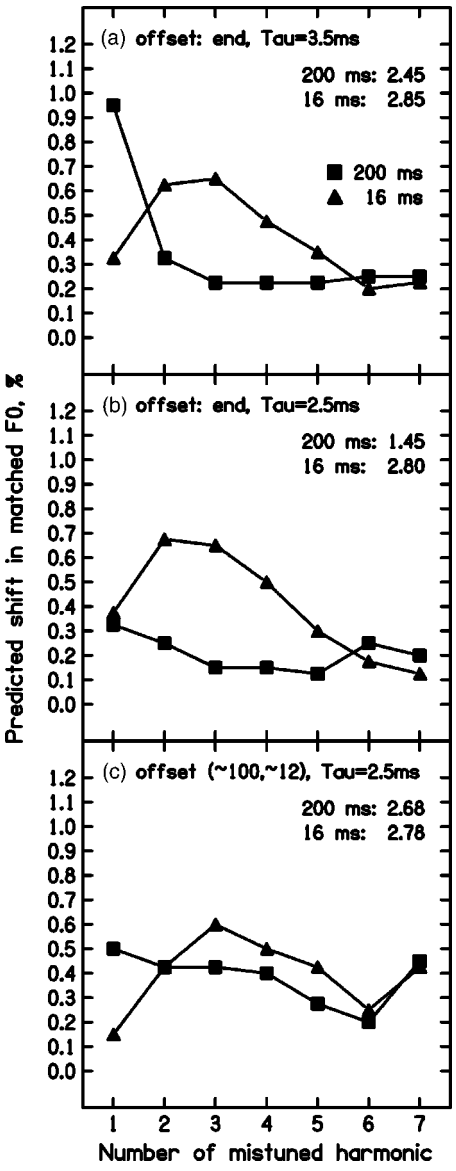


FIG. 2. Predictions of the pitch shifts from an autocorrelogram model of pitch perception (Meddis and Hewitt, 1991a, b; Meddis and O'Mard, 1997). Panels (a) and (b) show predictions for different values of the time constant τ (see text for details), with the time of calculation of the ACF (offset) fixed at the end of the stimulus. Panel (c) shows predictions when the offset was set to a time in the stimulus that, in the case of the harmonic stimuli, corresponded to a full number of periods for each individual F0. Otherwise, as in Fig. 1.

sonable agreement between predictions and data [shown in Fig. 1(f)]. Note also that in all cases the predicted sum of the pitch shifts produced by all harmonics is larger for the 16-ms than for the 200-ms duration. This is also in agreement with the data.

In summary, the accuracy of the predictions depends on the time constant and the offset chosen. The best fit between predictions and data was seen when the offset corresponded to a whole number of cycles for the harmonic stimuli. This was the method used before by Meddis and Hewitt (1991a, b) and Meddis and O'Mard (1997). However, it should be noted that this choice corresponds to assuming additional, but unformalized knowledge within the model; it is not obvious how the auditory system would determine this specific point in time. Reasonable predictions were also derived when the offset was fixed at the end of the stimulus and τ was equal to 3.5 ms.

At first sight, it seemed surprising that the autocorrelogram model could account for the observed interaction between duration and harmonic number. Some insight into the reasons for this can be obtained by investigating the effect of the 8-ms onset and offset ramps applied to our stimuli. The reasoning behind this was that, especially for the short stimulus duration, the ramps might lead to a ''deflation'' of the longer lags (lower frequencies) relative to the shorter lags (higher frequencies) in the calculation of the summary ACF; due to the ramps, the calculation of the ACF is based on less neural activity when the lags are long than when they are short (especially for the short duration) and this would reduce the contribution to the summary ACF from channels with low center frequencies relative to that from channels with high center frequencies. To test this idea, calculations were repeated for stimuli which had the same overall duration and 1-ms raised-cosine onset and offset ramps. Generally, shortening the ramps gave predictions that deviated more from the observed data pattern. Only when the offset corresponded to a complete number of cycles for the harmonic stimuli was there some qualitative agreement between the data and the predictions derived for stimuli with 1-ms ramps.

In conclusion, the observed effect of stimulus duration on the dominance region *can*, at least qualitatively, be predicted by the autocorrelation model of pitch perception (Meddis and Hewitt, 1991a, b; Meddis and O'Mard, 1997). The observed effect is thus consistent with both a pattern recognition model of pitch perception and a model that does not derive the residue pitch from estimates of the frequencies of the individual partials. It should be noted, however, that the predictions of the autocorrelogram model were highly dependent on the specific parameter values chosen; for example, choosing the offset as an integer number of cycles of the harmonic stimulus improved predictions, but this corresponds to providing information to the model that is not derived from the stimulus itself.

IV. EXPERIMENT 2: CONTRALATERAL PRESENTATION OF THE MISTUNED HARMONIC

The objective of the second experiment was to assess the effect of presenting the mistuned harmonic contralaterally to the remainder of the complex. It is not obvious how the pattern of pitch shifts obtained in this case would compare with that for monaural presentation. On the one hand, contralateral presentation would lead to an increase in peripheral resolution of the mistuned component which in turn (according to pattern recognition models of pitch perception) might lead to an increased contribution of that component to the overall pitch. This increase in contribution to pitch would be expected to be larger for the higher than for the lower harmonics. The reason for this expectation is as follows. For monaural presentation, the greater masking of higher harmonics causes them to have higher FDLs, when expressed as a percentage of their center frequency, than the lower harmonics (Moore *et al.*, 1984). In contrast, for contralateral

presentation, we would expect the pattern of FDLs to follow those for pure tones in isolation, and be lower at higher frequencies (Moore, 2003). Consequently, the reduction in FDL and therefore the increase in contribution to pitch, produced by contralateral presentation, should be much greater at higher than at lower harmonic numbers.

On the other hand, Darwin and Ciocca (1992) reported that the pitch shift for a 12-component complex tone with a duration of 90 ms was *slightly* but significantly smaller when a single mistuned harmonic (the fourth) was presented to the opposite ear to the remaining 11 harmonics than when presented to the same ear (see also Darwin *et al.*, 1992). Thus, in spite of the increased resolution of the mistuned harmonic when presented contralaterally, its contribution to the pitch decreased relative to that for monaural presentation, probably because it was more likely to be perceptually segregated. It is not obvious what the relative sizes of these counteracting effects are for different harmonic numbers and durations.

A. Stimuli and listeners

The stimuli were the same as in experiment 1, except that the mistuned harmonic was presented *only* contralaterally to the remainder of the complex. Six subjects participated. Four of them were run at both durations (16 and 200 ms), and three of those four had also participated in experiment 1. One subject, who had also participated in experiment 1, was only available for a short time and so took part in only the 16-ms duration condition. One new subject performed matches with the 200-ms stimuli in the dichotic and monaural conditions, but was unable to give reliable pitch estimates for the short duration. For some of the subjects who had participated in experiment 1, a single monaural condition was rerun after completing the dichotic conditions of experiment 2, to check the stability of the earlier results. This check showed that the results for the monaural condition were not significantly changed after participation in the dichotic experiment.

Subjects ranged in age from 22 to 30 years. Their quiet thresholds at octave frequencies between 250 and 4000 Hz were within 15 dB of the 1969 ANSI standard and were similar for the two ears. All of them had considerable musical experience.

B. Results

Figure 3 shows the results for the 200-ms duration for those subjects who were also run in the monaural condition. Panels (a)–(d) show the results for individual subjects, and panel (e) shows the mean results for all four subjects, together with the corresponding standard errors. The empty squares connected by the dashed lines show the pitch shifts in the dichotic condition. For comparison, the filled black squares connected by the solid lines show the results for the monaural condition; for three out of the four subjects these are replotted from Fig. 1. There was a trend for the pitch shifts to be smaller in the dichotic than in the monaural condition; the sum of the pitch shifts from all harmonics, expressed as a percentage of the F0, was only 1.65 in the dichotic condition, but was 2.41 in the monaural condition. However, an ANOVA showed that this trend was not significant [$F(1,3)=7.12$, $p=0.076$], presumably due to the results of one subject (RC), who showed similar shifts for monaural and dichotic presentation. The interaction between mode of presentation (monaural or dichotic) and harmonic number also was not significant.

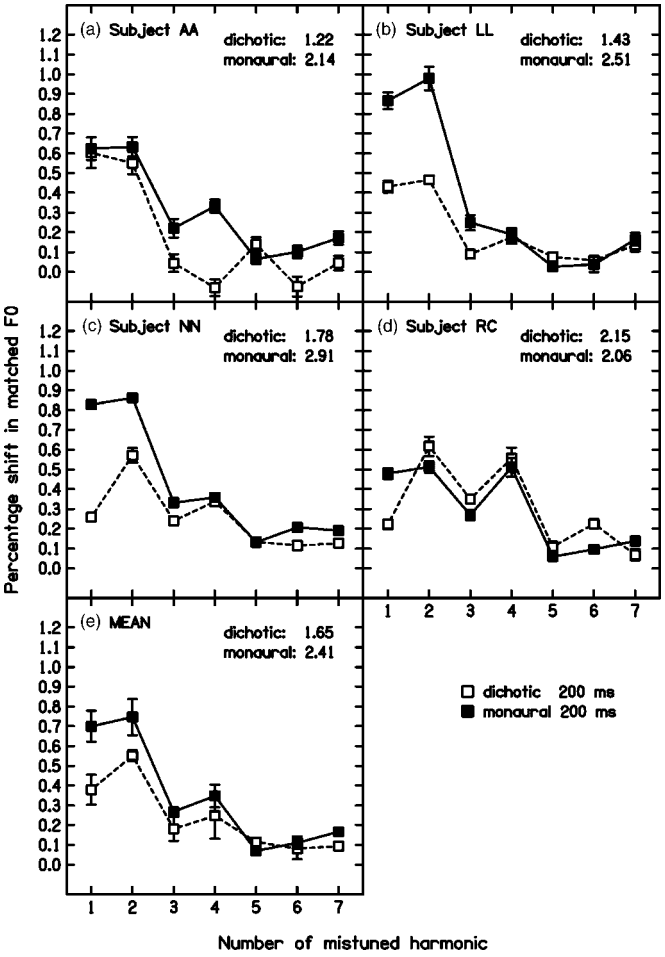


FIG. 3. Results of experiment 2 (contralateral presentation of the mistuned harmonic) and, for comparison, of experiment 1, for a stimulus duration of 200 ms. Panels (a)–(d) show the pitch matches for the four individual subjects, and panel (e) shows the mean over all subjects together with the corresponding standard errors. The empty squares connected by dashed lines and the filled squares connected by solid lines show pitch shifts obtained with dichotic presentation (experiment 2) and with monaural presentation (experiment 1), respectively. The sum of the pitch shifts for each harmonic, expressed as a percentage of the F0, is indicated in the upper-right corner in each panel, separately for each mode of presentation.

Figure 4 shows the results for the 16-ms duration for those subjects who were also run in the monaural condition. Individual data are shown in panels (a)–(d), and panel (e) shows the mean results. The empty triangles connected by the dashed lines show the pitch shifts for the dichotic condition. The results for the monaural condition are replotted from Fig. 1 and shown as the filled black triangles connected by the solid lines. For this short duration, there was a clear and highly significant effect of mode of presentation [$F(1,3)=312.81$, $p<0.001$]; the pitch shifts were markedly larger in the monaural than in the dichotic condition (the sum of the pitch shifts was 2.87 in the monaural condition and only 0.95 in the dichotic condition). There was also a signifi-

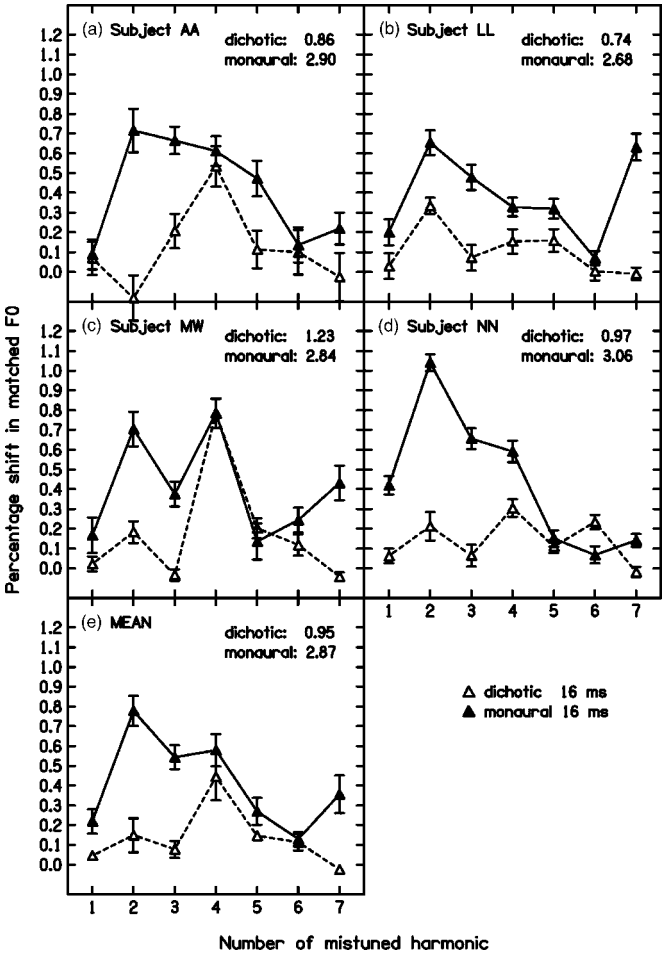


FIG. 4. As in Fig. 3, but for a stimulus duration of 16 ms. The empty triangles connected by dashed lines and the filled triangles connected by solid lines show pitch shifts obtained with dichotic presentation (experiment 2) and with monaural presentation (experiment 1), respectively.

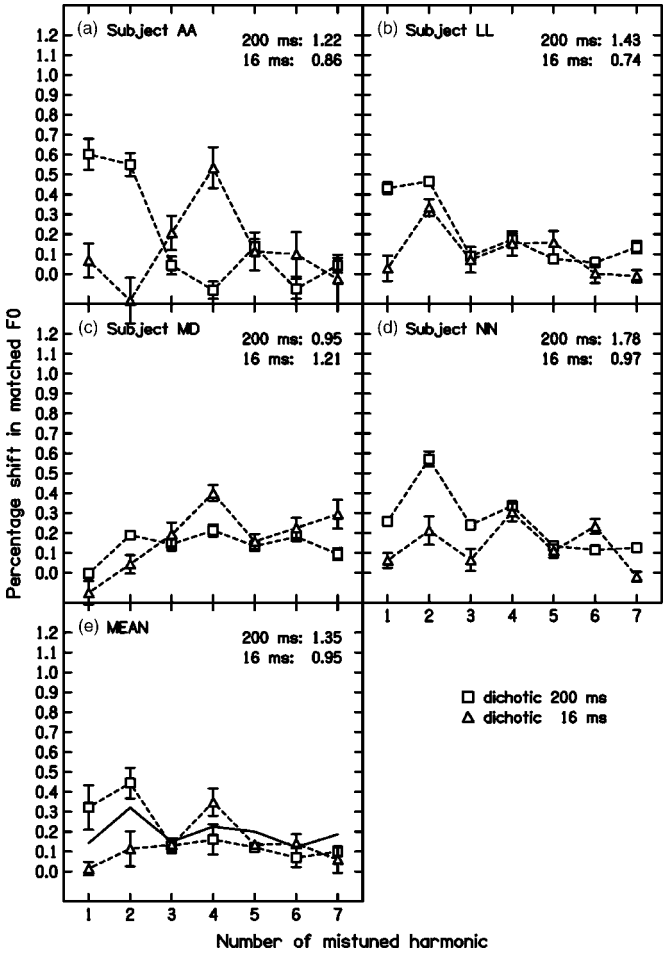


FIG. 5. Results of experiment 2 (contralateral presentation of the mistuned harmonic) for both stimulus durations. Panels (a)–(d) show pitch matches from four individual subjects, and panel (e) shows the mean over all subjects together with the corresponding standard errors. The squares and the triangles show pitch shifts for stimulus durations of 200 and 16 ms, respectively. The solid line in panel (e) shows the predictions of a pattern recognition model for the 16-ms duration, based on the observed pitch shifts for the 200-ms duration. See text for details.

cant interaction between the number of the mistuned harmonic and the mode of presentation [$F(6,18)=5.95$, $p=0.001$]. *Post hoc* contrasts based on Fisher's least significant difference procedure showed that pitch shifts were significantly larger for the monaural than for the dichotic condition when the second, third, or seventh harmonic was mistuned ($p<0.05$); for all other harmonics the difference was not significant.

Figure 5 shows the results for the four subjects who were tested in the dichotic condition using both durations. Panels (a)–(d) show the results of individual subjects, and panel (e) shows the mean results. The empty squares and the empty triangles show results for the 200- and 16-ms durations, respectively. The interaction between duration and harmonic number was of borderline significance [$F(6,18)=3.99$, $p=0.051$]. In contrast to the findings for experiment 1, there was a nonsignificant trend for the pitch shifts for the 16-ms duration to be *smaller* than those for the 200-ms duration [$F(1,3)=2.8$, $p=0.19$]. The direction of this trend was somewhat unexpected, and was opposite to the effect observed for monaural presentation, where the short duration led to significantly larger pitch shifts than the long duration.

Also indicated in Fig. 5(e), by the solid line, are the predictions of a pattern recognition model for the pitch shifts for the 16-ms duration. These predictions were based on the observed pitch shifts for the 200-ms duration in the dichotic condition. They were derived in the same way as in the monaural condition. In the dichotic condition, the constant scaling factor was 3.96, chosen such that the overall sum of the predicted pitch shifts for the dichotic 16-ms duration condition was equal to the overall sum of the observed pitch shifts for the dichotic 200-ms condition.

Qualitatively the predictions go in the right direction. For harmonics one and two, the predicted pitch shifts for the 16-ms duration were somewhat smaller than the observed pitch shifts for the 200-ms duration, as were the observed pitch shifts for the short duration (though this difference was not significant). However, the predicted pitch shifts were larger than the observed pitch shifts for harmonics one and two. For harmonics above the second, predicted pitch shifts for the 16-ms duration were generally somewhat larger than the observed pitch shifts for the 200-ms duration. Note that for each individual harmonic the effect of duration on the observed pitch shifts was nonsignificant (*t*-test).

Thus, as in the monaural case, the quantitative agree-

ment between predicted and observed pitch shifts for the 16-ms duration was moderate. In principle, the same possible reasons for this exist as mentioned for the monaural case. Qualitatively, the changes in the observed pitch shifts with decreased duration go in the predicted direction.

C. Discussion

For contralateral presentation of the mistuned harmonic, the effect of stimulus duration on the dominance region (interaction between harmonic number and duration) was only of borderline significance. The reason for this might be the unexpectedly large reduction in the size of the pitch shifts when the mistuned harmonic was presented contralaterally to the remainder of the complex rather than ipsilaterally; this effect of relative ear of entry was much larger for the 16-ms duration than for the 200-ms duration. As mentioned above, Darwin and Ciocca (1992) found a slight but significant reduction in the size of the pitch shift for contralateral presentation of the mistuned harmonic relative to monaural presentation. Their tones had a duration of 90 ms. We are not aware of any other study which investigated the effect of duration on the pitch shifts caused by mistuned harmonics for the case of contralateral presentation. A trend was observed for pitch shifts to be *smaller* for the short than for the long duration in the dichotic condition; the sum of the pitch shifts produced by each of the mistuned harmonics was 0.95% and 1.35% for the 16- and 200-ms durations, respectively. The expectation was that the *opposite* effect of duration on pitch shift would be seen, as usually it is assumed that individual components are perceptually more fused at short than at long durations (Moore *et al.*, 1986). Therefore, the sum of the pitch shifts would be expected to be larger for the short than for the long duration, as was observed with monaural presentation.

There are at least two possible explanations for this unexpected trend for the pitch shifts to be smaller for the short than the long duration when the mistuned component was presented contralaterally. First, it may be that the increase in perceptual fusion with decreasing stimulus duration is smaller for dichotic than for monaural presentation, i.e., grouping by common onset and offset might be less important for dichotic than for monaural presentation. However, we are not aware of any data suggesting this. In addition, because there was a trend for the pitch shifts to be smaller for the short than for the long duration in the dichotic condition, perceptual fusion would have to be *less* for the short than for the long duration in the dichotic condition. This seems rather unlikely.

The second possible explanation is that the pitch processor does not always assign nearly equal weights to inputs from the two ears. In the current experiment, subjects were instructed to match the pitch of the complex tone, not that of the separate tone. So, they had to pay attention to a specific ear, namely to the one to which the remainder of the target complex and the strictly harmonic comparison complex were presented. Possibly the weights depend on the extent to which attention is paid to one ear or the other, and—in this specific case—on whether there is enough time to switch attention between ears over the duration of the stimulus. For the 16-ms stimulus, the duration might have been too short for voluntary or involuntary shifts of attention to the ear receiving the single mistuned harmonic.

There are two reasons why attention to the ear being presented with the mistuned harmonic could increase pitch shifts. First, it could be that a contralateral component only affects the residue pitch when attention is paid to it. Second, it is theoretically possible that, for long(ish) stimulus durations, subjects can perform the task without using the residue pitch of the mistuned complex. Instead, the pitch matching might be performed in a rather more analytical way; subjects may compare the harmonic complex with the two distinct percepts of the mistuned component and the ''remainder'' of the mistuned complex. They could then try to minimize simultaneously the overall discrepancy between the harmonic complex and these two distinct percepts. The simultaneous minimization of *both* discrepancies is needed. If only the discrepancy between the mistuned harmonic and the corresponding harmonic in the comparison complex were minimized, then a harmonic mistuned by e.g. $\pm 3\%$ would lead to a pitch shift of $\pm 3\%$. If only the discrepancy between the remainder of the complex and the comparison complex were minimized, then a mistuned harmonic would not lead to any pitch shift. As the pitch shifts observed for a single mistuned harmonic are between those extremes, it follows that, if this analytical method is used, then both the mistuned harmonic and the remainder must be compared with the corresponding parts in the comparison complex.

The fact that a mistuned harmonic can still influence matches to the residue pitch even when it is presented before or after the remainder of the complex (Ciocca and Darwin, 1999) might indicate the use of such an analytical strategy. In the current study, where all partials were presented synchronously, pitch matches based on this strategy would only be possible when the duration is long enough to allow switching between the two percepts described above and to prevent strong perceptual fusion of the harmonics. Some of the subjects might have used such a strategy for the 200-ms duration in the dichotic condition, where the difference in ear of entry would enhance segregation between the mistuned harmonic and the remainder of the complex.

V. SUMMARY AND CONCLUSIONS

(1) For monaural stimulus presentation, the harmonic numbers that produced the largest shifts in residue pitch when they were mistuned by $\pm 3\%$ were higher for the 16-ms duration than for the 200-ms duration. This shift in the dominance region was predicted by pattern recognition models of pitch perception (Goldstein, 1973) and, in a simulation, was also found to be compatible with an autocorrelation model of pitch perception (Meddis and Hewitt, 1991a, b; Meddis and O'Mard, 1997). The implication is that the pitch of an inharmonic sound can depend on its duration.

(2) For monaural presentation, the sum of the pitch shifts from all harmonics was larger for the short than for the long duration, probably because the harmonics were perceptually more fused for the short than for the long duration.

(3) There was a marginally significant effect of stimulus duration on the dominance region when the mistuned harmonic was presented only contralaterally to the remainder of the complex. The results also showed a reduction in the overall pitch shifts for contralateral compared to ipsilateral presentation, which was highly significant for the short duration.
(4) For contralateral presentation of the mistuned harmonic, there was a nonsignificant trend for the sum of the pitch shifts produced by all harmonics to be larger for the long than for the short duration. This might indicate that the pitch processor does not always and automatically assign nearly equal weights to the inputs from the two ears. Rather the weights might depend on the distribution of attention to the two ears over the duration of the stimulus.

ACKNOWLEDGMENTS

This work was supported by EPSRC Grant No. GR/R65794/01. We thank Brian Moore for helpful discussions and comments on an earlier version of this paper. Thanks also to Chris Darwin, Christophe Micheyl, and Andrew Oxenham for useful comments.

[1]Throughout the paper, if appropriate, the Huynh–Feldt correction was applied to the degrees of freedom (Howell, 1997). In such cases, the corrected significance value and the original degrees of freedom are reported.
[2]If the assumption that the calculation of the ACF should be restricted to the time where the stimulus was present is dropped, then larger time constants can be used. The predictions of such an approach, with τ equal to 10 ms, the offset fixed at the end of the stimulus, and zero-padding before the onset of the 16 ms stimulus, showed qualitative agreement with the data.

Bilsen, F. A., and Ritsma, R. J. (**1967**). ''Repetition pitch mediated by temporal fine structure at dominant spectral regions,'' Acustica **19**, 114–116.
Ciocca, V., and Darwin, C. J. (**1999**). ''The integration of nonsimultaneous frequency components into a single virtual pitch,'' J. Acoust. Soc. Am. **105**, 2421–2430.
Dai, H. (**2000**). ''On the relative influence of individual harmonics on pitch judgment,'' J. Acoust. Soc. Am. **107**, 953–959.
Darwin, C. J. (**1992**). ''Listening to two things at once,'' in *The Auditory Processing of Speech: From Sounds to Words*, edited by M. E. H. Shouten (Mouton de Gruyter, Berlin).
Darwin, C. J., and Ciocca, V. (**1992**). ''Grouping in pitch perception: Effects of onset asynchrony and ear of presentation of a mistuned component,'' J. Acoust. Soc. Am. **91**, 3381–3390.
Darwin, C. J., Buffa, A., Williams, D., and Ciocca, V. (**1992**). ''Pitch of dichotic complex tones with a mistuned frequency component,'' in *Auditory Physiology and Perception*, edited by Y. Cazals, L. Demany, and K. Horner (Pergamon, Oxford).
Freyman, R. L., and Nelson, D. A. (**1986**). ''Frequency discrimination as a function of tonal duration and excitation-pattern slopes in normal and hearing-impaired listeners,'' J. Acoust. Soc. Am. **79**, 1034–1044.
Glasberg, B. R., and Moore, B. C. J. (**1990**). ''Derivation of auditory filter shapes from notched-noise data,'' Hear. Res. **47**, 103–138.
Goldstein, J. L. (**1973**). ''An optimum processor theory for the central formation of the pitch of complex tones,'' J. Acoust. Soc. Am. **54**, 1496–1516.
Goldstein, J. L., and Srulovicz, P. (**1977**). ''Auditory-nerve spike intervals as an adequate basis for aural frequency measurement,'' in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London).
Henning, G. B. (**1970**). ''Comparison of the effects of signal duration on frequency and amplitude discrimination,'' in *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G. F. Smoorenburg (Sijthoff, Leiden).
Howell, D. C. (**1997**). *Statistical Methods for Psychology* (Duxbury, Belmont, CA).
Liang, C.-A., and Chistovich, L. A. (**1961**). ''Frequency difference limens as a function of tonal duration,'' Sov. Phys. Acoust. **6**, 75–80.
Meddis, R. (**1986**). ''Simulation of mechanical to neural tranduction in the auditory receptor,'' J. Acoust. Soc. Am. **79**, 702–711.
Meddis, R. (**1988**). ''Simulation of auditory-neural tranduction: Further studies,'' J. Acoust. Soc. Am. **83**, 1056–1063.
Meddis, R., and Hewitt, M. (**1991a**). ''Virtual pitch and phase sensitivity studied using a computer model of the auditory periphery. I. Pitch identification,'' J. Acoust. Soc. Am. **89**, 2866–2882.
Meddis, R., and Hewitt, M. (**1991b**). ''Virtual pitch and phase sensitivity studied using a computer model of the auditory periphery. II. Phase sensitivity,'' J. Acoust. Soc. Am. **89**, 2883–2894.
Meddis, R., and O'Mard, L. (**1997**). ''A unitary model of pitch perception,'' J. Acoust. Soc. Am. **102**, 1811–1820.
Meddis, R., Hewitt, M. J., and Shackleton, T. M. (**1990**). ''Implementation details of a computational model of the inner hair-cell/auditory-nerve synapse,'' J. Acoust. Soc. Am. **87**, 1813–1816.
Micheyl, C., Moore, B. C. J., and Carlyon, R. P. (**1998**). ''The role of excitation-pattern cues and temporal cues in the frequency and modulation-rate discrimination of amplitude-modulated tones,'' J. Acoust. Soc. Am. **104**, 1039–1050.
Moore, B. C. J. (**1973**). ''Frequency difference limens for short-duration tones,'' J. Acoust. Soc. Am. **54**, 610–619.
Moore, B. C. J. (**1987**). ''The perception of inharmonic complex tones,'' in *Auditory Processing of Complex Sounds*, edited by W. A. Yost and C. S. Watson (Erlbaum, Hillsdale, NJ).
Moore, B. C. J. (**2003**). *An Introduction to the Psychology of Hearing, 5th Ed.* (Academic, San Diego).
Moore, B. C. J., Glasberg, B. R., and Peters, R. W. (**1985**). ''Relative dominance of individual partials in determining the pitch of complex tones,'' J. Acoust. Soc. Am. **77**, 1853–1860.
Moore, B. C. J., Glasberg, B. R., and Peters, R. W. (**1986**). ''Thresholds for hearing mistuned partials as separate tones in harmonic complexes,'' J. Acoust. Soc. Am. **80**, 479–483.
Moore, B. C. J., Glasberg, B. R., and Shailer, M. J. (**1984**). ''Frequency and intensity difference limens for harmonics within complex tones,'' J. Acoust. Soc. Am. **75**, 550–561.
Patterson, R. D., and Wightman, F. L. (**1976**). ''Residue pitch as a function of component spacing,'' J. Acoust. Soc. Am. **59**, 1450–1459.
Plomp, R. (**1967**). ''Pitch of complex tones,'' J. Acoust. Soc. Am. **41**, 1526–1533.
Ritsma, R. J. (**1967**). ''Frequencies dominant in the perception of the pitch of complex sounds,'' J. Acoust. Soc. Am. **42**, 191–198.
Ritsma, R. J. (**1970**). ''Periodicity detection,'' in *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G. F. Smoorenburg (Sijthoff, Leiden).
Sekey, A. (**1963**). ''Short-term auditory frequency discrimination,'' J. Acoust. Soc. Am. **35**, 682–690.
Terhardt, E. (**1972**). ''Zur Tonhöhenwahrnehmung von Klängen II. Ein Funktionsschema (Pitch of complex sounds II. A functional scheme),'' Acustica **26**, 187–199.
Terhardt, E. (**1974**). ''Pitch, consonance, and harmony,'' J. Acoust. Soc. Am. **55**, 1061–1069.
Terhardt, E. (**1979**). ''Calculating virtual pitch,'' Hear. Res. **1**, 155–182.
Terhardt, E., Stoll, G., and Seewann, M. (**1982a**). ''Algorithm for extraction of pitch salience from complex tonal signals,'' J. Acoust. Soc. Am. **71**, 679–688.
Terhardt, E., Stoll, G., and Seewann, M. (**1982b**). ''Pitch of complex signals according to virtual pitch theory,'' J. Acoust. Soc. Am. **71**, 671–678.
Wightman, F. L. (**1973**). ''The pattern-transformation model of pitch,'' J. Acoust. Soc. Am. **54**, 407–416.

Frequency dependency of the relationship between perceived auditory source width and the interaural cross-correlation coefficient for time-invariant stimuli

Russell Mason,[a] Tim Brookes, and Francis Rumsey
Institute of Sound Recording, University of Surrey, Guildford, Surrey, GU2 7XH, England

(Received 27 November 2003; revised 2 December 2004; accepted 6 December 2004)

Previous research has indicated that the relationship between the interaural cross-correlation coefficient (IACC) of a narrow-band sound and its perceived auditory source width is dependent on its frequency. However, this dependency has not been investigated in sufficient detail for researchers to be able to properly model it in order to produce a perceptually relevant IACC-based model of auditory source width. A series of experiments has therefore been conducted to investigate this frequency dependency in a controlled manner, and to derive an appropriate model. Three main factors were discovered in the course of these experiments. First, the nature of the frequency dependency of the perceived auditory source width of stimuli with an IACC of 1 was determined, and an appropriate mathematical model was derived. Second, the loss of perceived temporal detail at high frequencies, caused by the breakdown of phase locking in the ear, was found to be relevant, and the model was modified accordingly using rectification and a low-pass filter. Finally, it was found that there was a further frequency dependency at low frequencies, and a method for modeling this was derived. The final model was shown to predict the experimental data well. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1853113]

PACS numbers: 43.66.Pn, 43.66.Rq, 43.66.Ba, 43.55.Mc [AK] Pages: 1337–1350

I. INTRODUCTION

It has been shown that the results of measurements based on the interaural cross-correlation coefficient (IACC) can be used to predict a number of aspects of binaural perception. Whereas recent research has suggested that it is not an accurate representation of the physiology of the binaural hearing process (Brand *et al.*, 2002), the predictions of models of binaural perception based on the IACC have shown remarkable similarity to experimental data relating to lateralization (e.g., Sayers, 1964; Okano, 2000; Constan and Hartmann, 2001), binaural detection thresholds (e.g., Sayers and Cherry, 1957; Robinson and Jeffress, 1963; Bernstein and Trahiotis, 1992), and the perceived source width of an auditory stimulus (e.g., Chernyak and Dubrovsky, 1968; Plenge, 1972; Kurozumi and Ohgushi, 1983; Blauert and Lindemann, 1986).

The relationship between the IACC and binaural detection thresholds has been researched in detail, and the majority of relevant fundamental factors such as frequency dependency (e.g., Bernstein and Trahiotis, 1996b), and temporal resolution (e.g., Culling and Summerfield, 1998), have been investigated in a controlled manner. The relationship between the IACC and perceived auditory source width, however, has not been investigated in as much detail. The aim of this paper is to investigate the frequency dependency of the relationship between the IACC of an auditory stimulus and the perceived auditory source width of that stimulus, and to model this dependency such that it can be incorporated in an auditory source width measure.

In the late 1960s, Chernyak and Dubrovsky (1968) showed that there was a negative relationship between the IACC and the perceived auditory source width of noise stimuli that were presented over headphones. This relationship was also found by Keet (1968), who used musical signals that were reproduced over loudspeakers. This inverse relationship between the IACC and the perceived auditory source width has been confirmed by other investigations using a range of stimuli (Kurozumi and Ohgushi, 1983; Morimoto *et al.*, 1993). From this, a number of researchers have developed models, based on the IACC, to predict the perceived source width of auditory stimuli for sound reproduction (e.g., Sakai *et al.*, 1980; Tohyama and Suzuki, 1989) and auditorium acoustics (e.g., Ando, 1985; Morimoto and Iida, 1995; Hidaka *et al.*, 1995). This research has resulted in models that match perception reasonably well, and has led to the standardization of some methods (ISO 3382, 1997). However, in spite of the growing popularity of these models, there are arguments against their practicality (e.g., Griesinger, 1997; de Vries *et al.*, 2001), and there are also a number of fundamental factors that have not yet been fully addressed in a controlled manner.

One such factor is the effect of audio frequency on the relationship between the IACC and the perceived source width of an auditory stimulus. Investigations by a number of researchers have indicated the existence of this effect (e.g., Perrott *et al.*, 1980; Blauert and Lindemann, 1986; Yanagawa *et al.*, 1990), but methods to model it have not yet been derived.

The experiments that are reported in this paper investigate the frequency dependency of the relationship between the IACC and perceived auditory source width in detail, with the aim to identify and model the separate contributory components, in order to develop an objective auditory source

[a]Electronic mail: r.mason@surrey.ac.uk

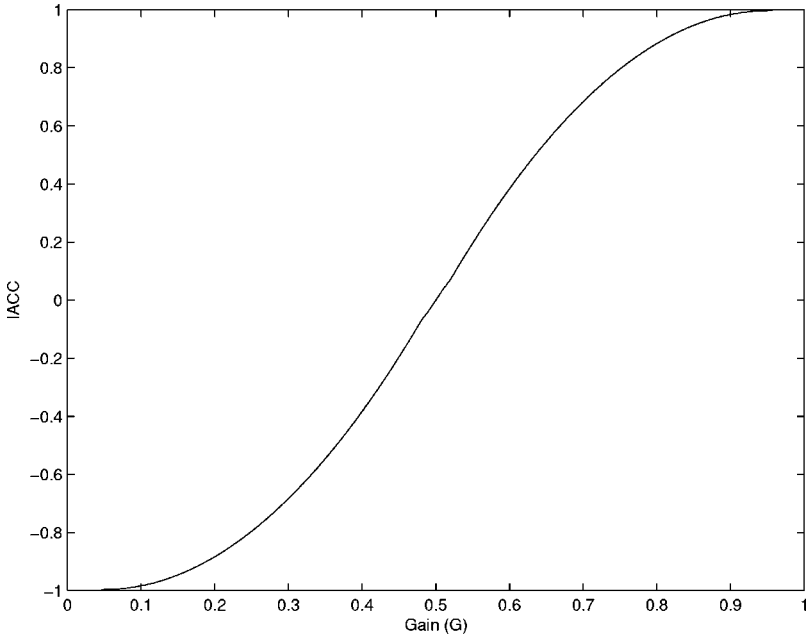


FIG. 1. Relationship between the gain value, *G*, and the resulting IACC for stimuli created using the method described in Eqs. (1)–(4).

width measure that is suitable for a wide range of stimulus types. Additionally, it is expected that the research reported in this paper will lead to the inclusion of an IACC-based auditory source width model in a more complete binaural hearing model, similar to that proposed by Morimoto *et al.* (1995). The experiments reported herein all employ continuous narrow-band stimuli, and it is acknowledged that further work may be required to adjust the resulting model for use with stimuli that are more externally valid. Detailed information on the IACC-based calculations used in this paper is contained in the Appendix.

II. METHOD

A. Stimuli

In order to undertake the experiment in a controlled manner and limit extraneous variables as much as possible, stimuli with specific characteristics were required. A detailed discussion of the requirements of the stimuli for the experiments that are described in this paper is provided by Mason *et al.* (2003); the pertinent information is summarized here. First, testing of frequency dependence requires stimuli with a narrow bandwidth—ideally within a single critical band, i.e., approximately 35 Hz at low frequencies and widening with increasing center frequency (Glasberg and Moore, 1990). Second, accurate mapping of a perceived effect to a given measured value requires stimuli that are time invariant in terms of that metric. As any measurement of time invariance is dependent on the temporal resolution of that measurement, the optimum integration window needs to be determined. Previous studies into the perception of temporal variations in IACC have suggested that this can be modeled by a time constant of between 35 and 243 ms, depending on the experiment setup and the individual subject (Grantham and Wightman, 1979; Boehnke *et al.*, 2002). For these experiments the shortest integration time (therefore the most critical value resulting from the most sensitive subjects) from previous studies was used, namely 35 ms. Finally, to be able to examine the effect of frequency dependency at a wide range of values of IACC requires stimuli whose IACC can be varied in small steps across the entire range of values while minimizing any other perceivable variations.

Previous experiments that investigated relationships between the IACC and perceived auditory source width used a wide range of stimuli, including music (e.g., Keet, 1968; Morimoto and Iida, 1995) and noise (e.g., Chernyak and Dubrovsky, 1968; Blauert and Lindemann, 1986). The main problem with these types of stimuli is that the IACC can vary a great deal over time, as shown by Mason *et al.* (2003), which makes it difficult to specify the actual IACC value to which a given perceptual effect relates, be it the maximum, minimum, or some type of average.

The stimuli chosen for use in the experiments described in this paper are based on suppressed carrier amplitude modulation with a time offset between the channels. Two signals, s_1 and s_2, are created as shown in Eqs. (1) and (2); where f_c is the carrier frequency, f_m is the modulation frequency, and t_m is the time offset defined as $[1/(4f_m)]$

$$s_1 = \sin(2\pi f_c t)\sin(2\pi f_m t), \tag{1}$$

$$s_2 = \sin(2\pi f_c[t+t_m])\sin(2\pi f_m[t+t_m]). \tag{2}$$

A two-channel output signal, *l* and *r*, is then created by mixing signals s_1 and s_2; one in phase and one phase inverted between the channels, as shown in Eqs. (3) and (4). Assuming that the stimuli are reproduced to the subjects over headphones, the IACC of the output signal can be altered by varying the gain value *G*. A range of values of *G* of 0 to 1 results in a range of values of IACC of -1 to 1, and it is possible to set the IACC to be any value within this range:

$$l = Gs_1 + (1-G)s_2, \tag{3}$$

$$r = Gs_1 - (1-G)s_2. \tag{4}$$

The relationship between the gain *G* and the resulting IACC is independent of f_c for the range used, but is not linear, as is shown in Fig. 1. This means that a linear variation in *G*

results in a small variation in the resulting IACC close to 1 and −1 where the ear is more sensitive to small changes, and a larger variation in the resulting IACC close to 0 where the ear is less sensitive to small changes (Gabriel and Colburn, 1981; Boehnke *et al.*, 2002).

The modulation frequency—f_m in Eqs. (1) and (2)—of this type of stimulus affects the bandwidth and the time invariance of the IACC in a contrary manner, which requires a compromise. For the experiments described in this paper, the value of f_m was set to be 20 Hz, which resulted in a bandwidth of 40 Hz (slightly wider than a critical band at low center frequencies) and a constant IACC over time for an integration window of 50 ms or longer (resulting in an IACC that is perceptually time invariant for the majority of listeners, excluding the most sensitive).

The center frequency of a stimulus is set by altering the carrier frequency, f_c in Eqs. (1) and (2). It was confirmed that the IACC of the stimuli could be adjusted by altering the value of G in Eqs. (3) and (4), and that this caused only minimal changes in the perceived timbre of the stimuli. The interested reader is referred to Mason *et al.* (2003) for more information.

The stimuli were generated using computer software, which allowed variation of G in real time in 128 equal steps between 0.5 and 1 (giving a range of IACC of 0 to 1 as shown in Fig. 1). The output signal was fed through a 16-bit DAC to a pair of Sennheiser HD-580 headphones.

The stimuli were reproduced at a sound pressure level (measured individually at each ear) corresponding to 70 phons, as derived from the equal loudness contours contained in ISO 226 (1987). Whereas the data contained in ISO 226 are intended for pure tone stimuli as opposed to these two-tone experiment stimuli, informal listening by the authors confirmed that the stimuli were all of similar loudness after alignment.

B. Procedure

The experiments were conducted using a ''method of adjustment'' paradigm (Gescheider, 1997), where the subjects were asked to adjust a variable stimulus to be the same perceived auditory source width as a given fixed stimulus. The fixed stimuli were presented individually, and were continuously looped so that the subjects could take as long as required to adjust the auditory source width of the variable stimulus. The subjects could switch between the variable and fixed stimuli whenever and as often as they required, and could move on to the next fixed stimulus when they were satisfied with their result. The switching between the stimuli was instantaneous, with a 20-ms cross-fade to avoid audible clicks. The user interface was configured so that the subjects could alter the IACC of the variable stimulus either by moving an on-screen slider with a mouse or by using a hardware slider. The subjects could switch between the stimuli using either the mouse or hardware buttons. The order of presentation of the stimuli was randomized by the computer to avoid biases caused by order effects; this also rendered the test double blind (i.e., neither the subjects nor the experimenter knew which fixed stimulus was which).

The benefits of the method of adjustment procedure over other experimental methods are that it is relatively efficient in terms of the amount of time required to undertake the experiment, and the intrasubject repeatability is high (Hesse, 1986). In addition, as the subjects have an active role in the experiment, it is less likely that their performance will suffer due to boredom or fatigue (Cardozo, 1965). However, there are two main drawbacks to the method of adjustment procedure. First, it is possible that in an experiment with a large number of repetitions that require similar adjustments to be made in each case, a bias will be caused by habituation and expectation (Gescheider, 1997). Second, the method lacks the robustness of other procedures, i.e., it is dependent on the judgment of the subjects that a match is ''good enough'' (Wier *et al.*, 1976), as opposed to a strict statistical method such as that used in a two-alternative forced choice (2AFC) experiment (Levitt, 1971). It is apparent that the effects of these drawbacks can be reduced by the use of carefully selected subjects and, in view of this, it was decided that the advantages of the procedure outweighed the disadvantages for these experiments.

C. Subjects

As the method of adjustment procedure is strongly dependent on the ability of the subject, a pilot experiment was used to select a small panel of suitable listeners. This used the same procedure and a subset of the stimuli from the main experiments, and also served to train the subjects in the experiment task and to enable them to become familiar with the types of stimuli that were used throughout. The subjects were selected based on their consistency (between judgments for multiple iterations of the same stimuli) and accuracy (of matching the IACC of a fixed stimulus that had the same center frequency as the variable stimulus). Seven subjects were selected from the 25 who undertook the pilot experiment. The selected subjects included one female and six males, with an age range of 21 to 35, one of which was the first author. They were either final year undergraduates, postgraduates, or staff from the Institute of Sound Recording, and could therefore be considered to be experienced critical listeners. The subjects were instructed that there were no correct answers in the experiment, and that they would only be rated on the consistency of their judgments.

III. EXPERIMENT 1—EFFECT OF FREQUENCY ON PERCEIVED AUDITORY SOURCE WIDTH WHEN IACC=1

A. Introduction

The research of Perrott *et al.* (1980) showed that the perceived auditory source width (which they termed ''tonal volume'') of an auditory stimulus was dependent on the center frequency, even for a simple diotic stimulus (i.e., a signal that is identical at each ear and therefore has an IACC of 1), such as a single sine tone. As it is known that the IACC has an inverse relationship with the perceived auditory source width, it can be assumed that stimuli with an IACC of 1 will represent the narrowest condition at a particular frequency. In view of this, the first experiment examined the effect of the frequency of a stimulus on its perceived auditory source

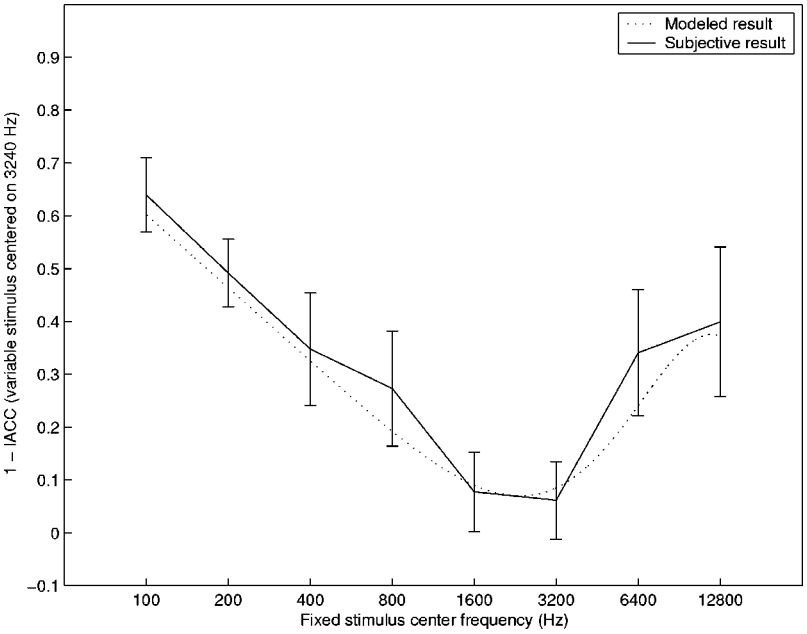


FIG. 2. Means and associated 95% confidence intervals of the 1-IACC of the variable stimulus centered on 3240 Hz that the subjects judged to have the same auditory source width as each fixed stimulus in experiment 1 (subjective result), together with the results predicted by Eq. (5) (modeled result).

width for an IACC of 1, in order to determine a baseline from which any changes in the perceived auditory source width caused by variations in the IACC can be calculated.

B. Stimuli

Stimuli based on amplitude modulation, as described above, were used in the experiment. The stimuli were created using Eqs. (1)–(4), with $f_m = 20$ Hz and $G = 1.0$. In order to examine the effect of the frequency of the stimulus on the perceived auditory source width, fixed stimuli with a range of values of f_c were used with octave spacing: 100, 200, 400, 800, 1600, 3200, 6400, and 12 800 Hz.

The variable stimulus that was used in the experiment was created using $f_m = 20$ Hz and $f_c = 3240$ Hz. The value of f_m was chosen to match the fixed stimuli, and the value of f_c was chosen (from informal listening by the experimenters) to give a range of variation of auditory source width for the variable stimulus that encompassed the range of widths present in the fixed stimuli.

C. Procedure

There were eight fixed stimuli in the experiment; for each of these the subjects were asked to adjust the variable stimulus until they judged it to have the same auditory source width as the fixed stimulus. The experiment comprised four repetitions, each presented in a different random order. All seven selected subjects undertook the full experiment.

D. Results

The results of the subjective experiment were analyzed by calculating the IACC of the variable stimulus that the subjects judged to have the same perceived auditory source width as each fixed stimulus. The means and associated 95% confidence intervals of these results are shown as the subjective result in Fig. 2, with the center frequency of each fixed stimulus along the x axis and the IACC value subtracted from 1 (1-IACC) of the matched variable stimulus on the y axis. The IACC scale is inverted in this manner to give a positive correlation with the perceived auditory source width, as discussed earlier. The wide variation in 1-IACC, resulting from changes in the center frequency of the fixed stimulus, equates to a wide variation in perceived auditory source width, despite the fact that the IACC was set to be 1 for all the fixed stimuli. It may be interpreted from these results that the low-frequency fixed stimuli were perceived to be widest, with a monotonic decrease in the auditory source width to approximately 1600 Hz. Above 3200 Hz the perceived auditory source width appears to rise again.

The results from this experiment can be modeled by a frequency-dependent equation, as shown in Eq. (5), where b predicts the 1-IACC of the variable stimulus whose auditory source width would be judged to match that of a fixed stimulus with center frequency f. This is shown in Fig. 2 as the modeled result:

$$b = \frac{\sin[(5\sqrt{f} + 4000)/100] + 1}{f^{0.2}} + 0.07, \quad 100 \leqslant f \leqslant 12800. \tag{5}$$

It is reasonable to expect that the perceived widths of two similar stimuli having similar center frequencies and the same IACC and loudness will be approximately equal. Therefore, when the center frequency of the fixed stimulus is close to that of the variable stimulus, it is reasonable to expect that the subjects will adjust the IACC of the variable stimulus to be similar to the IACC of the fixed stimulus. Based on this expectation, the 1-IACC of the variable stimulus adjusted to match the width of the fixed stimulus with a center frequency of 3200 Hz should be approximately 0, but it can be seen from Fig. 2 that in fact it has a mean value of 0.043. It is likely that this discrepancy was caused by a floor effect in the experiment due to the finite range of variation available with the variable stimulus. This may mean that the stimuli that were judged to be narrowest (those centered on

1600 and 3200 Hz) were actually narrower than the subjective results shown in Fig. 2 suggest and may also be the cause of the offset of 0.07 in the prediction model shown in Eq. (5). However, the likely error is small compared to the range of results for the other experimental stimuli, and therefore has not been corrected in the model.

It must be noted that the shape of the subjective result curve in Fig. 2 and the modeled result in Eq. (5) inevitably are partly influenced by the center frequencies of the stimuli that were used in the experiment, and that the use of a greater number of fixed stimuli with center frequencies between those used would allow for a more accurate model. Also, Eq. (5) has been derived to match the subjective results to IACC measurement results, and it may be necessary to revise this once further adaptations are made to the IACC calculation in later sections of the paper.

E. Discussion

The physiological explanation for these subjective results is as yet unknown; however, a mechanism for incorporating the results of this experiment into a model of auditory source width perception needs to be derived. Since the stimuli used in this experiment all have a 1-IACC of 0, simply adding the baseline function shown in Eq. (5) to the measured 1-IACC should provide a realistic measure of perceived auditory source width for these stimuli. The validity of this approach when applied to the measured results of stimuli with a range of IACC values is tested implicitly in the experiments described below.

IV. EXPERIMENT 2—EFFECT OF BREAKDOWN OF PHASE LOCKING ON PERCEPTION OF IACC

A. Introduction

As mentioned in the Introduction, it is expected that the results of the research that are reported in this paper will be implemented as part of a more complete model of binaural hearing. One feature that is commonly found in such models is a simulation of the breakdown of phase locking in the neural output of the ear above a certain frequency (Stern and Trahiotis, 1997).

Detailed models have been developed to mimic the limited firing rate of the neural synapses (Meddis *et al.*, 1990), though this effect is commonly approximated by the use of half-wave rectification followed by a low-pass filter (e.g., Karjalainen, 1996). Bernstein and Trahiotis (1996b) investigated the effect of a half-wave rectification and low-pass filtering preprocess on IACC-based calculations used to predict the binaural detection performance of a group of subjects for a range of stimuli. They found that the addition of this preprocess led to predictions which matched their empirical results quite accurately. As binaural detection and perception of auditory source width appear to be similar in their association with the IACC, it is likely that this type of simulation of the breakdown of phase locking will be applicable to prediction of the perceived auditory source width of stimuli.

However, the majority of auditory source width models that are based on the IACC do not employ such a simulation. In view of this, it is of interest to investigate whether phase locking has any effect on auditory source width perception and, if so, whether simulation of this using rectification and filtering is an appropriate addition to an auditory source width model. If this is the case, then the optimum filter characteristics must also be determined.

B. Stimuli

The amplitude-modulation-based stimuli that were used for the previous experiment have an interesting property that can be employed to investigate the effect of the breakdown of phase locking on the perceived auditory source width of a stimulus. If a stimulus with an IACC of 0 is created [therefore $G=0.5$ in Eqs. (3) and (4)] when $f_c=(2n-1)f_m$ where n is an integer, then each output channel contains only one tonal component, the frequency of which is f_c+f_m in one channel and f_c-f_m in the other. This means that there is no in-channel amplitude modulation, and the decorrelation of the two channels is due to the constantly varying phase difference between them caused by the different frequency tones in each channel. Conversely, when $G=0.5$ and $f_c=2nf_m$, the two tonal components are combined in each of the output channels, which causes in-channel amplitude modulation in each of them. In this case the decorrelation is caused by the fluctuations in the interchannel level difference over time as well as a change in the interchannel phase difference every quarter of the modulation period. This can be seen in the plots of the examples of each stimulus type shown in Fig. 3. For ease of reference, stimuli with $f_c=(2n-1)f_m$ will be referred to as type 1, and stimuli with $f_c=2nf_m$ will be referred to as type 2.

For both of these types of stimuli, the IACC is 0. However, the IACC of the amplitude envelopes of the two signals is very different, where type 1 stimuli have an envelope IACC of 1, and type 2 stimuli have an envelope IACC that is much lower (the actual value of the IACC will be dependent on the method used to extract the envelope of the signal). This is shown in detail by Mason *et al.* (2003). While the envelope differences between type 1 and type 2 signals may not be perceived at lower frequencies, it is possible that they will be perceived at higher frequencies where the breakdown of phase locking may cause the human perceptual system to fail to track the fine detail of the stimuli.

For this experiment, the stimuli were created using the same suppressed carrier amplitude modulation technique that was used in the previous experiment, though in this case all the fixed stimuli used $G=0.5$ which resulted in an IACC of 0. The center frequencies (f_c) of the fixed stimuli were chosen to encompass the range of frequencies around which previous research indicated the breakdown of phase locking may start to take effect. Ten center frequencies (values of f_c) were used for each stimulus type—type 1: 180, 300, 420, 580, 740, 940, 1180, 1460, 1820, and 2260 Hz; type 2: 200, 320, 440, 600, 760, 960, 1200, 1480, 1840, and 2280 Hz. This gave 20 fixed stimuli for the experiment.

In order to compare like with like, two variable stimuli were used, one of each type. The frequencies of the variable stimuli were chosen to be well below the frequency at which the breakdown of phase locking may start to have an effect,

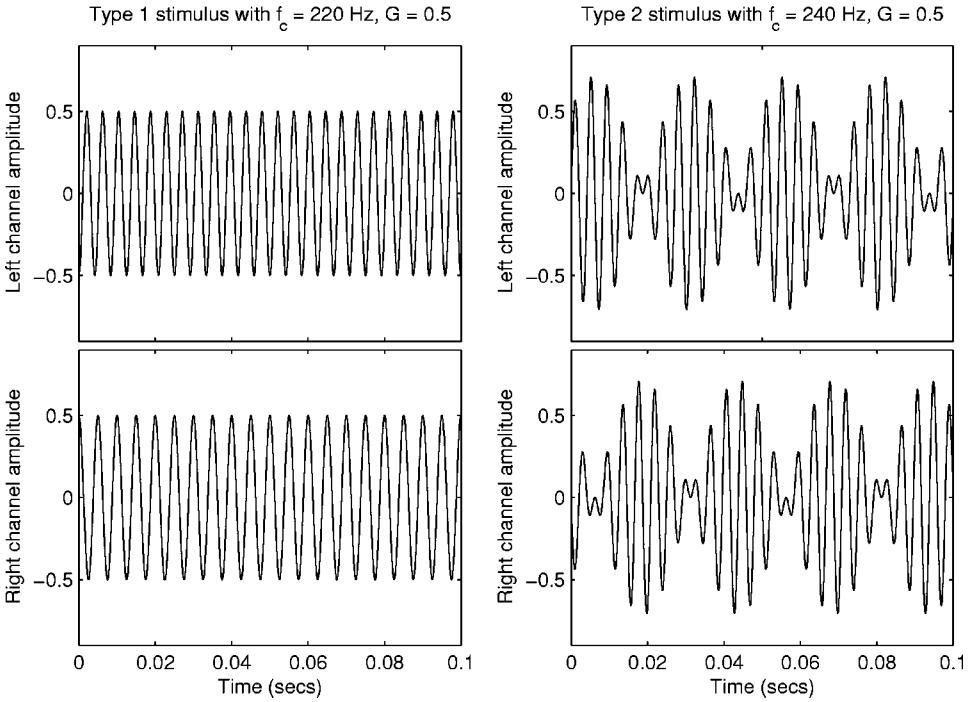


FIG. 3. An example signal of type 1 with a center frequency of 220 Hz and a gain (G) of 0.5 in the left plot, and an example signal of type 2 with a center frequency of 240 Hz and a gain (G) of 0.5 in the right plot.

and to be able to cover as much of the range of auditory source widths of the fixed stimuli as possible, as judged by the authors in informal listening. The type 1 variable stimulus had a value of f_c of 220 Hz and the type 2 variable stimulus had a value of f_c of 240 Hz. As for the previous experiment, the value of f_m was fixed for all the stimuli at 20 Hz.

C. Procedure

There were 20 fixed stimuli in the experiment; for each of these the subjects were asked to adjust the corresponding variable stimulus until they judged it to have the same auditory source width as the fixed stimulus. The experiment comprised two repetitions, each presented in a different random order. All seven selected subjects undertook the full experiment.

D. Results

The results of this subjective experiment were analyzed by calculating the IACC of the variable stimulus that the subjects judged to have the same perceived auditory source width as each fixed stimulus. The means and associated 95% confidence intervals of these results are shown in Fig. 4, with the center frequency of each fixed stimulus along the x axis and 1-IACC of the matched variable stimulus on the y axis, with the left plot showing the results for the type 1 stimuli and the right plot showing the results for the type 2 stimuli. In both plots, an increase on the y axis can be interpreted as an increase in the perceived auditory source width.

As for the previous experiment, it is reasonable to expect that the subjects would adjust the IACC of the variable stimulus to be similar to the IACC of the fixed stimulus with a similar center frequency, i.e., in this case that the mean 1-IACC of the variable stimuli adjusted to match the width of the fixed stimuli with center frequencies of 180 and 200 Hz would be approximately 1. It can be seen in Fig. 4 that the variable stimuli matched to the width of the fixed stimuli centered on 180 and 200 Hz have IACCs of 0.89 and 0.8, respectively. It is likely that the discrepancy between the expected and actual results is partially due to a ceiling effect in the experiment due to the finite range of variation available with the variable stimulus. The fact that there appears to be a greater difference between the expected and actual results for the type 2 stimuli than for the type 1 stimuli is explained later.

If the breakdown of phase locking in the ear does need to be taken into account in an IACC-based model of auditory source width, then this will be evident as differences between the results at similar frequencies for the two types of stimuli, especially at higher frequencies. It can be seen in Fig. 4 that the results for the type 1 stimuli above approximately 1 kHz indicate that these were perceived to be significantly narrower than the type 2 stimuli with similar center frequencies. This indicates that the breakdown of phase locking does indeed affect the relationship between the IACC and perceived auditory source width, and that this should therefore be taken into account in a model.

The most suitable type of low-pass filter was then identified as follows. Each fixed stimulus used in the experiment was passed through a trial filter and its IACC was calculated to give the modeled result. The variable stimulus that each subject matched to it was analyzed similarly, and the means and associated 95% confidence intervals were calculated to give the subjective result. The results from the two sets of filtered IACC measurements were compared, with an appropriate offset to compensate for any outstanding frequency-dependent differences between the variable and fixed stimuli. This procedure was repeated using a number of different trial filter types, including that proposed by Bernstein and Trahiotis (1996b). The best fit between the subjective and modeled results was found with the use of half-wave rectification

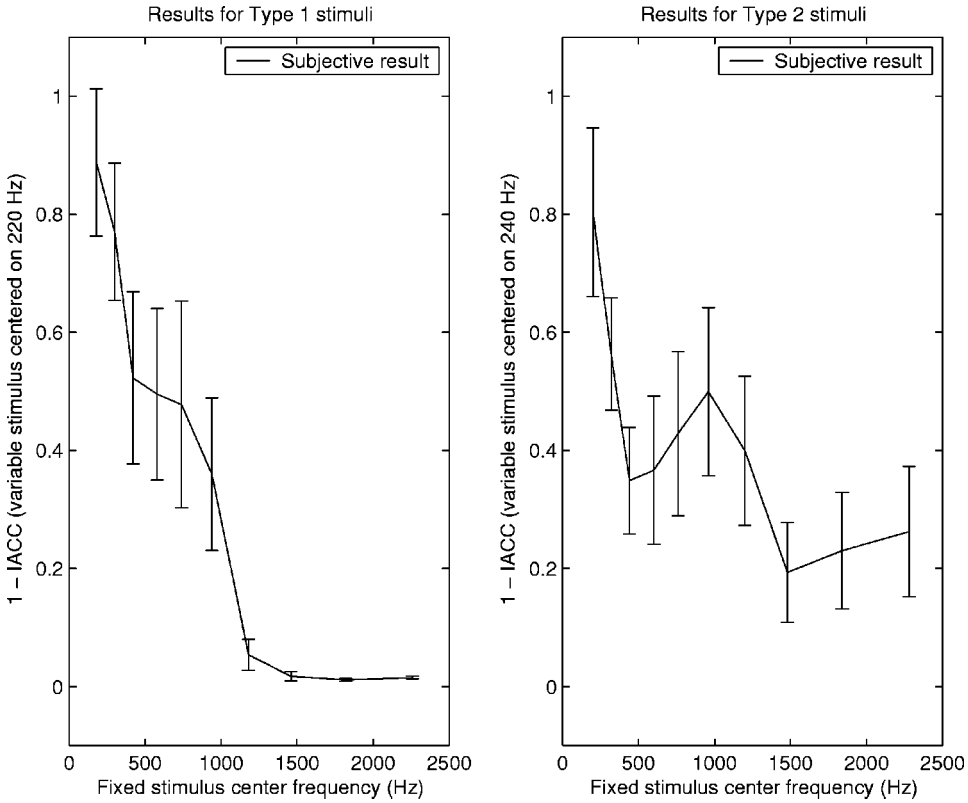


FIG. 4. Means and associated 95% confidence intervals of the 1-IACC of the variable stimulus that the subjects judged to have the same auditory source width as each fixed stimulus in experiment 2, with type 1 stimuli (variable stimulus centered on 220 Hz) on the left and type 2 stimuli (variable stimulus centered on 240 Hz) on the right.

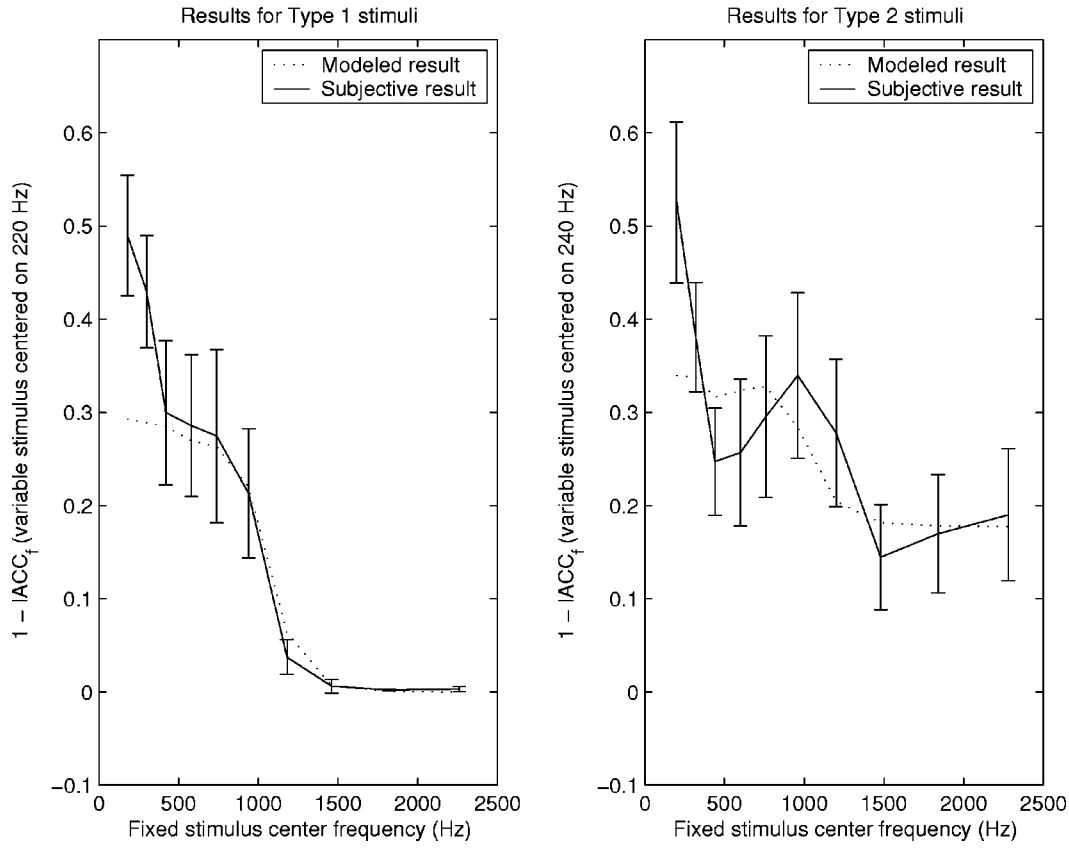


FIG. 5. Means and associated 95% confidence intervals of the 1-IACC$_f$ of the variable stimulus that the subjects judged to have the same auditory source width as each fixed stimulus in experiment 2 (subjective result), together with the 1-IACC$_f$ of each fixed stimulus with the addition of an offset (modeled result), with type 1 stimuli (variable stimulus centered on 220 Hz) on the left and type 2 stimuli (variable stimulus centered on 240 Hz) on the right.

followed by a sixth-order Butterworth low-pass filter with a cutoff frequency of 1 kHz. The relevant comparison is shown in Fig. 5 (the term 1-$IACC_f$ refers to a calculation made after half-wave rectification and filtering as described in the Appendix).

It can be seen in Fig. 5 that in terms of the shape of the subjective results in the region of 1 kHz there is a difference between the two stimulus types. The authors currently cannot offer an explanation for this, and the optimum filter type was chosen based on matching the range of the 95% confidence intervals for as many data points as possible. It is also apparent from Fig. 5 that the modeled results do not accurately match the subjective results at low frequencies; it was therefore necessary to ignore the results at low frequencies in order to obtain a good fit. This mismatch at low frequencies may be caused by a further frequency dependency in the relationship between the IACC and the perceived auditory source width. However, further investigation is required to verify this.

E. Discussion

It must be noted that half-wave rectification of a signal causes the possible range of IACC values to reduce. An audio signal that has an IACC of 0 before half-wave rectification will have an IACC of approximately 0.4 after half-wave rectification. In a similar manner, an audio signal that has an IACC of -1 before half-wave rectification will have an IACC of approximately 0 after half-wave rectification. This difference in the measured IACC values is partly caused by the fact that half-wave rectified signals have a nonzero mean (i.e., they contain a dc offset). The effect that this has on the measured result could be negated by subtracting the mean value of each of the input signals before calculation, resulting in a cross-covariance calculation as defined by Lynn and Fuerst (1998). However, Bernstein and Trahiotis (1996a) showed that the cross-correlation calculation (that they termed the normalized correlation) gave a better match to the perceived properties of high-frequency signals than the cross-covariance calculation (that they termed the normalized covariance). For the purposes of the model of perceived auditory source width that is under development in this paper, using the cross-covariance calculation would result in an overestimation of the dissimilarity of the high-frequency signals that have been half-wave rectified and low-pass filtered.

It is possible that the use of a more physiologically accurate model of the transduction process of the ear will give an improved match to the subjective data, compared to half-wave rectification and low-pass filtering. However, this will be at the expense of an increase in the required computation, and the improvement in the match to the subjective data is likely to be small.

In view of this reduction in the range of possible IACC values, the function that was derived at the end of the previous experiment needs to be modified. In order to update this previous model, the analysis of that experiment was repeated using the 1-$IACC_f$ calculation (including the half-wave rectification and low-pass filtering). This analysis resulted in a similar equation to that shown in Eq. (5). The result, now in terms of 1-$IACC_f$, is shown in Eq. (6), where b predicts the 1-$IACC_f$ of the variable stimulus whose auditory source width would be judged to match that of a fixed stimulus with center frequency f:

$$b = \left[\frac{\sin[(5.25\sqrt{f} + 4000)/100] + 0.75}{f^{0.2}} + 0.05 \right] / 3.8,$$

$$100 \leqslant f \leqslant 12800. \tag{6}$$

F. Comparison between type 1 and type 2 variable stimuli

In order to allow comparison between the results for the two types of stimuli, an additional experiment was conducted. Fixed type 2 stimuli were created with a center frequency (f_c) of 240 Hz and IACC values of 0.0, 0.2, 0.4, 0.6, 0.8, and 1.0. The subjects were then asked to match the perceived auditory source width of these stimuli using a type 1 variable stimulus with a center frequency (f_c) of 220 Hz. The same seven selected subjects undertook this experiment, and they were asked to adjust the variable stimulus until they judged it to have the same auditory source width as each fixed stimulus in turn, as for the previous experiments. The experiment comprised two repetitions, each with the fixed stimuli presented in a different random order.

The results of this experiment were analyzed by calculating the IACC of the variable stimulus that the subjects judged to have the same perceived auditory source width as each fixed stimulus. The means and associated 95% confidence intervals of these results are shown in Fig. 6, with the 1-IACC of each fixed stimulus (type 2 centered on 240 Hz) along the x axis and the 1-IACC of the matched variable stimulus (type 1 centered on 220 Hz) on the y axis.

These results indicate that for the variable type 1 stimulus to have the same auditory source width as the fixed type 2 stimulus with a given 1-IACC value, the former requires a lower 1-IACC value than the latter. In addition, it appears that the perceived auditory source width of the fixed type 2 stimuli with 1-IACC values of 0.8 and 1 are both matched by the same type 1 stimulus. These results can be interpreted as follows. First, it appears that the type 1 stimulus with a given 1-IACC value will be perceived to be slightly wider than the type 2 stimulus with the same 1-IACC value. Second, it appears that the auditory source width of the type 2 stimulus does not increase if the 1-IACC is increased from approximately 0.8 to 1. It is likely that the latter effect is the cause of the larger discrepancy between the expected and actual results for the type 2 stimulus at 200 Hz compared to the type 1 stimulus at 180 Hz. While the discrepancy between the IACCs is greater, the fact that there appears to be little change in the perceived width between a 1-IACC of 0.8 and a 1-IACC of 1 for the associated variable stimulus may mean that the perceptual discrepancy is small. For the stimuli that were tested, a reasonable conversion from the 1-IACC results for the type 2 stimuli to the 1-IACC results for the type 1 stimuli can be made by multiplying the former by 0.85.

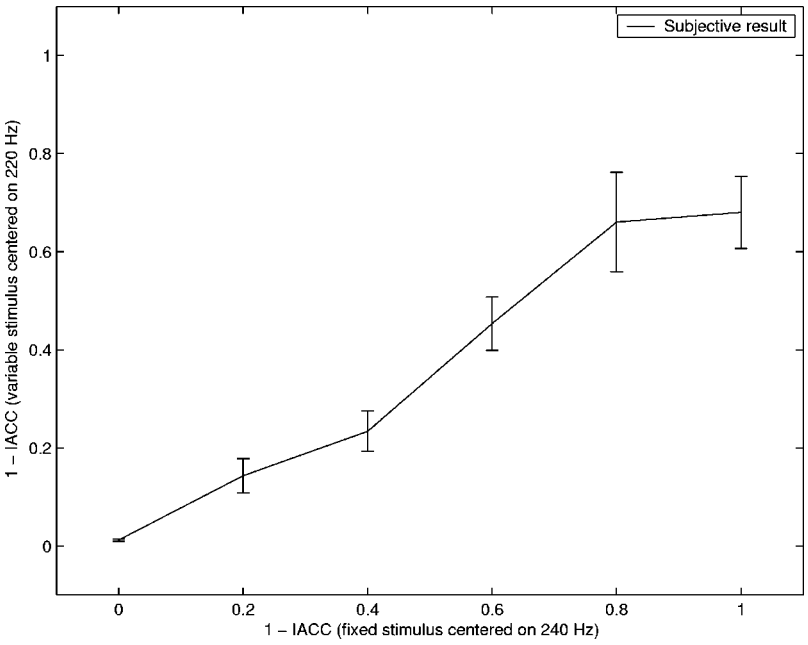


FIG. 6. Means and associated 95% confidence intervals of the 1-IACC of the type 1 variable stimulus with a center frequency of 220 Hz that the subjects judged to have the same auditory source width as the type 2 fixed stimuli with a center frequency of 240 Hz and a range of 1-IACC values.

V. EXPERIMENT 3—INVESTIGATION OF LOW-FREQUENCY STIMULI WITH IACC=0

A. Introduction

For some time, the relationship between the IACC and the perceived auditory source width at low frequencies has been considered to be a problem. For instance, signals that contain spectral information solely below approximately 500 Hz generally exhibit little variance in IACC compared to signals that contain spectral information solely above this range, and at low frequencies natural sounds rarely have an IACC much less than 1 (Hidaka *et al.*, 1995). This is thought to be due to the small physical distance between the ears compared to the wavelength of the sound for low-frequency stimuli, though this is complicated by the diffraction and scattering of the sound by the head (Lindevald and Benade, 1986). Tohyama and Suzuki (1989) found that the modeling equation derived by Cook *et al.* (1955) could be used to predict the IACC of a diffuse field, but only by using an equivalent ear spacing of 32.8 cm as opposed to the true interaural distance of approximately 16 cm.

Although the IACC of low-frequency stimuli (below a few hundred Hertz) rarely varies far from a value of 1, research has shown that low-frequency stimuli can vary significantly in their perceived auditory source width (Morimoto and Maekawa, 1988; Griesinger, 1997; Martens, 1999). Griesinger (1992) suggests that it is possible that the human auditory system compensates for the lack of decorrelation at low frequencies. The disparity between low-frequency IACC and perceived auditory source width has led previous investigations into IACC-based models of perceived auditory source width to recommend ignoring the results at low frequencies (Hidaka *et al.*, 1995). However, it is important to deal with the disparity in order to create a model that can accurately predict the perceived auditory source width of a wide range of stimuli.

In view of the above, an experiment was conducted to investigate whether the relationship between the IACC and the perceived auditory source width is dependent on frequency at low frequencies, and, if so, to derive an appropriate model.

B. Stimuli

The amplitude-modulation-based stimuli described in Eqs. (1)–(4) were again used for this experiment. All of the fixed stimuli had a value of G of 0.5, giving an IACC of 0. Fourteen different center frequencies (values of f_c) were used for each type of stimulus—type 1: 100, 140, 180, 260, 300, 340, 380, 420, 460, 580, 660, 740, 820, and 940 Hz; type 2: 120, 160, 200, 280, 320, 360, 400, 440, 480, 600, 680, 760, 840, and 960 Hz. In order to reduce the number of judgments required in the experiment, a number of the fixed stimuli that were identical to those in the previous experiment were omitted from this experiment, and for these the results were taken from the previous experiment. However, four of the fixed stimuli were identical between the two experiments in order to test the repeatability of the results and to confirm that the results from the two experiments could be combined.

As for the previous experiment, two types of variable stimuli were employed in order that the experiment compared like with like with regard to stimulus type. The same variable stimuli were used as in the previous experiment (type 1 $f_c = 220$ Hz and type 2 $f_c = 240$ Hz), in order to be able to combine the results from the two experiments. However, it is possible that the maximum perceivable auditory source width available with these variable stimuli is not as great as the auditory source width of the lower frequency fixed stimuli, which may cause a ceiling effect in the results. This will be considered in the analysis and discussion of the results.

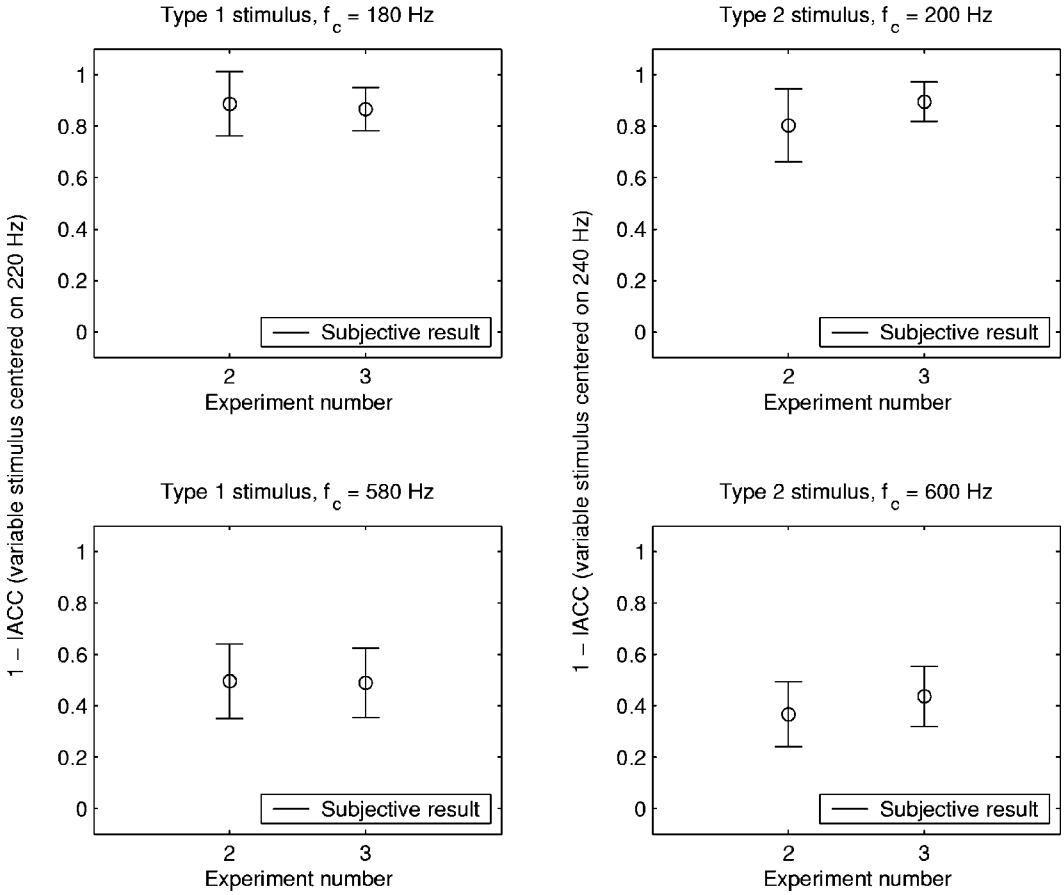


FIG. 7. Means and associated 95% confidence intervals of the 1-IACC of the variable stimulus that the subjects judged to have the same auditory source width as each fixed stimulus that was identical in experiments 2 and 3, separated by the stimulus type and the fixed stimulus center frequency.

C. Procedure

There were 20 fixed stimuli in the experiment; for each of these the subjects were asked to adjust the corresponding variable stimulus until they judged it to have the same auditory source width as the fixed stimulus. The experiment comprised two repetitions, each presented in a different random order. All seven selected subjects undertook the full experiment.

D. Results

The first stage of the analysis involved an examination of the similarity between the results of this experiment and the results of the previous experiment for the four repeated fixed stimuli, in order to determine whether the results from the two experiments could be combined. As for the previous experiments, the analysis involved calculating the IACC of the variable stimulus that the subjects judged to have the same perceived auditory source width as each fixed stimulus. The means and associated 95% confidence intervals of these results, shown in Fig. 7, display no statistically significant differences between the two experiments, which means that the data from the two experiments can be combined.

The means and associated 95% confidence intervals of the combined results are shown in Fig. 8, with the measured 1-IACC of the variable stimulus that the subjects considered to be the same auditory source width as each fixed stimulus on the *y* axis, and the center frequency of the fixed stimulus on the *x* axis, separated by the stimulus type.

The results shown in Fig. 8 support those from the previous experiment, in that the perceived auditory source width of a stimulus appears to increase as its center frequency falls below approximately 400 Hz. Above 400 Hz, the perceived auditory source width appears to be almost constant up to the frequency at which the breakdown of phase locking starts to take effect. It must be noted that the perceived auditory source width of the stimuli at low frequencies may be greater than that indicated in Fig. 8, as it is possible that the results below approximately 200 Hz were subject to a ceiling effect caused by the limited maximum auditory source width of the variable stimulus.

When investigating a suitable method for modeling this frequency dependency, it was logical to start with the research of Lindevald and Benade (1986) and Tohyama and Suzuki (1989). As the results of these two studies were almost identical over the frequency range tested, the simpler formula of Tohyama and Suzuki was employed, the equivalent of which is shown in Eq. (7), where *s* is the mean IACC in a diffuse field, *f* is the frequency, *d* is equivalent to the distance between the ears (calculated by Tohyama and Suzuki to be 0.328 m), and *c* is the speed of sound in air:

$$s = \mathrm{sinc}(2fd/c). \tag{7}$$

The accuracy of the match between the experiment results

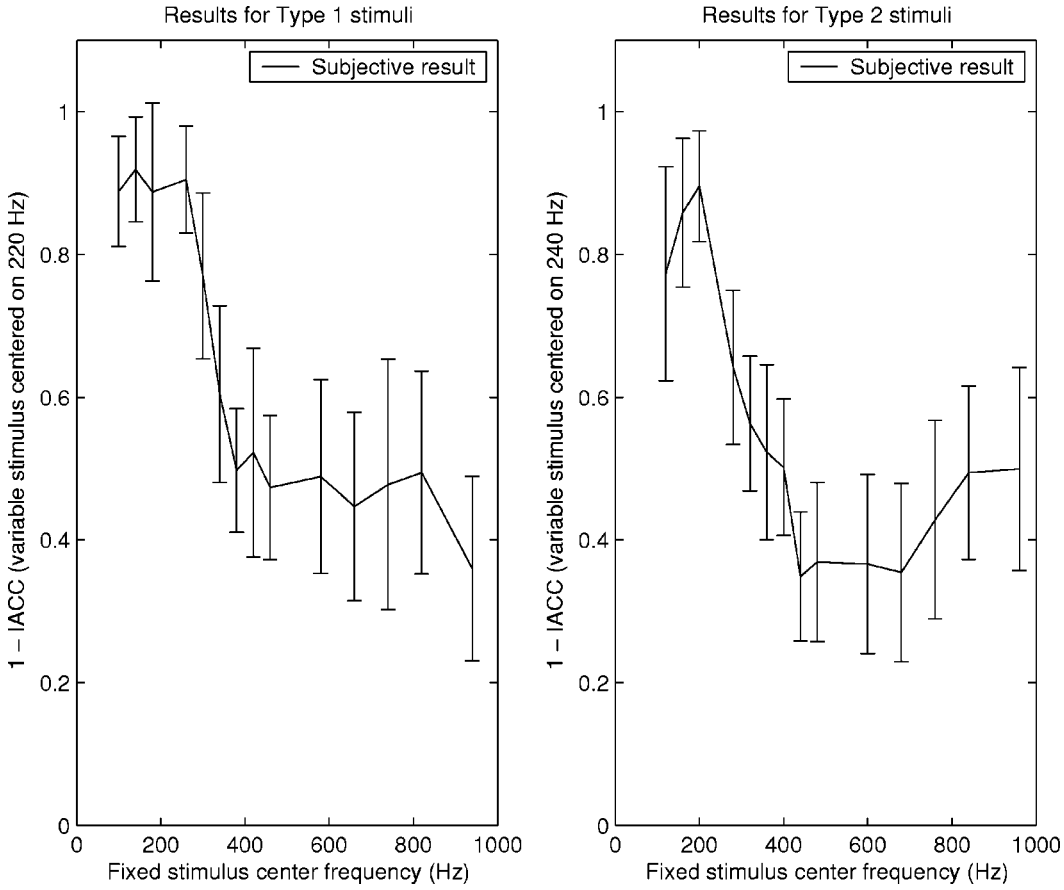


FIG. 8. Means and associated 95% confidence intervals of the 1-IACC of the variable stimulus that the subjects judged to have the same auditory source width as each fixed stimulus in experiment 3, with type 1 stimuli (variable stimulus centered on 220 Hz) on the left and type 2 stimuli (variable stimulus centered on 240 Hz) on the right.

and the predictions from this formula was evaluated as follows. An IACC was calculated for each of the fixed stimuli after preprocessing with the half-wave rectification and low-pass filtering derived from experiment 2 (giving a result in terms of 1-$IACC_f$). This value was then modified by multiplying it by s [Eq. (7)] (giving a result in terms of 1-$IACC_{fs}$). Finally, in order to examine the effect of the baseline compensation derived from experiment 1, b [Eq. (6)] was added (giving a result in terms of 1-$IACC_{fsb}$). This is termed the modeled result. The variable stimulus that the subjects matched to each fixed stimulus was also processed using this method (giving a result in terms of 1-$IACC_{fsb}$), and the means and associated 95% confidence intervals were calculated to give what is termed the subjective result.

The use of a frequency-dependent multiplier based on s, calculated using values of $d=0.328$ m and $c=330\ \mathrm{ms}^{-1}$, improved the match between the subjective and modeled results. However, it was found that using values of $d=0.4$ m and $c=330\ \mathrm{ms}^{-1}$ in Eq. (8) resulted in a closer match. The results of this are shown in Fig. 9:

$$s=\max(\mathrm{sinc}(2fd/c),0)/3+0.5. \tag{8}$$

It is apparent from the results shown in Fig. 9 that the modifications made to the basic IACC calculation [half-wave rectification and low-pass filtering of the input signal, multiplication of the result by the low-frequency compensation s of Eq. (8), and addition of the baseline compensation b of Eq. (6)] mean that the modeled results closely match the subjective results judgments in most cases. However, it is also apparent that the modeled results for the low-frequency stimuli (below approximately 200 Hz) are higher than the subjective results. As noted above, it is possible that the limited maximum perceived auditory source width of the variable stimulus caused the auditory source width of these low-frequency fixed stimuli to be underestimated. In view of this, it is possible that the matching at low frequencies is actually better than these results suggest.

E. Discussion

It is unclear why this experiment shows a better match between the modeled results and subjective results for a value of d in Eq. (8) of 0.4 compared to the value of 0.328 that matched the results of modeled measurements in a diffuse field in previous studies. If it is assumed that the frequency dependency at low frequencies is due to the human perceptual system compensating for the limited range of IACC values encountered at those frequencies, then it is possible that the difference between the actual interaural distance and the interaural distance that gives the most accurate prediction of the subjective results is due to overcompensation by the auditory system. However, it is not possible, from the current results, to accurately determine the cause of this effect.

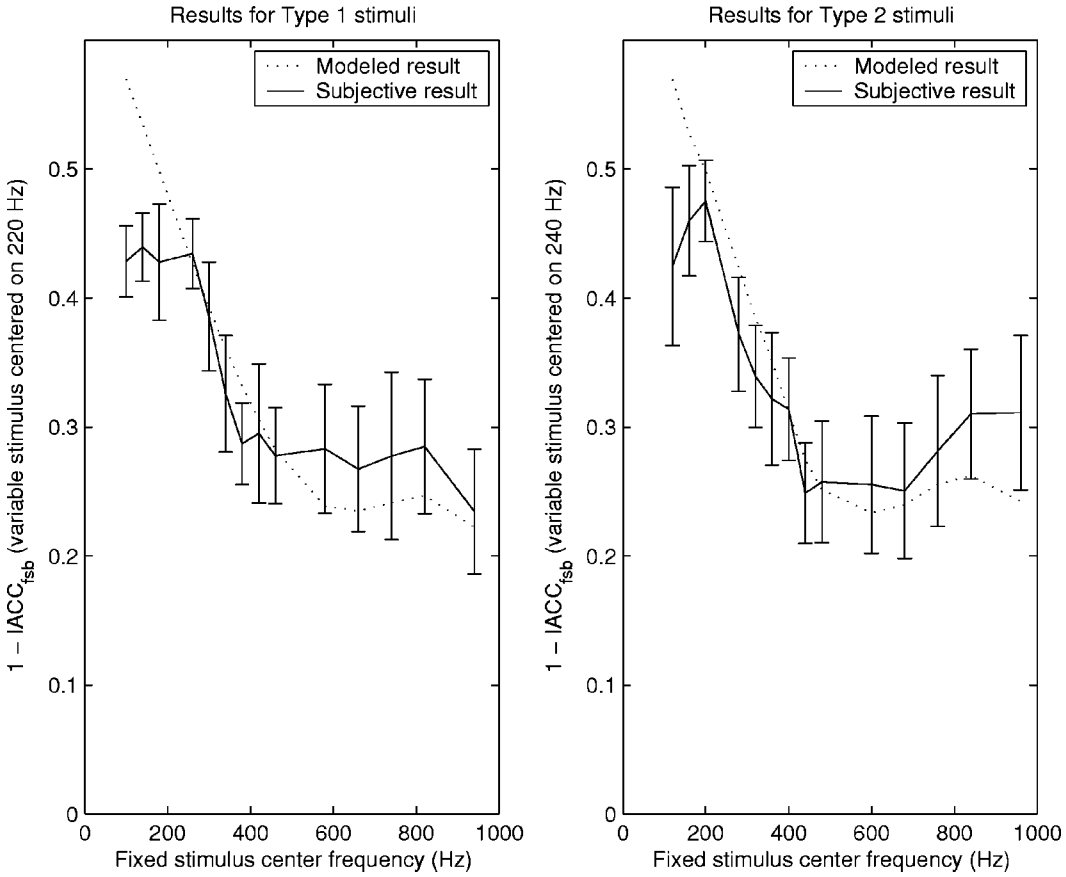


FIG. 9. Means and associated 95% confidence intervals of the 1-IACC$_{fsb}$ of the variable stimulus that the subjects judged to have the same auditory source width as each fixed stimulus in experiment 3 (subjective result), together with the 1-IACC$_{fsb}$ of each fixed stimulus (modeled result), with type 1 stimuli (variable stimulus centered on 220 Hz) on the left and type 2 stimuli (variable stimulus centered on 240 Hz) on the right.

VI. CONCLUSIONS

Three experiments were conducted to investigate features of the frequency dependency of the relationship between the IACC of an auditory stimulus and its perceived auditory source width. Three important findings emerged.

First, the perceived auditory source width of a stimulus with an IACC of 1 decreases monotonically as its frequency increases from a hundred to a few thousand Hertz, beyond which the perceived auditory source width increases again. The effect on the perceived auditory source width has been modeled with a simple formula, whose application in the subsequent experiments indicates that it provides a reasonable match to the subjective results.

Second, for IACC-based models of perceived auditory source width, it is important to take into account the loss of fine temporal detail at higher audio frequencies that results from the breakdown of phase locking in the neural output of the ear. This loss of detail causes the perceived auditory source width of higher frequency stimuli to be dependent on the IACC of the envelope of the signal, rather than the signal itself. This effect can be modeled by passing the input signal through a half-wave rectifier and a sixth-order low-pass Butterworth filter with a cutoff frequency of 1 kHz prior to measuring the IACC of the signal.

Finally, for low-frequency stimuli, human auditory perception appears to make up for the limited range of IACC values encountered in real situations by introducing an additional frequency dependency in the relationship between the IACC and the perceived auditory source width at these frequencies. This can be modeled by multiplying the IACC by a frequency-dependent factor derived from the expected range of IACC values that may be encountered in natural sound fields.

These three findings can be combined into a single generic model termed 1-IACC$_{fsb}$, which is described by Eq. (9), where w is a prediction of the perceived auditory source width, l_f and r_f are the left and right binaural signals that have been half-wave rectified and filtered with a sixth-order Butterworth low-pass filter with a cutoff frequency of 1 kHz, s is the frequency-dependent weighting factor defined in Eq. (8), and b is the frequency-dependent baseline function defined in Eq. (6):

$$w = \left[1 - \max_{-1\ \mathrm{ms} < \tau < 1\ \mathrm{ms}} \left(\frac{\int_{t1}^{t2} l_f(t) r_f(t+\tau)\, dt}{\left[\int_{t1}^{t2} l_f^2(t)\, dt \int_{t1}^{t2} r_f^2(t)\, dt \right]^{1/2}} \right) \right] s + b. \tag{9}$$

VII. DISCUSSION

The experiments that are described in this paper used stimuli whose characteristics were tightly controlled and constant over time. These were employed so that it was possible to match an objective value to a subjective effect as accurately as possible. However, the disadvantage of using

tightly controlled stimuli is that they are less externally valid (i.e., they are dissimilar to natural audio signals).

The stimuli had a narrow bandwidth and were time invariant, both in terms of their spectral content and their IACC. The perceived auditory source width of stimuli whose characteristics vary over time has not yet been determined. Additionally, the optimum method for combining measurements that are made in a number of separate frequency bands has not yet been derived. These factors must be investigated in order to further develop the model such that it can reasonably be used to predict perceived auditory source width for naturally occurring audio signals.

This paper only considered stimuli that have positive IACC values, as it has been shown that signals with a negative IACC are perceived to have differing spatial properties compared to equivalent signals with a positive IACC (Plenge, 1972). Further research is required into the perceived effect of stimuli with a negative IACC to allow this factor to be modeled.

The model that has been derived from the results of these experiments gives a prediction of the perceived auditory source width of the stimuli on a scale that allows comparison between stimuli, but is not related to any physical measurement of width. It would be useful to be able to predict the perceived auditory source width of stimuli in terms of a physical distance across for sounds perceived to be within the head, or in terms of a subtended angle for sounds perceived to be outside the head. Further research is required to convert the results of the measurement model to such a scale.

ACKNOWLEDGMENTS

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC), UK, Grant No. GR/R55528/01. The authors would like to thank Slawomir Zielinski and Søren Bech for their comments on earlier versions of this paper. The authors would also like to thank Armin Kohlrausch and two anonymous reviewers for their helpful comments.

APPENDIX: CALCULATION DETAILS

A number of different types of IACC-based calculation have been used in this paper. This appendix explains these in detail.

The basis for each calculation is the normalized cross-correlation function (NCC). In order for this to be equivalent to the IACC, it is essential that the signals analyzed are identical to those at the ears of the listener. The NCC is shown in Eq. (A1), where x and y are the two signals whose correlation is to be calculated, $t1$ to $t2$ is the period over which the correlation is calculated, and τ is an offset between the two signals under measurement:

$$\mathrm{NCC}(\tau)=\frac{\int_{t1}^{t2}x(t)y(t+\tau)dt}{[\int_{t1}^{t2}x^2(t)dt\int_{t1}^{t2}y^2(t)dt]^{1/2}}. \tag{A1}$$

The NCC is calculated across a range of τ that is equivalent to ± 1 ms, and the IACC is taken to be the maximum value across this range. The values of $t1$ and $t2$ are set to give a window length of 50 ms, and measurements are made at 1.5-ms steps throughout the input signal to quantify any variations in IACC over time. However, only the mean of these results is ultimately used. For the stimuli that were used in the experiment, it was found that the variations in IACC over time measured in this manner were no more than ± 0.01. The input signal is not filtered or weighted unless otherwise stated.

The variations on this basic calculation were derived from the results of the experiments and are denoted using subscript characters. A list of these is given below.

IACC—basic IACC calculation.

1-IACC—basic IACC calculation with the result then subtracted from 1.

1-$\mathrm{IACC_f}$—the input signal is half-wave rectified and low-pass filtered using a sixth-order Butterworth filter with a cutoff frequency of 1 kHz before using the basic IACC calculation with the result then subtracted from 1.

1-$\mathrm{IACC_{fs}}$—the input signal is half-wave rectified and low-pass filtered using a sixth-order Butterworth filter with a cutoff frequency of 1 kHz before using the basic IACC calculation with the result then subtracted from 1 and multiplied by the frequency-dependent weighting shown in Eq. (8).

1-$\mathrm{IACC_{fsb}}$—the input signal is half-wave rectified and low-pass filtered using a sixth-order Butterworth filter with a cutoff frequency of 1 kHz before using the basic IACC calculation with the result subtracted from 1, multiplied by the frequency-dependent weighting shown in Eq. (8), and then added to the frequency-dependent baseline function shown in Eq. (6).

All of the measurements described in the paper were made by analyzing the two-channel experimental stimuli in the digital domain. In order to ensure that the characteristics of the stimuli that the subjects heard were similar to those measured in the digital domain, a number of stimuli were reproduced over the system used in the experiment and captured using a head and torso simulator (HATS). It was found that the measurements made in the digital domain were within ± 0.02 of those measurements made of the signals that were captured with the HATS.

Ando, Y. (**1985**). *Concert Hall Acoustics* (Springer-Verlag, Berlin).

Bernstein, L. R., and Trahiotis, C. (**1992**). "Discrimination of interaural envelope correlation and its relation to binaural unmasking at high frequencies," J. Acoust. Soc. Am. **91**, 306–316.

Bernstein, L. R., and Trahiotis, C. (**1996a**). "On the use of the normalized correlation as an index of interaural envelope correlation," J. Acoust. Soc. Am. **100**, 1754–1763.

Bernstein, L. R., and Trahiotis, C. (**1996b**). "The normalized correlation: Accounting for binaural detection across center frequency," J. Acoust. Soc. Am. **100**, 3774–3784.

Blauert, J., and Lindemann, W. (**1986**). "Spatial mapping of intracranial auditory events for various degrees of interaural coherence," J. Acoust. Soc. Am. **79**, 806–813.

Boehnke, S. E., Hall, S. E., and Marquardt, T. (**2002**). "Detection of static and dynamic changes in interaural correlation," J. Acoust. Soc. Am. **112**, 1617–1626.

Brand, A., Behrend, O., Marquardt, T., McAlpine, D., and Grothe, B. (**2002**). "Precise inhibition is essential for microsecond interaural time difference coding," Nature (London) **417**, 543–547.

Cardozo, B. L. (**1965**). "Adjusting the method of adjustment: SD vs DL," J. Acoust. Soc. Am. **37**, 786–792.

Chernyak, R. I., and Dubrovsky, N. A. (**1968**). "Pattern of the noise images and the binaural summation of loudness for the different interaural corre-

lation of noise,'' Proc., 6th Int. Con. Acoust., A53–A56.
Constan, Z. A., and Hartmann, W. M. (**2001**). ''Sound localization by interaural time differences at high frequencies,'' J. Acoust. Soc. Am. **109**, 2485.
Cook, R. K., Waterhouse, R. V., Berendt, R. D., Edelman, S., and Thompson, M. C. (**1955**). ''Measurement of correlation coefficients in reverberant sound fields,'' J. Acoust. Soc. Am. **27**, 1072–1077.
Culling, J. F., and Summerfield, Q. (**1998**). ''Measurements of the binaural temporal window using a detection task,'' J. Acoust. Soc. Am. **103**, 3540–3553.
de Vries, D., Hulsebos, E., and Baan, J. (**2001**). ''Spatial fluctuation in measures for spaciousness,'' J. Acoust. Soc. Am. **110**, 947–954.
Gabriel, K. J., and Colburn, H. S. (**1981**). ''Interaural correlation discrimination: I. Bandwidth and level dependence,'' J. Acoust. Soc. Am. **69**, 1394–1401.
Gescheider, G. A. (**1997**). *Psychophysics, the Fundamentals* (Erlbaum, Hillsdale, NJ).
Glasberg, B. R., and Moore, B. C. J. (**1990**). ''Derivation of auditory filter shapes from notched-noise data,'' Hear. Res. **47**, 103–138.
Grantham, D. W., and Wightman, F. L. (**1979**). ''Detectability of a pulsed tone in the presence of a masker with time-varying interaural correlation,'' J. Acoust. Soc. Am. **65**, 1509–1517.
Griesinger, D. (**1992**). ''IALF—Binaural measures of spatial impression and running reverberance,'' Aud. Eng. Soc. Preprints, Vol. 92, No. 3292.
Griesinger, D. (**1997**). ''The psychoacoustics of apparent source width, spaciousness and envelopment in performance spaces,'' Acustica **83**, 721–731.
Hesse, A. (**1986**). ''Comparison of several psychophysical procedures with respect to threshold estimates, reproducibility and efficiency,'' Acustica **59**, 263–273.
Hidaka, T., Beranek, L. L., and Okano, T. (**1995**). ''Interaural cross-correlation, lateral fraction, and low- and high-frequency sound levels as measures of acoustical quality in concert halls,'' J. Acoust. Soc. Am. **98**, 988–1007.
ISO (**1987**). ''Acoustics—Normal equal-loudness level contours,'' ISO 226.
ISO (**1997**). ''Acoustics—Measurement of the reverberation time of rooms with reference to other acoustical parameters,'' ISO 3382.
Karjalainen, M. (**1996**). ''A binaural auditory model for sound quality measurements and spatial hearing studies,'' Proc., Int. Conf. Acoust. Sp. Sig. Proc. (ICASSP), pp. 985–988.
Keet, W. de V. (**1968**). ''The influence of early lateral reflections on the spatial impression,'' Proc. 6th Int. Con. Acoust., pp. E53–E56.
Kurozumi, K., and Ohgushi, K. (**1983**). ''The relationship between the cross-correlation coefficient of two-channel acoustic signals and sound image quality,'' J. Acoust. Soc. Am. **74**, 1726–1733.
Levitt, H. (**1971**). ''Transformed up-down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.
Lindevald, I. M., and Benade, A. H. (**1986**). ''Two-ear correlation in the statistical sound fields of rooms,'' J. Acoust. Soc. Am. **80**, 661–664.
Lynn, P. A., and Fuerst, W. (**1998**). *Introductory Digital Signal Processing with Computer Applications* (Wiley, Chichester).
Martens, W. L. (**1999**). ''The impact of decorrelated low-frequency reproduction on auditory spatial imagery: Are two subwoofers better than one?'' Proc., 16th Int. Aud. Eng. Soc. Conf., pp. 67–77.
Mason, R., Brookes, T., and Rumsey, F. (**2003**). ''Creation and verification of a controlled experimental stimulus for investigating selected perceived spatial attributes,'' Aud. Eng. Soc. Preprints, Vol. 114, No. 5771.
Meddis, R., Hewitt, M. J., and Shackleton, T. M. (**1990**). ''Implementation details of a computational model of the inner hair-cell/auditory-nerve synapse,'' J. Acoust. Soc. Am. **87**, 1813–1816.
Morimoto, M., and Iida, K. (**1995**). ''A practical evaluation method of auditory source width in concert halls,'' J. Acoust. Soc. Jpn. (E) **16**, 59–69.
Morimoto, M., and Maekawa, Z. (**1988**). ''Effects of low frequency components on auditory spaciousness,'' Acustica **66**, 190–196.
Morimoto, M., Iida, K., and Furue, Y. (**1993**). ''Relation between auditory source width in various sound fields and degree of interaural cross-correlation,'' Appl. Acoust. **38**, 291–301.
Morimoto, M., Ueda, K., and Kiyama, M. (**1995**). ''Effects of frequency characteristics of the degree of interaural cross-correlation and sound pressure level on the auditory source width,'' Acustica **81**, 20–25.
Okano, T. (**2000**). ''Image shift caused by strong lateral reflections, and its relation to inter-aural cross correlation,'' J. Acoust. Soc. Am. **108**, 2219–2230.
Perrott, D. R., Musicant, A., and Schwethelm, B. (**1980**). ''The expanding-image effect: The concept of tonal volume revisited,'' J. Aud Res. **20**, 43–55.
Plenge, G. (**1972**). ''Über das problem der im-kopf-lokalisation (On the problem of in-head localization),'' Acustica **26**, 241–252.
Robinson, D. E., and Jeffress, L. A. (**1963**). ''Effect of varying the interaural noise correlation on the detectability of tonal signals,'' J. Acoust. Soc. Am. **35**, 1947–1952.
Sakai, S., Asahi, N., Gotoh, T., Yoshii, F., and Akiya, H. (**1980**). ''On the simplified measurement of the interaural crosscorrelation coefficient and its effects on sound imaging and quality,'' Aud. Eng. Soc. Preprints, Vol. 67, No. 1697.
Sayers, B. M. (**1964**). ''Acoustic-image lateralization judgments with binaural tones,'' J. Acoust. Soc. Am. **36**, 923–926.
Sayers, B. M., and Cherry, E. C. (**1957**). ''Mechanism of binaural fusion in the hearing of speech,'' J. Acoust. Soc. Am. **29**, 973–987.
Stern, R. M., and Trahiotis, C. (**1997**). ''Models of binaural perception,'' in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. R. Anderson (Erlbaum, Hillsdale, NJ), Chap. 24, pp. 499–531.
Tohyama, M., and Suzuki, A. (**1989**). ''Interaural cross-correlation coefficients in stereo-reproduced sound fields,'' J. Acoust. Soc. Am. **85**, 780–786.
Wier, C. C., Jesteadt, W., and Green, D. M. (**1976**). ''A comparison of method-of-adjustment and forced-choice procedures in frequency discrimination,'' Percept. Psychophys. **19**, 75–79.
Yanagawa, H., Anazawa, T., and Itow, T. (**1990**). ''Interaural correlation coefficient and their relation to the perception of subjective diffuseness,'' Acustica **71**, 230–232.

Speech and melody recognition in binaurally combined acoustic and electric hearing

Ying-Yee Kong[a)]
Hearing and Speech Research Laboratory, Department of Cognitive Sciences, University of California–Irvine, Irvine, California 92697

Ginger S. Stickney
Hearing and Speech Research Laboratory, Department of Otolaryngology—Head and Neck Surgery, University of California–Irvine, Irvine, California 92697

Fan-Gang Zeng[b)]
Hearing and Speech Research Laboratory, Departments of Anatomy and Neurobiology, Biomedical Engineering, Cognitive Sciences, and Otolaryngology—Head and Neck Surgery, University of California–Irvine, Irvine, California 92697

(Received 24 July 2003; revised 10 December 2004; accepted 10 December 2004)

Speech recognition in noise and music perception is especially challenging for current cochlear implant users. The present study utilizes the residual acoustic hearing in the nonimplanted ear in five cochlear implant users to elucidate the role of temporal fine structure at low frequencies in auditory perception and to test the hypothesis that combined acoustic and electric hearing produces better performance than either mode alone. The first experiment measured speech recognition in the presence of competing noise. It was found that, although the residual low-frequency (<1000 Hz) acoustic hearing produced essentially no recognition for speech recognition in noise, it significantly enhanced performance when combined with the electric hearing. The second experiment measured melody recognition in the same group of subjects and found that, contrary to the speech recognition result, the low-frequency acoustic hearing produced significantly better performance than the electric hearing. It is hypothesized that listeners with combined acoustic and electric hearing might use the correlation between the salient pitch in low-frequency acoustic hearing and the weak pitch in the envelope to enhance segregation between signal and noise. The present study suggests the importance and urgency of accurately encoding the fine-structure cue in cochlear implants.
 [DOI: 10.1121/1.1857526]

PACS numbers: 43.66.Ts, 43.71.Ky [PFA] Pages: 1351–1361

I. INTRODUCTION

Cochlear implants have successfully restored partial hearing in severely hearing-impaired individuals. Recent studies have reported that many implant users can recognize 70%–80% of sentences presented in quiet. However, understanding speech in noise and music appreciation still remains a challenge for most implant users, due to the limitations of the electrode design and the signal processing scheme employed in current cochlear implants.

The poor speech perception in noise and music appreciation in cochlear-implant listeners are mainly due to their inability to encode pitch. The limited spectral resolution, especially the inaccurate encoding of low-frequency information, is believed to be the main reason for their poor pitch perception performance. Low-frequency information is important for both musical and voice pitch perception. It has been shown that speech recognition in the presence of a competing talker can be achieved by segregating the components of each voice using the fundamental frequency ($F0$) as a cue. Brokx and Nooteboom (1982) showed that listeners could identify the keywords in sentences more accurately against a background of competing speech by increasing the difference in $F0$. The $F0$ cue was shown to be effective in segregating competing voices even at low signal-to-noise ratios when the target speech did not show distinct peaks in the spectrum (Summerfield and Culling, 1992).

A pitch percept in normal auditory system can be elicited by either the place mechanism with resolved low-numbered harmonics or by the temporal mechanism following the temporal fine structure of the input signal. However, both pitch encoding mechanisms fail in current cochlear implants. The low-frequency information is neither appropriately represented by the place of stimulation nor by the temporal fine structure of the neural firing pattern. First, the relatively shallow insertion depth of present electrode arrays severely limit the transfer of low-frequency spectral information. The average insertion depth for the Nucleus implant was estimated to be 20 mm (Ketten *et al.*, 1998), which corresponds to the acoustic frequency lower limit of about 1000 Hz (Greenwood, 1990). Even with the latest electrode designs (such as the Clarion HiFocus, Nucleus Contour, and Med-El Combi40+), which are intended to provide a deeper insertion of up to 30 mm, there is still no guarantee that

[a)]Current address: MRC Cognition and Brain Sciences Unit, 15 Chaucer Rd., Cambridge, CB2 2EF, UK; electronic mail: ying-yee.kong@mrc-cbu.cam.ac.uk
[b)]Electronic mail: fzeng@uci.edu

low-frequency neurons can be stimulated due to both reduced nerve survival in deafened individuals and nontonotopic distribution of low-frequency neurons in the cochlea (Nadol *et al.*, 1989; Linthicum *et al.*, 1991). Second, low-frequency temporal information is not appropriately encoded in current speech processing strategies. All current coding strategies, except for the analog-based strategies such as the Compressed Analog (CA, Eddington, 1980) and Simultaneous Analog Stimulation (SAS, Kessler, 1999), extract only the temporal envelope of incoming signals from 6 to 22 frequency bands using a low-pass filter with a cutoff frequency below 500 Hz and amplitude modulate it to a fixed-high-rate pulsatile carrier. In these strategies, the low-frequency temporal information, namely the slowly-varying envelope ($<$50 Hz) and the periodicity cues (50–500 Hz) (Rosen, 1992), can be preserved in the temporal envelope, but they are encoded in the ''wrong places,'' i.e., locations in the cochlea that are tuned to higher frequencies. Furthermore, fine structure information of the input signal, which is the phase information defined mathematically by the Hilbert transform (Hilbert, 1912), is discarded in such processing schemes due to the usage of a fixed-rate carrier. While the low-frequency information conveyed by the temporal envelope can support speech recognition in quiet (e.g., Shannon *et al.*, 1995), it is not sufficient to support speech recognition in noise with limited spectral cues (e.g., Fu *et al.*, 1998; Zeng and Galvin, 1999; Qin and Oxenham, 2003; Stickney *et al.*, 2004) and robust pitch perception (e.g., Burns and Viemeister, 1981; Green *et al.*, 2002; Kong *et al.*, 2004).

As the audiological criteria for implant candidacy have become less stringent, individuals with substantial residual low-frequency hearing have received cochlear implants. Recent development of short-electrode arrays allows the preservation of low-frequency acoustic hearing in these patients (von Ilberg *et al.*, 1999; Gantz and Turner, 2003). For those who are implanted with the conventional long-electrode arrays, low-frequency acoustic information is also available by combining electric hearing with acoustic hearing from the nonimplanted ear (Dooley *et al.*, 1993; Tyler *et al.*, 2002). Availability of these individuals allows a unique opportunity to study the role of fine-structure information at low frequencies in auditory perception, particularly in tasks that depend on pitch perception (i.e., music perception and speech recognition in the presence of a competing talker).

Previous studies on speech perception with binaurally combined acoustic and electric hearing revealed mixed results (for adults, see Dooley *et al.*, 1993; Armstrong *et al.*, 1997; Tyler *et al.*, 2002; Ching *et al.*, 2004; for children, see Chmiel *et al.*, 1995; Ching *et al.*, 2001). Chmiel *et al.* (1995) reported significantly better speech performance in quiet with a combined use of hearing aids and cochlear implants in three of the six subjects. Similar results were also reported by Armstrong *et al.* (1997) and Ching *et al.* (2001 and 2004), showing better sentence and phoneme recognition performance with combined acoustic and electric hearing in both quiet and in multi-talker babble noise at a 10 dB signal-to-noise ratio. Anecdotally, some implant users reported that the additional low-frequency acoustic information improved both sound quality and sound localization (Armstrong *et al.*, 1997; Tyler *et al.*, 2002). Moreover, two of the three subjects tested in Tyler *et al.* (2002) reported that the acoustic and electric signals fused to form one integrated sound image. Potential incompatibility between acoustic and electric hearing has also been reported. For example, Tyler *et al.* (2002) reported that one of their subjects heard the acoustic and electric stimuli as separate sound sources. Blamey *et al.* (1996 and 2000) demonstrated a pitch mismatch and differences in the dynamic range and the shape of the iso-loudness curves between the acoustically and electrically stimulated ears. Dooley *et al.* (1993) also reported that some subjects discontinued using their hearing aids or cochlear implants after implantation. However, for those patients who adapted to both devices, the incompatibility between the two percepts did not seem to interfere with their speech recognition in both quiet and noise (Dooley *et al.*, 1993; Tyler *et al.*, 2002).

The first study by von Ilberg *et al.* (1999) on combined acoustic and electric hearing with specially designed short-electrodes showed better speech recognition in quiet with the additional low-frequency acoustic hearing compared to electric hearing alone. The range of improvement was 4 to 70 percentage points depending on the filtering configuration of the cochlear implant. A recent study by Turner and colleagues (2004) on short-electrodes also showed significant benefits of additional low-frequency acoustic hearing in speech recognition in noise. They compared the speech reception thresholds of spondee words in different noise backgrounds (steady-state noise versus competing sentences) in implant users with ''short-electrodes'' and traditional ''long-electrodes.'' They reported that speech reception thresholds improved by 15 dB in the competing talker background and 5 dB in the steady-state noise background in combined hearing recipients with the short-electrode implants compared to the traditional long-electrode users. They concluded that the better speech recognition performance in multi-talker babble noise with the additional low-frequency acoustic hearing was attributed to the ability of the listeners to take advantage of the voice differences between the target and the masker speech.

One of the reasons that current cochlear implant listeners have great difficulty in understanding speech in a fluctuating background of other talkers (Nelson *et al.*, 2003; Stickney *et al.*, 2004) is their impaired pitch perception ability. The goal of the present study was to investigate how residual low-frequency hearing from the nonimplanted ear provides information that is necessary for pitch perception, and in turn improves speech and music perception in cochlear implant listeners. Two experiments were conducted to reveal the role of low-frequency acoustic hearing in realistic listening situations that are exceptionally challenging for cochlear implant users. The first experiment was designed to evaluate speech recognition in the presence of another speech sound in three listening conditions: hearing aid (HA) alone, cochlear implant (CI) alone, and cochlear implant plus hearing aid (CI +HA). The second experiment was designed to evaluate melody recognition with primarily pitch cues in cochlear-implant users in the same three listening conditions. We hypothesized that the additional low-frequency acoustic infor-

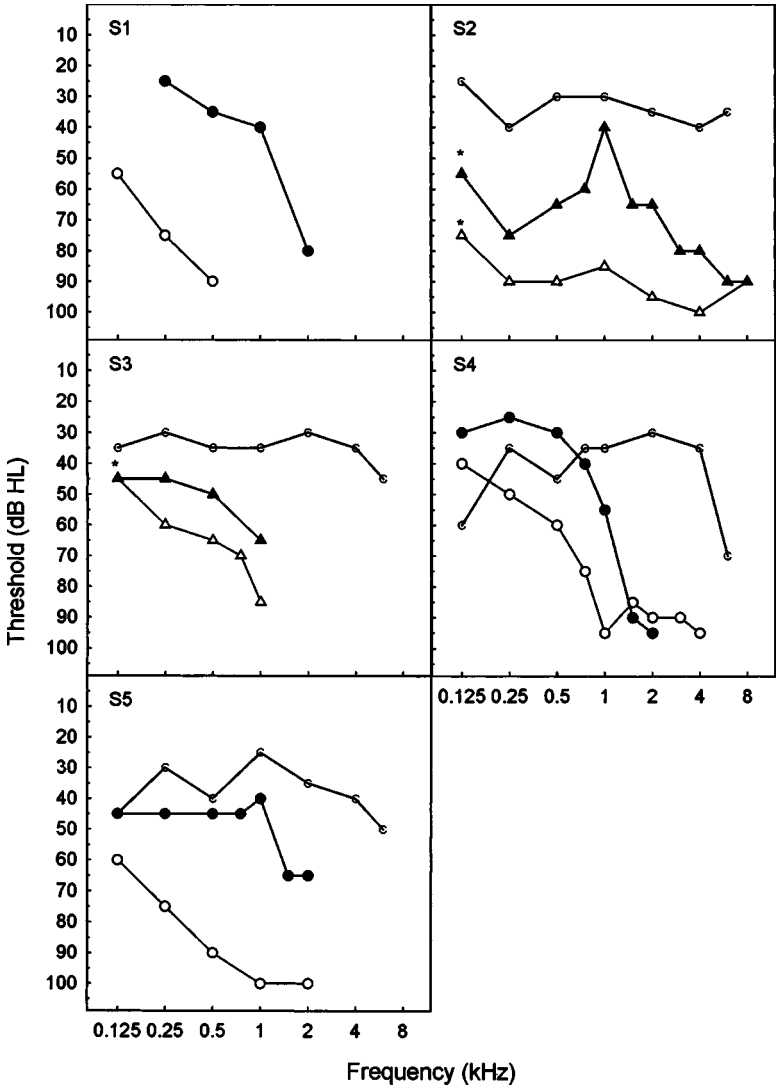


FIG. 1. Aided (closed symbols) and unaided (open symbols) thresholds in the nonimplanted ear (circles for the right ear; triangles for the left ear) and cochlear implant ear (indicated as "C"). Only thresholds at or below 100 dB HL were shown. The asterisk above the symbol indicates vibrotactile response. Implant thresholds in subject S1 were not obtained.

mation from the nonimplanted ear would provide more accurate pitch information to aid perceptual segregation of competing voices and to contribute significantly to musical pitch perception.

II. SPEECH RECOGNITION IN NOISE

A. Methods

1. Subjects

Four cochlear-implant subjects with significant residual acoustic hearing in the nonimplanted ear were recruited to participate in this study. They were two females and two males, with ages ranging from 49 to 79. Figure 1 shows their aided and unaided thresholds in the nonimplanted ear and the thresholds in the implanted ear. Their unaided thresholds showed moderate to profound loss at frequencies from 125 to 8000 Hz, but their aided thresholds showed only mild to severe loss at frequencies at or below 1000 Hz. The aided threshold averaged over 125, 250, and 500 Hz was 30, 70, 48, and 28 dB HL for subjects S1, S2, S3, and S4, respectively. While all other subjects had better aided thresholds below 1000 Hz, subject S2 had the lowest threshold (40 dB HL) at 1000 Hz and poorer thresholds below 1000 Hz. Thresholds from the cochlear implant for subjects S2, S3, and S4 were between 25 and 70 dB HL from 125 to 6000 Hz. The implant thresholds for S1 were not tested. Three out of the four subjects (S1, S3, and S4) continued to use their hearing aids on a daily basis, whereas S2 discontinued using his hearing aid after implantation in spite of residual hearing in the nonimplanted ear. Subject S2 did not use his hearing aid because of poor speech recognition rather than any perceived incompatibility between his hearing aid and cochlear implant.

All subjects were postlingually deafened and had at least one year of implant usage at the time of the test. They were native speakers of American English. Table I shows additional information regarding hearing history and implant type. Two subjects had the Clarion device, with S1 having the Clarion precurved electrode and S2 having the Clarion Hi Focus II with the positioner. The remaining two subjects (S3 and S4) used the Nucleus 24 device. Subject S1 used two different speech processing strategies: simultaneous analog stimulation (SAS) and multiple pulsatile sample (MPS), depending on the listening situations. Subject S2 and S4 used the continuous interleaved sampling (CIS) strategy, and S3 used the advanced combination encoder (ACE) strategy.

TABLE I. Biographical information on five cochlear-implant subjects.

Subject	Age[a]	Mus.[b]	Age onset[c]	Etiology	Yrs. exp.[d]	Device[e]	Strategy[f]	HA use[g]	Consonant[h] (%)	Vowel[i] (%)
S1	49	>20	4	Unknown	4	Clarion precurved	SAS/MPS	Y	46/54	51/40
S2	50	0	25	Unknown	2	Clarion HiFocus II	CIS	N	86	68
S3	69	0	37	Unknown	3	Nucleus 24	ACE	Y	54	51
S4	79	0	36	Unknown	1	Nucleus 24	CIS	Y	58	45
S5	19	0	3	Unknown	3	Clarion precurved	SAS	Y	41	33

[a]Age of the subject at the time of the experiments.
[b]Years of formal musical training.
[c]Age at the onset of hearing loss.
[d]Years of experience with the implant.
[e]Implant type.
[f]Processing strategy in the speech processor used during the experiments.
[g]Consistent use of hearing aid in the nonimplanted ear after implantation.
[h]Score (% correct) on consonant recognition in quiet in /aCa/ context.
[i]Score (% correct) on vowel recognition in quiet in /hVd/ context.

2. Stimuli

A subset of IEEE sentences (1969) recorded by Hawley *et al.* (1999) was used in this experiment. Each list consisted of ten sentences with five keywords per sentence. The target sentence was spoken by a male voice. Another sentence (competing sentence) spoken by a different male talker or by a female talker was used as a masker. The same competing sentence was used throughout testing (''Port is a strong wine with a smoky taste''). The target sentence was either presented alone or in the presence of the masker. The target and masker had the same onset, but the masker's duration was always longer than the target sentence. The target sentence was presented at approximately 65 dBA whereas the level of the masker varied from 45 dB to 65 dBA to produce five signal-to-noise ratios (SNR): +20, +15, +10, +5, and 0 dB.

3. Procedure

Subjects were evaluated under three listening conditions: hearing aid (HA) alone, cochlear implant (CI) alone, and cochlear implant with hearing aid (CI+HA). The subject's cochlear implant was turned off in the HA alone condition, and their hearing aid was turned off and the nonimplanted ear was plugged in the CI alone condition. These three listening conditions were evaluated in random order for each subject.

All tests were performed in a double-walled sound-treated booth. Both the target and masker sentences were presented via a loud speaker directly in front of the subject. Subjects used their own hearing aid and cochlear implant volume and sensitivity settings during the entire test session. All subjects were tested with the male masker, with subjects S2, S3, and S4 also being tested in a second test session with the additional female masker. S1 was not available for the second test session with the female masker. Prior to the test session, subjects were presented with two practice sessions of ten sentences each binaurally. In the first practice session, subjects were presented with sentences in quiet. The second practice session was used to familiarize listeners listening to the target sentence in the presence of the masker. In this practice session, two sentences were presented for each of the five SNR conditions used in the actual experiment. In the test session, each subject was presented with all five SNR conditions in a random order. There were 10 randomized sentences (5 keywords each), for a total of 50 keywords per SNR and 50 sentences for the test session. The subjects typed their responses at the keyboard and were encouraged to guess if unsure. Responses were collected and scored in terms of the number of words correctly identified using MATLAB software.

B. Results

Figure 2 shows percent correct scores as a function of SNR for sentence recognition in the presence of the male masker for both individual and average data (bottom right panel). Panels S1a and S1b represent results from subject S1 using the SAS and the MPS strategy, respectively. Results from the three listening conditions, hearing aid (HA) alone, cochlear implant (CI) alone, and the combined devices (CI +HA), are represented by closed squares, closed circles, and open triangles, respectively.

Both the individual and average data show the same trend: the hearing aid alone produced essentially zero speech recognition across different SNRs [$F(4,16)=1.88$, $p>0.05$], while the cochlear implant alone and the combined devices produced monotonically increasing performance as a function of SNR [CI alone: $F(4,16)=10.98$, $p<0.01$; CI +HA: $F(4,16)=23.09$, $p<0.001$]. The most interesting finding is that the combined hearing produced significantly better performance than the CI alone particularly in the higher SNR conditions by an average of 8 percentage points at a 15 dB SNR [$F(1,4)=10.78$, $p<0.05$] and 20 percentage points at a 20 dB SNR [$F(1,4)=27.13$, $p<0.01$].

Figure 3 shows sentence recognition scores with the female masker from subjects S2, S3, and S4. Similar to the male masker condition, qualitative trends were observed with the female masker: (1) the HA alone produced essentially zero speech recognition, (2) both the CI alone and the combined devices produced monotonically increasing perfor-

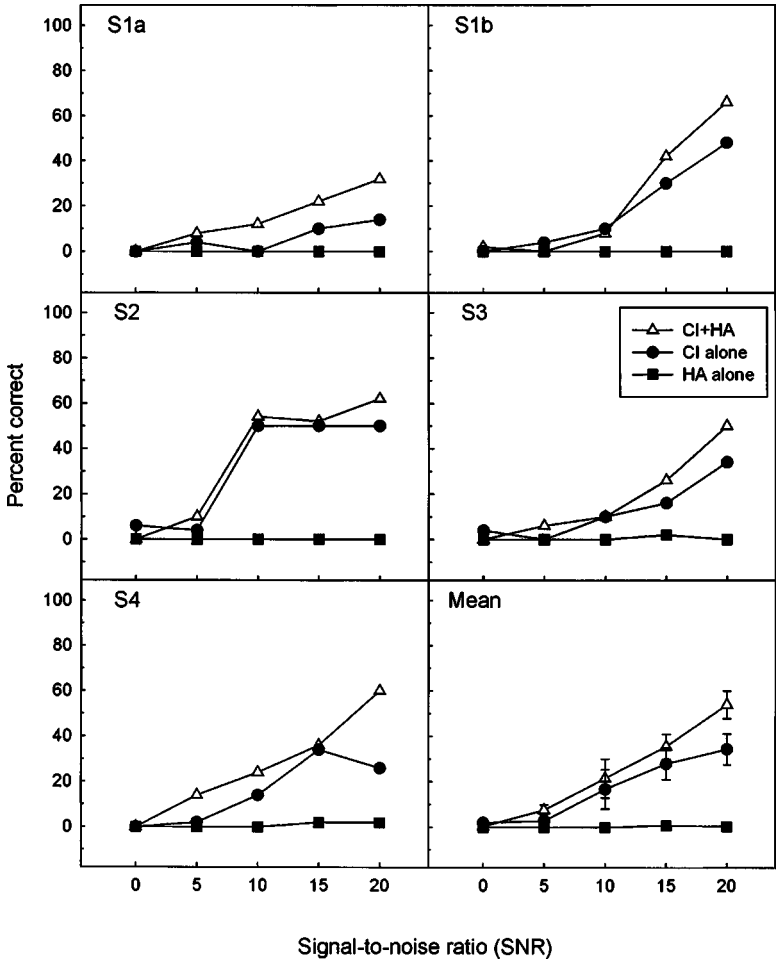


FIG. 2. Individual and mean sentence recognition scores (% correct) as a function of signal-to-noise ratio (SNR) with the male masker. The three functions in each panel represent the CI+HA (open triangles), CI alone (closed circles), and HA alone (closed squares) conditions. The vertical error bars in the mean graph represent the standard error of the mean.

mance as a function of SNR, and (3) the combined hearing produced better performance than either mode alone. The benefits of combined hearing compared to CI alone were observed in all subjects, with average improvement ranging from 8 to 25 percentage points from 0 to 20 dB SNR. However, due to the limited number of subjects and the large

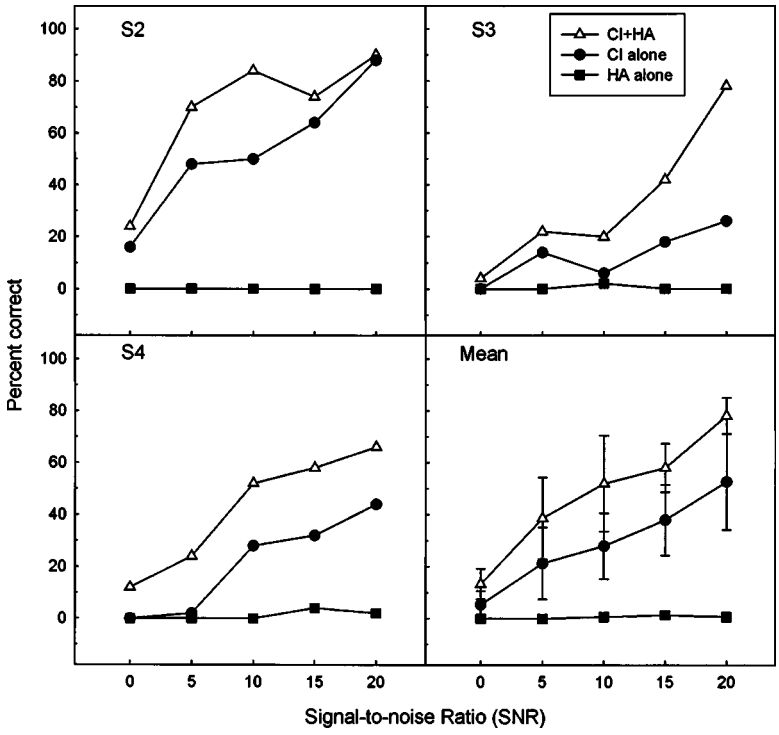


FIG. 3. Sentence recognition scores (% correct) as a function of SNR for subjects S2, S3, and S4 with the female masker.

intersubject variability of performance, the superior performance of combined hearing over CI alone was only found to be significant at 15 [$F(1,2)=19.10$, $p<0.05$] and 10 dB SNRs [$F(1,2)=18.75$, $p<0.05$]. Large differences between the male and the female maskers were observed in the combined hearing condition at all SNRs with an average of 21 percentage points better with the female masker than with the male masker. Significant difference between the two maskers in combined hearing was found in more challenging SNRs at 15 [$F(1,2)=100.00$, $p<0.01$] and 10 dB SNRs [$F(1,2)=19.24$, $p<0.05$]. In contrast, with the CI alone, the improvement of speech recognition with the female masker compared to the male masker was relatively small (average 9 percentage points) and no significant difference between maskers was found at any SNR [$F(1,2)=2.07$, $p>0.1$]. The improvement of the combined hearing over the CI alone was much greater for the female masker (average 19 percentage points) than the male masker (average 7 percentage points). For example, subject S2 improved by about 5 percentage points with the male masker, but the improvement with the female masker was 26 and 34 percentage points at 5 and 10 dB SNR, respectively. Similarly, subject S3 improved by 15 or less percentage points with the male masker, but the improvement with the female masker was 24 and 52 percentage points at 15 and 20 dB SNR, respectively.

III. MELODY RECOGNITION

A. Methods

1. Subjects

Five cochlear-implant subjects, including the same four cochlear-implant subjects from experiment I and an additional subject S5 participated in the melody recognition experiment. Only S1 had extensive musical training, while the rest had very limited music experience. Subject S5 was a non-native speaker of English, but he attended kindergarten in the United States at age 6. He reported learning all the melodies, except one, used in the experiment at a young age and was able to hum the tunes of these melodies. He was implanted with the Clarion precurved electrode and was using the SAS processing strategy. Like most of the subjects in this experiment, S5 continued to use his hearing aid on a daily basis. His average aided threshold for 125, 250, and 500 Hz was 45 dB HL.

2. Stimuli

Three sets of 12 familiar melodies, played by single notes, were generated using a software synthesizer (ReBirth RB-338, version 2.0.1). For each melody, rhythmic information was removed by using notes of the same duration (quarter notes with 350 ms in duration) with a silent period of 150 ms between notes. Therefore, pitch was the only available cue for melody recognition. Each melody consisted of 12–14 notes of its initial phrase. Three sets of the twelve melodies were generated in low-, mid-, and high-frequency ranges. In the low-frequency melody condition, all melodies were within a frequency range from 104 (G#2) to 261 Hz (C4), whereas the mid- (208 to 523 Hz) and high-range (414 to 1046 Hz) melodies were one and two octaves above the low-range melodies, respectively. The largest semitone difference between the highest and the lowest notes of the melody was 16 and the smallest difference was 7. Table II shows the titles of the melodies used in this experiment and the frequency components of each melody (for detailed information, see Kong *et al.*, 2004).

TABLE II. The 12 familiar melodies and their frequency ranges.

Melody[a]	Low range	Mid range	High range	Largest interval[b]	Int extent[c]
1	131–220 (C3-A3)	261–440 (C4-A4)	522–880 (C5-A5)	5th	9
2	122–184 (B2-F#3)	245–369 (B3-F#4)	490–738 (B4-F#5)	m3rd	7
3	110–174 (A2-F3)	220–348 (A3-F4)	440–696 (A4-F5)	4th	8
4	110–184 (A2-F#3)	220–369 (A3-F#4)	440–738 (A4-F#5)	6th	9
5	110–174 (A2-F3)	220–348 (A3-F4)	440–696 (A4-F5)	4th	7
6	131–196 (C3-G3)	261–392 (C4-G4)	522–784 (C5-G5)	m3rd	7
7	110–220 (A2-A3)	220–440 (A3-A4)	440–880 (A4-A5)	Octave	12
8	146–220 (D3-A3)	292–440 (D4-A4)	584–880 (D5-A5)	4th	7
9	131–196 (C3-G3)	261–392 (C4-G4)	522–784 (C5-G5)	5th	7
10	103–261 (G#2-C4)	207–523 (G#3-C5)	414–1046 (G#4-C6)	m6th	16
11	104–233 (G#2-A#3)	207–466 (G#3-A#4)	414–932 (G#4-A#5)	4th	14
12	110–184 (A2-F#3)	220–369 (A3-F#4)	440–738 (A4-F#5)	5th	9

[a]1=Twinkle, Twinkle, Little Star; 2=This Old Man; 3=She'll be Coming Round the Mountain; 4=Old MacDonald Had a Farm; 5=Lullaby, and Good Night; 6=Mary Had a Little Lamb; 7=Take Me Out to the Ball Game; 8=London Bridge is Falling Down; 9=Happy Birthday; 10=Star Spangled Banner; 11=Auld Lang Syne; 12=Yankee Doodle.
[b]Largest interval in melody (m=minor).
[c]Range in semitones between the highest and the lowest notes in the melodies.

3. Procedure

All subjects were tested in three listening conditions (HA alone, CI alone, and CI+HA) and three melody conditions (low, mid, and high) for a total of 9 conditions. For the HA alone and CI alone conditions, stimuli were presented at the subject's most comfortable level while they wore their hearing aid or cochlear implant at their usual settings. For the combined CI+HA condition, the presentation level was set the same as in the HA alone condition while the speech processor volume was adjusted to achieve the most comfortable loudness. The presentation level ranged from 70 to 85 dB SPL.

The titles of the 12 melodies were displayed on a computer screen and the subject was asked to choose the melody that was presented. A practice session with feedback was given before the actual test. For each experimental condition, melodies were presented three times in random order. Repetition of the stimulus was not allowed and visual feedback regarding the correct response was given immediately after

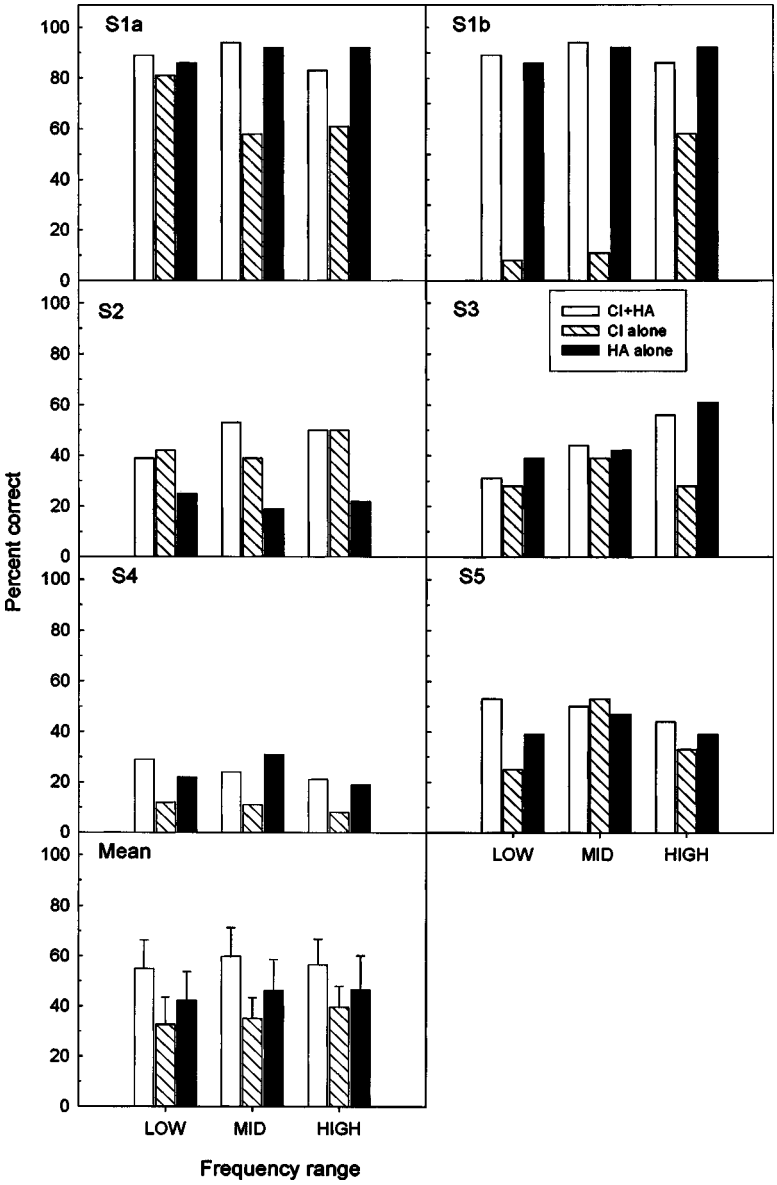


FIG. 4. Individual and mean melody recognition scores (% correct) for the three listening (CI alone, HA alone, and CI+HA) and three melody conditions (Low, Mid, and High). Vertical error bars represent the standard error of the mean.

the subject's response. As in experiment I, all three melody and all three listening conditions were presented in random order.

B. Results

Figure 4 shows individual and mean melody recognition results for the three melody (low, mid, and high) and listening conditions (HA alone=closed bars, CI alone=slanted bars, and CI+HA=open bars). Panels S1a and S1b represent results from subject S1 using the SAS and MPS strategies, respectively. Melody recognition performance varied remarkably from subject to subject in all listening conditions. Performance ranged from an average of 19% for S4 to 90% for S1 in the HA alone condition, from 8% for S4 to 81% for S1 in the CI alone condition, and from 21% for S4 to 92% for S1 in the CI+HA condition. Consistent with Kong *et al.* (2004), a difference in processing strategies was observed, with the SAS strategy producing better melody recognition than CIS-type strategy. For subject S1, her SAS strategy produced 73, 47, and 3 percentage points better performance than her MPS strategy for the low-, mid-, and high-range melodies, respectively. However, inconsistent with earlier studies on melody recognition, the melody recognition performance with cochlear implants alone in some of the subjects was considerably better than the chance performance level (e.g., Gfeller *et al.*, 2002; Kong *et al.*, 2004). It should be noted that the melody recognition performance with the mid-range melodies reported in Kong *et al.* (2004) was primarily obtained from cochlear implant users with the older devices (Clarion precurved and Nucleus-22) and with the envelope extraction processing strategies, namely MPS, CIS, and SPEAK. Preliminary data from our laboratory on a small group of users implanted with Nucleus 24, Clarion HiFocus II, and Med-El devices showed a different level of performance, with some performing similarly to the older device users at chance level and others in the range of 40%–80% correct. The reasons for this remarkable difference in performance between the newer and older devices will need further investigation, but it is not in the scope of discussion in this study.

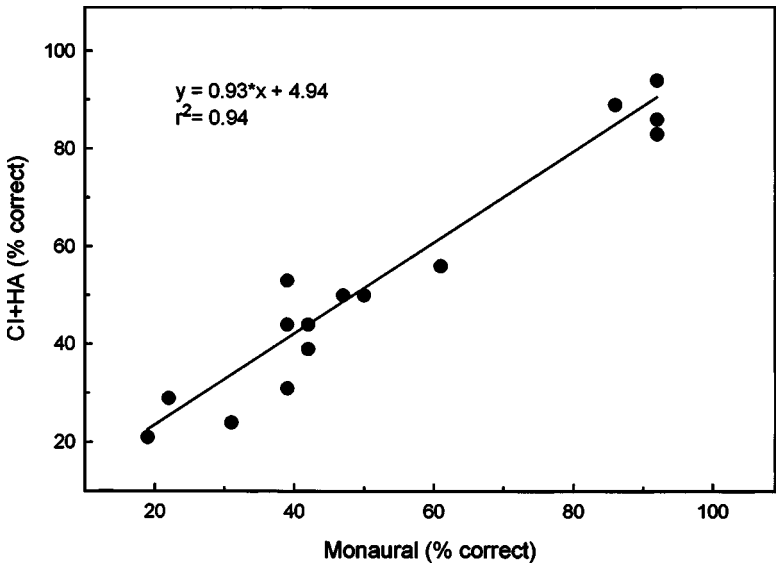


FIG. 5. Correlation of melody recognition (% correct) with binaural (CI+HA) and monaural (HA alone or CI alone) stimulation. Monaural data from S1a, S1b, S3, S4, and S5 are from the HA alone condition, but from the CI alone condition for S2. The slope of the regression line is 0.93 and the intercept is 4.94.

With the HA alone, the average melody recognition performance across all subjects and conditions was 45% correct. This is in direct contrast with the score of 0% obtained for the speech recognition in noise task from the first experiment. On average, HA alone produced an average of 17 percentage points better melody recognition than the averaged CI alone performance, but showed similar performance to the combined hearing condition. These patterns of results were observed in four out of the five subjects. The only exception was subject S2, who discontinued the regular use of his hearing aid after implantation and had very unusually poor aided thresholds in the frequency range ($<$1000 Hz) that was tested in this experiment. Due to the large intersubject differences in the melody recognition scores, a repeated measures ANOVA did not show significant difference between the HA alone and CI alone performance. Nevertheless, the trend of better performance with the HA alone than the CI alone was found in 14 out of the 18 cases, the probability of obtaining this result by chance is only 1.2%.

IV. DISCUSSION

A. Comparison between speech and melody recognition

Cochlear implant speech recognition performance in quiet has improved with advances in technology, but speech recognition in competing backgrounds and music perception remains challenging for implant users. One of the reasons for their poor performance in speech recognition in noise and music perception is their impaired pitch perception ability caused by both the limitations of the electrode design and the signal processing scheme employed in current cochlear implants. We hypothesize that providing the additional fine-structure information at low frequencies via the nonimplanted ear may allow for better encoding of pitch, which in turn can improve music appreciation and enhance speech recognition in competing backgrounds.

The present study showed that speech recognition in noise improved with combined acoustic and electric hearing compared to electric hearing alone, consistent with findings in earlier combined hearing studies (Armstrong *et al.*, 1997; Ching *et al.*, 2001 and 2004; Tyler *et al.*, 2002) and the results reported regarding ''short-electrode'' cochlear implants (Turner *et al.*, 2004). Turner *et al.* (2004) showed significantly lower spondee word reception thresholds in the presence of two simultaneously presented sentences than in steady-state white noise in ''short-electrode'' users, but not in the traditional long-electrode users. In contrast to speech recognition, the advantage of combined hearing was not observed in the melody recognition task. Instead, the performance with combined hearing was determined by the better ear (i.e., acoustic ear in S1, S3, S4, and S5, implant ear in S2), as indicated by Fig. 5 showing a highly significant correlation between the binaural and the best monaural conditions [$r^2=0.94$, $p<0.001$] and close to the unit slope (0.93) of the linear regression function.

The differential speech and music results reinforce the recently reported dichotomies in auditory perception, i.e., the fine-structure cue at low frequencies dominates pitch perception while the envelope cue dominates speech recognition (Smith *et al.*, 2002). The present results demonstrate this dichotomy with opposite patterns of results between hearing aid and cochlear implant performance in speech and melody recognition: the hearing aid (containing the fine-structure cue at low frequencies) produced no speech recognition but significant melody recognition, while the cochlear implant (containing the temporal envelope cue) produced significant speech recognition but relatively poor melody recognition. The inability to recognize speech with only low-frequency information is consistent with classic articulation index studies, where low-pass filtered speech ($\lesssim$800 Hz) was relatively unintelligible (e.g., French and Steinberg, 1947; Pavlovic *et al.*, 1986). However, this additional low-frequency acoustic information, when combined with the cochlear implant, produced significantly better speech recognition in noise than the implant alone condition.

B. Auditory segregation and grouping

The superior speech recognition performance in binaurally combined hearing over the CI alone may arise from the

benefits of (1) binaural processing including the binaural squelch effect (Carhart, 1965; Colburn, 1977) and/or diotic summation, a small benefit arising from listening with two ears compared to one ear with the identical signal and noise (Day *et al.*, 1988), (2) a monaurally based grouping and segregation mechanism, or (3) a combination of both. However, we argue that the presently observed improved speech recognition in noise with binaurally combined acoustic and electric hearing cannot be due to the binaural advantage. First, there are apparently no preserved level and phase differences between the acoustic and electric hearing, as required by the traditional binaural squelch. Both the target speech and masker were presented directly in front of the subjects in our study. Second, the advantage from diotic summation is small (Cox *et al.*, 1981) and it results mainly in better speech recognition in quiet (Kaplan and Pickett, 1981). This cannot account for the considerably large improvement of speech recognition in noise (averaged 19 percentage points with the female masker) with the combined hearing in our study. Third, similar improvement was obtained with combined acoustic and electric hearing on the same side with the short-electrode implant, providing evidence strongly against the binaural advantage hypothesis. Fourth, speech recognition in noise was improved more with the female masker than with the male masker, suggesting a monaurally based grouping and segregation mechanism.

Previous studies in which speech recognition in noise improved with the separation of the fundamental frequency have demonstrated the importance of voice pitch cues for segregating speech from competing backgrounds (e.g., Brokx and Nooteboom, 1982; Gardner *et al.*, 1989; Assmann and Summerfield, 1990). During voicing, the pulsing of the vocal folds gives rise to a consistent pattern of periodicity in the time wave form and corresponding harmonicity in the spectrum. Different from acoustic hearing, low harmonics cannot be resolved in current cochlear implants. The only pitch information available in the implants is from the reduced salience pitch cue provided by the temporal envelope (Burns and Viemeister, 1981; Faulkner *et al.*, 2000; Green *et al.*, 2002). The nonsalience of temporal envelope pitch can be demonstrated by the much poorer discriminability of modulation frequency and electric pulse rate than pure-tone frequency discrimination (Formby, 1985; Grant, 1998; Zeng, 2002). Thus, we hypothesize that the pitch difference in the temporal envelope is not robust enough to reliably separate the target and masker, particularly when both are dynamic speech sounds (e.g., Green *et al.*, 2002). We further hypothesize that the fine-structure information at low frequencies in the combined acoustic and electric hearing provides better $F0$ information that allows the cochlear-implant users to segregate the target from the masker. In the present study, the average fundamental frequency was 108 Hz for the target, 136 Hz for the male masker, and 219 Hz for the female masker (measured by the STRAIGHT program, courtesy of Kawahara, 1997). The significantly better speech recognition performance with the female masker compared to the male masker supported the idea that the availability of the fundamental frequency cue in combined acoustic and electric hearing was critical for separating the target speech from the masker speech.

The encoding of voice pitch in normal-hearing listeners can be achieved by the place coding or temporal coding mechanism, or both. A number of models have been proposed to investigate the auditory and perceptual processes by which normal-hearing listeners utilize the $F0$ difference when identifying the constituents of double vowels. Assmann and Summerfield (1990) tested different models to predict performance in normal-hearing listeners in identifying concurrent vowels with different fundamental frequencies. They reported that the place-time models, which estimated voice pitch using a periodicity analysis of the wave forms in each channel, were superior to the place models in the context of vowel identification. A purely temporal model by Meddis and Hewitt (1992) could even predict the improvement of segregation of simultaneous vowels as a function of $F0$ difference based on the pooled periodicity information which were summed across channels. These models suggested that temporal information, namely the periodicity cues, are critical for the segregation of competing sound sources.

The underlying mechanism for segregating competing sounds with combined acoustic and electric hearing is unclear. We propose that the segregation of target speech from the masker is based on the temporal periodicity cues in both the acoustic and electric signals. While the periodicity cue carried in the envelope alone does not provide sufficient $F0$ sensitivity to perceptually segregate target speech from the masker (Faulkner *et al.*, 2000; Green *et al.*, 2002), the presence of the additional salient temporal fine-structure cue at low frequencies in acoustic hearing, which is correlated with the periodicity cue in the temporal envelope in electric hearing, increases perceptual segregation between the signal and noise as well as improves grouping of the signal and that of the noise. This hypothesis is consistent with the recently reported poor (23% correct) speaker identification performance (Vongphoe and Zeng, 2004) and the absence of talker effect for speech recognition in the presence of a competing talker (Stickney *et al.*, 2004) in cochlear implant users. Even though there is no direct evidence to support this hypothesis at this stage, several predictions can be made to test its validity in the future. For example, should fundamental frequency be the main cue used by low-frequency acoustic hearing to improve electric hearing, we would predict minimal improvement for voiceless speech segments. Additionally, any mismatch between the fundamental frequency provided by acoustic hearing and the temporal envelope provided by electric hearing would result in a reduced benefit in combined hearing.

V. CONCLUSIONS

The present study implicates a dichotomy between the envelope and fine-structure cues at low frequencies in speech and melody recognition. The temporal envelope cue is sufficient for speech recognition, but not for melody recognition. On the other hand, the fine-structure cue at low frequencies is sufficient for pitch perception, but not for speech recognition. However, when the fine-structure cue in acoustic hear-

ing is combined with the envelope cue in electric hearing, significant improvement can be observed in speech recognition in a competing background. The greatest improvement was observed when the target and the masker had the largest difference in fundamental frequency, suggesting a monaurally based grouping mechanism rather than a binaurally based mechanism for the observed advantage with the combined acoustic and electric hearing.

The present study suggests the importance of appropriately encoding the fine-structure cue in cochlear implants. Although this fine-structure cue at low frequencies produces negligible intelligibility for speech recognition in quiet, it is critical for music perception, speech recognition in noise, and other listening situations including speaker identification and sound source segregation.

ACKNOWLEDGMENTS

We thank our cochlear-implant subjects for their time and dedication. We also thank Kaibao Nie for programming the user interface, J. Ackland Jones for generating the music stimuli and assisting in data collection and manuscript preparation, and Peter Assmann and Chris Turner for helpful suggestions. This work was supported in part by the National Institutes of Health, Department of Health and Human Services (A research supplement award to Y.-Y.K., F32 DC005900-01 to G.S.S., and 2 RO1-DC-02267 to F.G.Z.).

Armstrong, M., Pegg, P., James, C., and Blamey, P. (**1997**). ''Speech perception in noise with implant and hearing aid,'' Am. J. Otolaryngol. **18**, S140–S141.

Assmann, P. F., and Summerfield, Q. (**1990**). ''Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies,'' J. Acoust. Soc. Am. **88**, 680–696.

Blamey, P. J., Dooley, G. J., James, C. J., and Parisi, E. S. (**2000**). ''Monaural and binaural loudness measures in cochlear implant users with contralateral residual hearing,'' Ear Hear. **21**, 6–17.

Blamey, P. J., Dooley, G. J., Parisi, E. S., and Clark, G. M. (**1996**). ''Pitch comparsions of acoustically and electrically evoked auditory sensations,'' Hear. Res. **99**, 139–150.

Brokx, J. P. L., and Nooteboom, S. G. (**1982**). ''Intonation and the perception separation of simultaneous voices,'' J. Phonetics **10**, 23–36.

Burns, E. M., and Viemeister, N. F. (**1981**). ''Played again SAM: Further observations on the pitch of amplitude-modulated noise,'' J. Acoust. Soc. Am. **70**, 1655–1660.

Carhart, R. (**1965**). ''Monaural and binaural discrimination against competing sentences,'' Int. J. Audiol. **4**, 5–10.

Ching, T. Y. C., Incerti, P., and Hill, M. (**2004**). ''Binaural benefits for adults who use hearing aids and cochlear implants in opposite ears,'' Ear Hear. **25**, 9–21.

Ching, T. Y. C., Psarros, C., Hill, M., Dillon, H., and Incerti, P. (**2001**). ''Should children who use cochlear implants wear hearing aids in the opposite ear?,'' Ear Hear. **22**, 365–380.

Chmiel, R., Clark, J., Jerger, J., Jenkins, H., and Freeman, R. (**1995**). ''Speech perception and production in children wearing a cochlear implant in one ear and a hearing aid in the opposite ear,'' Ann. Otol. Rhinol. Laryngol. Suppl. **166**, 314–316.

Colburn, H. S. (**1977**). ''Theory of binaural interaction based on auditory-nerve data. II. Detection of tones in noise,'' J. Acoust. Soc. Am. **61**, 525–533.

Cox, R., DeChicchis, A. R., and Wark, D. (**1981**). ''Demonstration of binaural advantage in audiometric test rooms,'' Ear Hear. **2**, 194–201.

Day, G., Browning, G., and Gatehouse, S. (**1988**). ''Benefit from binaural hearing aids in individuals with a severe hearing impairment,'' Br. J. Audiol. **22**, 273–277.

Dooley, G. J., Blamey, P. J., Seligman, P. M., Alcantara, J. I., Clark, G. M., Shallop, J. K., Arndt, P., Heller, J. W., and Menapace, C. M. (**1993**). ''Combined electrical and acoustical stimulation using a bimodal prosthesis,'' Arch. Otolaryngol. Head Neck Surg. **119**, 55–60.

Eddington, D. K. (**1980**). ''Speech discrimination in deaf subjects with cochlear implants,'' J. Acoust. Soc. Am. **68**, 885–891.

Faulkner, A., Rosen, S., and Smith, C. (**2000**). ''Effects of the salience of pitch and periodicity information on the intelligibility of four-channel vocoded speech: implications for cochlear implants,'' J. Acoust. Soc. Am. **108**, 1877–1887.

Formby, C. (**1985**). ''Differential sensitivity to tonal frequency and to the rate of amplitude modulation of broadband noise by normally hearing listeners,'' J. Acoust. Soc. Am. **78**, 70–77.

French, N. R., and Steinberg, J. C. (**1947**). ''Factors governing the intelligibility of speech sounds,'' J. Acoust. Soc. Am. **19**, 90–119.

Fu, Q.-J., Shannon, R. V., and Wang, X. (**1998**). ''Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing,'' J. Acoust. Soc. Am. **104**, 3586–3596.

Gantz, B. J., and Turner, C. W. (**2003**). ''Combining acoustic and electrical hearing,'' Laryngoscope **113**, 1726–1730.

Gardner, R. B., Gaskill, S. A., and Darwin, C. J. (**1989**). ''Perceptual grouping of formants with static and dynamic differences in fundamental frequency,'' J. Acoust. Soc. Am. **85**, 1329–1337.

Gfeller, K., Turner, C., Mehr, M., Woodworth, G., Fearn, R., Knutson, J. F., Witt, S., and Stordahl, J. (**2002**). ''Recognition of familiar melodies by adult cochlear implant recipients and normal-hearing adults,'' Cochlear Implants Int. **3**, 29–53.

Grant, K. W., Summers, V., and Leek, M. R. (**1998**). ''Modulation rate detection and discrimination by normal-hearing and hearing-impaired listeners,'' J. Acoust. Soc. Am. **104**, 1051–1060.

Green, T., Faulkner, A., and Rosen, S. (**2002**). ''Spectral and temporal cues to pitch in noise-excited vocoder simulations of continuous-interleaved-sampling cochlear implants,'' J. Acoust. Soc. Am. **112**, 2155–2164.

Greenwood, D. D. (**1990**). ''A cochlear frequency-position function for several species—29 years later,'' J. Acoust. Soc. Am. **87**, 2592–2605.

Hawley, M. L., Litovsky, R. Y., and Colburn, H. S. (**1999**). ''Speech intelligibility and localization in multi-source environment,'' J. Acoust. Soc. Am. **105**, 3436–3448.

Hilbert, D. (**1912**). *Grundzuge einer Allgemeinen Theorie der linearen Integralgleichungen* (Teubner, Leipzig).

IEEE (**1969**). ''IEEE recommended practice for speech quality measurement,'' IEEE Trans. Audio Electroacoust. **17**, 225–246.

Kaplan, H., and Pickett, J. (**1981**). ''Effects of dichotic/diotic versus monotic presentation on speech understanding in noise in elderly hearing-impaired listeners,'' Ear Hear. **2**, 202–207.

Kawahara, H. (**1997**). ''Speech representation and transformation using adaptive interpolation of weighted spectrum: VOCODER revisited,'' Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Vol. 2, pp. 1303–1306.

Kessler, D. K. (**1999**). ''The Clarion mulit-strategy cochlear implant,'' Ann. Otol. Rhinol. Laryngol. **108**, 8–16.

Ketten, D. R., Skinner, M. W., Wang, G., Vannier, M. W., Gates, G. A., and Neely, J. G. (**1998**). ''In vivo measures of cochlear length and insertion depth of nucleus cochlear implant electrode arrays,'' Ann. Otol. Rhinol. Laryngol. Suppl. **175**, 1–16.

Kong, Y.-Y., Cruz, R., Jones, J. A., and Zeng, F.-G. (**2004**). ''Music perception with temporal cues in acoustic and electric hearing,'' Ear Hear. **25**, 173–185.

Linthicum, F. H., Fayad, J., Otto, S. R., Galey, F. R., and House, W. F. (**1991**). ''Cochlear implant histopathology,'' Am. J. Otolaryngol. **12**, 245–311.

Meddis, R., and Hewitt, M. J. (**1992**). ''Modeling the identification of concurrent vowels with different fundamental frequencies,'' J. Acoust. Soc. Am. **91**, 233–245.

Nadol, J. B., Young, Y.-S., and Glynn, R. J. (**1989**). ''Survival of spiral ganglion cells in profound sensorineural hearing loss: Implications for cochlear implantation,'' Ann. Otol. Rhinol. Laryngol. **98**, 411–416.

Nelson, P., Jin, S.-H., Carney, A., and Nelson, D. (**2003**). ''Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners,'' J. Acoust. Soc. Am. **113**, 961–968.

Pavlovic, C. V., Studebaker, G. A., and Sherbecoe, R. L. (**1986**). ''An articulation index based procedure for predicting the speech recognition performance of hearing-impaired individuals,'' J. Acoust. Soc. Am. **80**, 50–57.

Qin, M. K., and Oxenham, A. J. (**2003**). ''Effects of simulated cochlear-implant processing on speech reception in fluctuating masker,'' J. Acoust. Soc. Am. **114**, 446–454.
Rosen, S. (**1992**). ''Temporal information in speech and its relevance for cochlear implants,'' Philos. Trans. R. Soc. London, Ser. B **336**, 367–373.
Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). ''Speech recognition with primarily temporal cues,'' Science **270**, 303–304.
Smith, Z. M., Delgutte, B., and Oxenham, A. J. (**2002**). ''Chimaeric sounds reveal dichotomies in auditory perception,'' Nature (London) **416**, 87–90.
Stickney, G. S., Zeng, F.-G., Litovsky, R., and Assmann, P. (**2004**). ''Cochlear implant speech recognition with speech masker,'' J. Acoust. Soc. Am. **116**, 1081–1091.
Summerfield, Q., and Culling, J. F. (**1992**). ''Auditory segregation of competing voices: Absence of effects of FM or AM coherence,'' Philos. Trans. R. Soc. London, Ser. B **336**, 357–366.
Turner, C. W., Gantz, B. J., Vidal, C., and Behrens, A. (**2004**). ''Speech recognition in noise for cochlear implant listeners: Benefits of residual acoustic hearing,'' J. Acoust. Soc. Am. **115**, 1729–1735.
Tyler, R. S., Parkinson, A. J., Wilson, B. S., Witt, S., Preece, J. P., and Noble, W. (**2002**). ''Patients utilizing a hearing aid and a cochlear implant: speech perception and localization,'' Ear Hear. **23**, 98–105.
Von Ilberg, C., Kiefer, J., Tillein, J., Pfenningdorff, T., Hartmann, R., Sturzebecher, E., and Klinke, R. (**1999**). ''Electric-acoustic stimulation of the auditory system,'' ORL J. Otorhinolaryngol. Relat. Spec. **61**, 334–340.
Vongphoe, M., and Zeng, F.-G. (**2004**). ''Speaker recognition with temporal cues in acoustic and electric hearing,'' J. Acoust. Soc. Am. (submitted).
Zeng, F.-G. (**2002**). ''Temporal pitch in electric hearing,'' Hear. Res. **174**, 101–106.
Zeng, F.-G., and Galvin, III, J. J. (**1999**). ''Amplitude mapping and phoneme recognition in cochlear implant listeners,'' Ear Hear. **20**, 60–74.

Simulations of temporal patterns of oral airflow in men and women using a two-mass model of the vocal folds under dynamic control

Jorge C. Lucero[a)]
Department of Mathematics, University of Brasilia, Brasilia DF 70910-900, Brazil

Laura L. Koenig[b)]
Haskins Laboratories, 270 Crown Street, New Haven, Connecticut, 06511 and Long Island University, Brooklyn, New York 11201-8423

(Received 30 May 2004; revised 4 December 2004; accepted 7 December 2004)

In this study we use a low-dimensional laryngeal model to reproduce temporal variations in oral airflow produced by speakers in the vicinity of an abduction gesture. It attempts to characterize these temporal patterns in terms of biomechanical parameters such as glottal area, vocal fold stiffness, subglottal pressure, and gender differences in laryngeal dimensions. A two-mass model of the vocal folds coupled to a two-tube approximation of the vocal tract is fitted to oral airflow records measured in men and women during the production of /aha/ utterances, using the subglottal pressure, glottal width, and *Q* factor as control parameters. The results show that the model is capable of reproducing the airflow records with good approximation. A nonlinear damping characteristics is needed, to reproduce the flow variation at glottal abduction. Devoicing is achieved by the combined action of vocal fold abduction, the decrease of subglottal pressure, and the increase of vocal fold tension. In general, the female larynx has a more restricted region of vocal fold oscillation than the male one. This would explain the more frequent devoicing in glottal abduction–adduction gestures for /h/ in running speech by women, compared to men. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1853235]

PACS numbers: 43.70.Aj, 43.70.Bk, 43.70.Gr [DOS] Pages: 1362–1372

I. INTRODUCTION

Our purpose in this paper is to explore the capability of low-dimensional vocal fold models to reproduce vocal fold vibration onset–offset patterns observed experimentally during speech. At the same time, it analyzes how the vibratory behavior of the vocal folds is controlled during speech by individual speakers. The analysis follows an inverse dynamic approach, based on previous modeling studies by McGowan *et al.* (1995). Basically, this approach consists in fitting a model of the glottal source and vocal tract to collected speech records. The fit is performed by computing the time-varying laryngeal parameters that best reproduce the given speech signal at the model output. It must be done in a dynamic fashion, since the characteristics of the targeted speech signal, such as frequency and amplitude, vary in time. In this way, it represents a more demanding test for the vocal fold model than usual simulations of sustained voice, in which the laryngeal configuration is static. The model must be capable of simulating the pattern of variation of the speech signal, and within restrictions of smoothness and physiologically realistic ranges of the control parameters.

We consider airflow signals collected during productions of the utterance /aha/. This utterance is convenient for studying laryngeal control during speech, since it involves an open vocal tract with little upper articulatory movement (Koenig, 2000). Glottal abduction and subsequent adduction, combined with variations of other laryngeal biomechanical parameters such as subglottal pressure and vocal fold stiffness, yield elevated airflow, breathy voicing, and, in some cases, a cessation of voicing during /h/. By fitting a model of the larynx to the airflow data, we seek to infer the temporal control patterns of those parameters. Our purpose is to determine control strategies of voicing onset and offset used by speakers, and detect possible differences between female and male speakers. In addition, we intend to provide modeling support to our work on the development of speech motor control in children (e.g., Koenig, 2000; Koenig and Lucero, 2002).

For the vocal fold model, we adopt a version of the popular two-mass model of Ishizaka and Flanagan (1972). This is a simple model that has been widely adapted to yield realistic simulations of glottal source behavior (e.g., Koizumi *et al.*, 1987; Lous *et al.*, 1998; Pelorson *et al.*, 1994; de Vries *et al.*, 2002).

We will first describe the collected airflow data, followed by the laryngeal and vocal tract models and the fitting algorithm. Our results present the best fits of the model to the recorded data.

II. DATA

We collected data from eight normal English speaking subjects: four men and four women. All the subjects were between 20 and 40 years old, with mean ages of 27 and 28 years for men and women, respectively. The data consisted of oral airflow signals, recorded using a Rothenberg mask

[a)]Electronic mail: lucero@mat.unb.br
[b)]Electronic mail: koenig@haskins.yale.edu

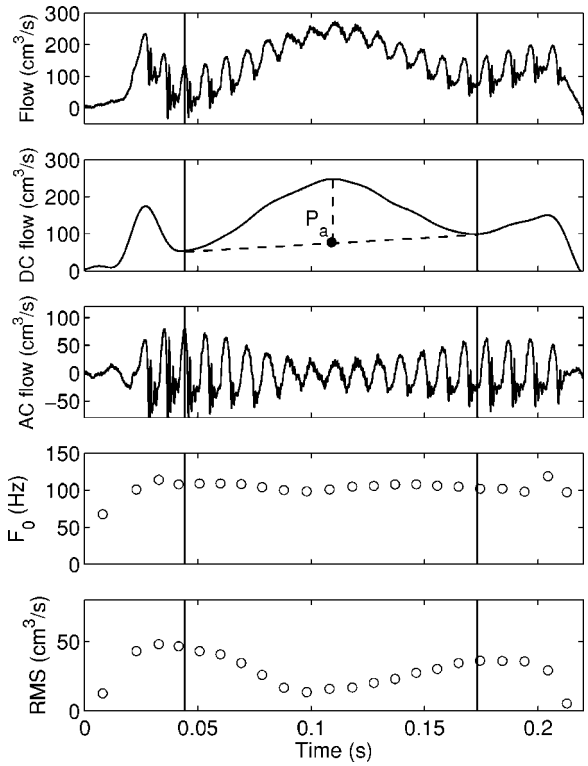


FIG. 1. Airflow data from subject MJB (male). From top to bottom: recorded airflow, dc component, ac component, fundamental frequency F_0, and rms value of ac amplitude. The vertical full lines mark the period selected as a target for the simulations. P_a defines a reference point (see the text).

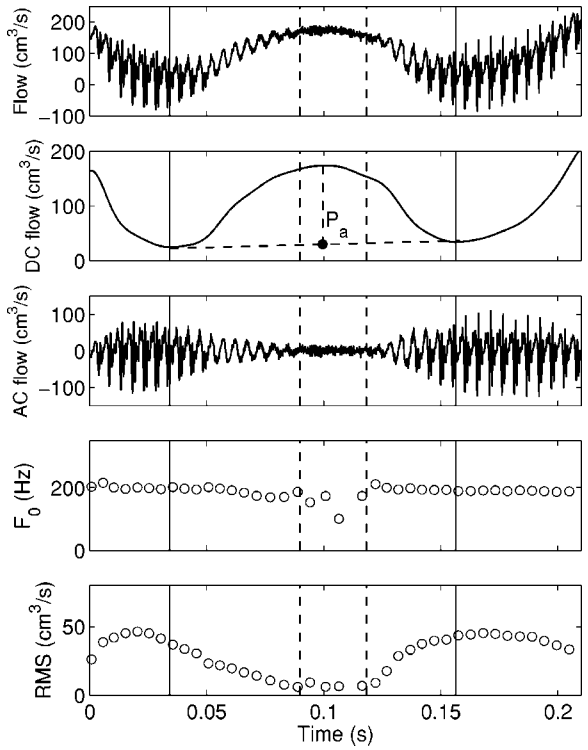


FIG. 2. The airflow data from subject FJH (female). From top to bottom: recorded airflow, dc component, ac component, fundamental frequency F_0, and rms value of ac amplitude. The vertical full lines mark the period selected as a target for the simulations, and the dashed lines mark the devoiced period. P_a defines a reference point (see the text).

while the subjects repeated the utterance "A papa hopper," with stress on the third syllable. The signals were filtered at 4.8 kHz, and sampled at 10 kHz. All productions were auditorily assessed to verify that they were perceptually acceptable versions of the intended utterance. We selected ten typical records from each subject, of various loudness levels, and showing both voiced and devoiced /h/.

The data was read into Matlab software, which was used for all further processing.

The dc flow component of each record was obtained by low-pass filtering it at a 50 Hz cutoff to eliminate glottal pulses, using a sixth-order Butterworth filter. The ac flow component was next computed as the difference between the unfiltered original signal and the filtered signals. The individual cycles of the ac component were then identified by using a zero-crossing algorithm with low pass filtering (Titze and Liang, 1993). From each cycle, we computed its rms value, and the fundamental frequency, as the inverse of its temporal length.

Figure 1 shows airflow data for a male subject, for the portion corresponding to the utterance /aha/.

Let us recall that airflow might be roughly considered, as directly proportional to the glottal area, when other parameters are constants. According to Titze (1988),

$$\Delta P = \frac{k_t \rho U_g^2}{2A_g^2}, \tag{1}$$

where ΔP is the transglottal pressure, k_t is a transglottal pressure coefficient, ρ is the air density, U_g is the airflow (volume velocity), and A_g the glottal area. Assuming also that the glottal air flow is equal to the oral flow, and that the transglottal pressure is constant, then the airflow results are proportional to the glottal area. Thus, the dc flow plot in Fig. 1 (second panel from the top) may be considered to represent the neutral glottal area, clearly showing the abduction–adduction gesture for the production of /h/.

In this case, the /h/ is fully voiced. The ac amplitude (the third panel from the top) decreases a bit during the glottal abduction–adduction, and there is almost no variation of the fundamental frequency F_0.

Figure 2 shows data corresponding to a female subject. In this case, the /h/ is devoiced. The dashed lines mark the period that we considered devoiced. In all cases, devoiced periods were identified as a region in which the computed values of F_0 were erratic (sudden jumps and discontinuities) and the rms values of ac amplitude reached a constant level close to zero (noise level), as may be seen in the F_0 and rms plots. Such regions were identified and marked by a visual inspection of the results.

We can also note a slight hysteresis effect for voice offset–onset in this example: voice stops and starts at different levels of the dc flow (169.5 and 152.2 cm^3/s, respectively). Assuming that the dc flow roughly represents the glottal neutral area, then we may say that voice offset and onset are achieved at different laryngeal configurations (different glottal areas). Further, the conditions for voice onset are more restricted than those for voice offset: to restart voice, the vocal folds must be driven closer together than the position at which voice stopped. This hysteresis effect has

TABLE I. Summary of recorded data. The table shows mean values of airflow parameters, grouped in fully voiced /h/ and devoiced /h/ for each subject. T is the length of the target region.

Subject		F_0 (Hz)	AC (cm^3/s)	dc (cm^3/s) base	dc (cm^3/s) peak	T (s)	Records
			Female subjects				
FEM	voiced	205	47.6	71.5	205.9	0.164	6
	devoiced	188	29.0	73.1	233.0	0.172	4
FGR	voiced	188	36.0	57.5	160.7	0.159	9
	devoiced	199	48.2	64.4	347.6	0.205	1
FJH	voiced	197	30.5	30.8	123.5	0.124	6
	devoiced	198	34.5	28.0	153.2	0.131	4
FRS	voiced	238	94.2	203.4	503.9	0.162	3
	devoiced	242	116.3	139.8	730.7	0.162	7
			Male subjects				
MJB	voiced	121	45.4	69.0	292.5	0.136	6
	devoiced	114	32.4	81.0	288.9	0.146	4
MJW	voiced	140	182.3	69.7	830.0	0.133	4
	devoiced	133	134.9	4.4	572.7	0.140	6
MSFA	voiced	123	61.6	71.2	342.4	0.144	8
	devoiced	102	44.0	71.9	362.4	0.162	2
MSO	voiced	149	25.5	68.2	230.8	0.158	4
	devoiced	155	25.9	71.4	277.5	0.170	6

been modeled using a subcritical Hopf bifurcation phenomenon (Lucero, 1999a), and it is common in cases of flow-induced vibrations (Thompson and Stewart, 1986).

In both figures, the vertical full lines mark the glottal abduction–adduction region, which was used later as target to fit the two-mass model. This region was defined between the two minima of the dc flow immediately before and after the /h/ peak. In the dc flow plots, P_a defines a weighted average between the two minima, and was used as a base reference value.

Table I summarizes the characteristics (mean values) of the recorded data. The dc base value is the weighted average indicated as P_a in Figs. 1 and 2, and the peak value corresponds to the peak glottal abduction. The ac rms and F_0 values were computed using a weighted average of values at the start and end of the target period, similarly to the dc base value. Main differences between male and female subjects appear in the fundamental frequency (higher in women than men, with no overlap between groups), and in the peak airflow value (higher in men than in women, with the exception of female subject FRS).

III. MODELS

A. Vocal folds

The larynx was modeled using a modified version of the two-mass model of the vocal folds (Ishizaka and Flanagan, 1972). Figure 3 (top) shows a sketch of the model. Each vocal fold is represented by two mass-damper-spring systems (m_1-b_1-s_1 and m_2-b_2-s_2), coupled through a spring (k_c). The two vocal folds are assumed identical, and they move symmetrically with respect to the glottal midline, in the horizontal direction.

When the glottis is open, the equations of motion may be written as

$$m_1\ddot{x}_1 + b_1(x_1,\dot{x}_1) + s_1(x_1) + k_c(x_1 - x_2) = f_1, \tag{2}$$

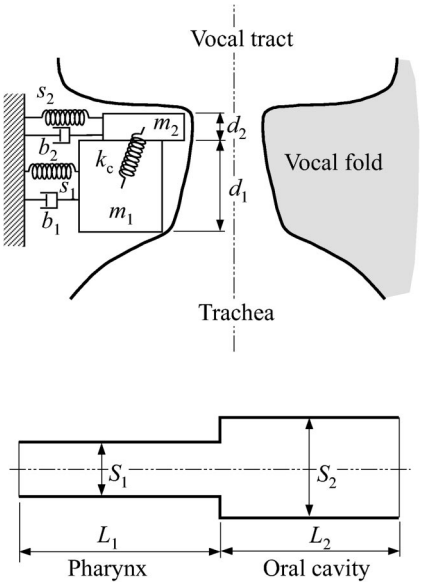


FIG. 3. Two-mass model of the vocal folds (top) and vocal tract model for vowel /a/ (bottom).

$$m_2\ddot{x}_2 + b_2(x_2,\dot{x}_2) + s_2(x_2) + k_c(x_2 - x_1) = f_2,$$

where b_i, s_i, f_i, with $i=1, 2$, denote the forces related to the tissue damping, elasticity, and the airflow, respectively, m_i are the masses, x_i are their horizontal displacements measured from a rest (neutral) position $x_0 > 0$, and k_c is the coupling stiffness. As in the work of Ishizaka and Flanagan work, we adopted a cubic characteristic for the tissue elastic forces, of the form

$$s_i(x_i) = k_i x_i(1 + 100x_i^2), \quad i = 1,2, \tag{3}$$

where k_i are stiffness coefficients and x_i is measured in cm.

The stiffness coefficients and masses were computed through a Q scaling factor (Ishizaka and Flanagan, 1972): $k_i = Q\hat{k}_i$, $k_c = Q\hat{k}_c$, and $m_i = \hat{m}_i/Q$. This Q factor may be regarded as a scaling factor for the natural frequencies of the model, and it provides a convenient way to control the oscillation frequency. It will be used later as one of the control parameters of the model.

For the damping forces, instead of the usual linear term $r_i\dot{x}_i$, we adopted a nonlinear characteristic of the form

$$b_i(x_i,\dot{x}_i) = r_i(1 + 150|x_i|)\dot{x}_i, \quad i = 1,2, \tag{4}$$

where r_i are damping coefficients. The reason was the need to limit the amplitude of the vocal fold oscillation as the glottal width increased. As an illustration, Fig. 4 shows a simulation result using a linear damping characteristic, and a glottal half-width increase from 0.009 to 0.102 cm. An increase in the amplitude of the glottal pulses when the glottal abduction begins may be noted. This increase is a shortcoming of the standard two-mass model, since it is not present in the recorded airflow data. It was also noted by McGowan *et al.* (1995) in their modeling work. It appears because, prior to the glottal area increase, the collision between the opposite vocal folds is the major factor that limits the vocal fold oscillation. When the glottal width increases, at the beginning of the abduction gesture, the distance from the vocal fold rest position to the midline increases, and thus there is

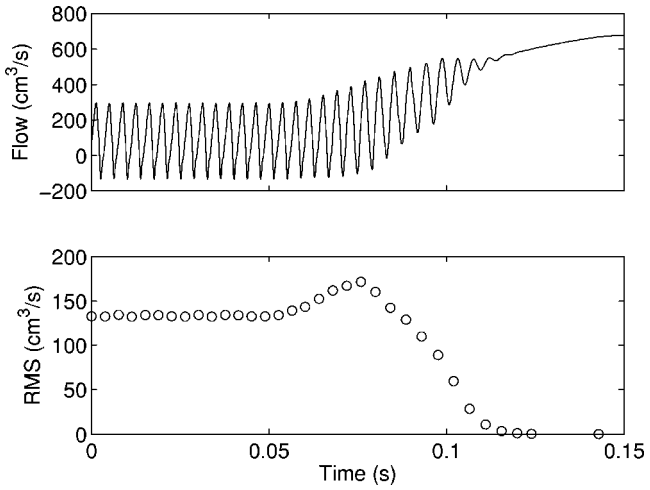


FIG. 4. Simulated airflow for vocal fold abduction, and linear vocal fold damping. Top: simulated airflow. Bottom: rms value of the ac component.

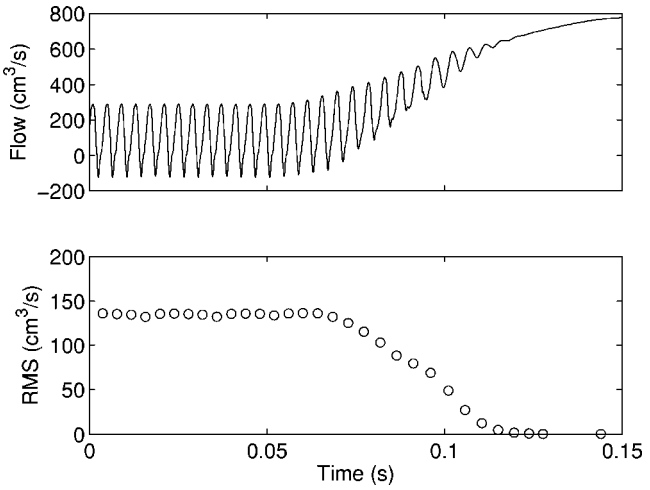


FIG. 5. Simulated airflow for vocal fold abduction, and nonlinear vocal fold damping. Top: simulated airflow. Bottom: rms value of the ac component.

more space for the oscillation. Eventually, for larger glottal widths, other factors come to limit the oscillation, such as the spring nonlinear characteristics and aerodynamic effects, and so the oscillation amplitude becomes smaller.

We found it necessary to eliminate this flow increase, to get reasonably good fits of the model to the data. McGowan *et al.* (1995) solved this issue by incorporating a representation of the anterior–posterior dimension of the vocal fold into the two-mass model. The posterior region was assigned a glottal area wider than the anterior region, and it was not allowed to vibrate. According to the authors, with that rudimentary upgrade of the two-mass model into a three-dimensional model, it was possible to abduct the vocal folds without increasing the amplitude of the airflow pulses.

Looking for alternative simpler solutions, we found that a nonlinear damping characteristic of the general form $r_i(1+\kappa|x_i|)\dot{x}_i$, where κ is a coefficient, produces the same result. The factor $r_i(1+\kappa|x_i|)$ acts as an equivalent damping coefficient, dependent on the displacement x_i. So, the larger the oscillation amplitude, the larger the damping on the system. This imposes a limit to the oscillation amplitude, by increasing the losses of the energy that fuels the oscillation, at large amplitudes. Note also that, at very low amplitudes $(x_i \to 0)$ the damping factor approaches a linear characteristic. This has the advantage that the threshold pressure and other conditions to start the vocal fold oscillation, which depend on the damping term, are not affected. Another strategy to limit the oscillation amplitude could be to increase the nonlinear part of the stiffness characteristics; however, this approach would also make the oscillation frequency more dependent on the oscillation amplitude. Using the damping term, on the other hand, does not have a direct effect on the oscillation frequency, and provides a simple and convenient way to control the oscillation amplitude through parameter κ. Figure 5 shows a simulation result with the same configuration of the model as for Fig. 4, but with a nonlinear damping characteristic (with $\kappa=150$), and an increased subglottal pressure in order to get the same airflow amplitude. The value of $\kappa=150$ was selected by inspection of the simulation results.

The adoption of a nonlinear damping term for the fold tissues is also in agreement with experimental data. It is known that the vocal fold tissues, like soft tissues in general, have strong nonlinearities in their viscoelastic properties (Alipour-Haghighi and Titze, 1985; Chan and Titze, 2000; Titze, 1994). More specifically, data obtained by Alipour-Haghighi and Titze (1985, Fig. 4) show that the time constant of tissue relaxation curves increases with the level of strain imposed. This result may be modeled by a damping factor that increases with strain, as in Eq. (4). Note that a linear damping term combined with the nonlinear stiffness in Eq. (3), which is the standard version of the two-mass model, would produce the opposite effect: relaxation curves with time constants that decrease with the level of strain. Experimental measures in other soft tissues and neuromuscular systems have also been interpreted as demonstrating strain-dependent viscous damping (e.g., Bilston *et al.*, 2001; Gielen and Houk, 1984; Lin and Rymer, 2001; Nasseri *et al.*, 2002). The proper form of the damping term in the vocal fold model is beyond the objectives of this paper. Here we have adopted the form in Eq. (4), because it provided good results when fitting the model to the data. However, other alternatives are possible, which should be evaluated in future research.

The contact between the opposite vocal folds is modeled as in the work of Ishizaka and Flanagan work (1972). We assume that at their rest position, both masses are at a distance x_0 from the glottal midline. Then, each mass i collides with its opposite counterpart at a displacement $x_i=-x_0$. During contact, the stiffness is increased,

$$s_i(x_i)=k_i x_i(1+100x_i^2)+3k_i(x_i+x_0)[1+500(x_i+x_0)^2],$$
$$x_i<-x_0 \quad \text{and} \quad i=1,2; \tag{5}$$

and the damping coefficient r_i is increased by adding 1 to the damping ratio: if $r_i=2\xi_i\sqrt{k_i m_i}$ is the value of the damping coefficient when the glottis is open, for a given damping ratio ξ_i, then during contact of mass i it assumes the value $r_i=2(\xi_i+1)\sqrt{k_i m_i}$.

B. Glottal and vocal tract aerodynamics

The glottal aerodynamics for the open glottis is modeled using a version of the Ishizaka and Flanagan equations (1972), updated with experimental results and a simplified version of the boundary layer model (Pelorson *et al.*, 1994, 1995; Lous *et al.*, 1998). For clarity, below we briefly describe the main equations, and leave the details for the cited references.

Letting the subglottal pressure be P_s, then the drop to pressure P_{11} at the glottal entry (at the lower edge of mass m_1) is, according to Bernoulli's equation,

$$P_s - P_{11} = \frac{\rho u_g^2}{2a_1^2}, \tag{6}$$

where ρ is the air density, u_g the volume velocity of glottal airflow, and a_1 is the cross-sectional lower glottal area, given by $a_1 = 2l_g(x_1 + x_0)$, where l_g is the vocal fold length. Note that flow contraction (*vena contracta*) is not considered, following Pelorson *et al.* (1995).

Along mass m_1, pressure drops to a value P_{12} due to air viscosity, given by

$$P_{11} - P_{12} = \frac{12\mu d_1 l_g^2 u_g}{a_1^3}, \tag{7}$$

where μ is the air viscosity, and d_1 is the width of mass m_1.

Two cases must be considered next, according to the glottal shape. Let us consider first the case in which the glottis is convergent or slightly divergent, i.e., $a_1 > k_s a_2$, where $k_s > 1$ is a suitable constant, and $a_2 = 2l_g(x_2 + x_0)$ is the cross-sectional upper glottal area. At the boundary between both masses there is a pressure variation, given by

$$P_{21} - P_{12} = \frac{\rho u_g^2}{2}\left(\frac{1}{a_1^2} - \frac{1}{a_2^2}\right), \tag{8}$$

where P_{21} is the air pressure at the lower edge of mass m_2. Next, there is a pressure drop along mass m_2 due to air viscosity, similar to Eq. (7),

$$P_{21} - P_{22} = \frac{12\mu d_2 l_g^2 u_g}{a_2^3}, \tag{9}$$

where P_{22} is the pressure at the glottal exit. At this point, due to the abrupt area expansion, the flow detaches from the glottal wall and forms a jet stream. We assume that all energy is lost in the stream due to turbulence (Pelorson *et al.*, 1994), and so

$$P_{22} = P_0, \tag{10}$$

where P_0 is the pressure input to the vocal tract.

When the glottis is more divergent, with $a_1 \leqslant k_s a_2$, we assume that the point of flow separation from the glottal wall moves inside the glottis and occurs at the boundary between both masses. This is a gross simplification of the boundary layer model, which shows that the point of flow separation moves upstream the glottis as a continuous function of the glottal angle of divergence. According to the Pelorson *et al.* (1994, 1995) results, in a divergent glottis the ratio between the glottal area at the point of airflow separation and the minimum glottal area becomes asymptotically constant at high Reynolds numbers. In measurements on a physical model of the larynx with a cylindrical profile for the vocal folds, they obtained the approximate ratio of 1.1. In a previous work (Lucero, 1999a), the flow separation equations were solved, assuming a linear variation of the glottal area along the glottis, and an approximate ratio of 1.3 was obtained. Here, we consider an intermediate value and thus assume $k_s = 1.2$, following also Lous *et al.* (1998).

In this case, again we assume that all airflow energy is lost due to turbulence from the point of detachment, and so

$$P_{21} = P_{22} = P_0. \tag{11}$$

Finally, the forces f_i acting on the masses are computed as $f_i = (P_{i1} + P_{i2})/2$.

The forces for the case of glottal closure are computed as follows (Ishizaka and Flanagan, 1972):

$$f_1 = d_1 l_g P_s, \quad x_1 \leqslant -x_0 \quad \text{or} \quad x_2 \leqslant -x_0; \tag{12}$$

$$f_2 = \begin{cases} d_2 l_g P_s, & \text{if } x_1 > -x_0, \quad x_2 \leqslant -x_0; \\ 0, & \text{if } x_1 \leqslant -x_0; \end{cases} \tag{13}$$

The vocal tract was represented by a standard two-tube configuration for vowel /a/ (Flanagan, 1972; Titze, 1994), shown in Fig. 3. Its equations were derived using a transmission line analogy, terminated in a radiation load of a circular piston in an infinite baffle. The elements of the transmission line were computed from the cross-sectional areas S_1, S_2 and lengths L_1, L_2 of the vocal tract tubes, using the standard equations of the analogy (Ishizaka and Flanagan, 1972; Flanagan, 1972).

We asumed that the shape of the vocal tract did not change over the utterance /aha/. This is something of an oversimplification: Both formants F1 and F2 were significantly higher across speakers in the stressed vowel compared to the unstressed vowel. Nevertheless, the mean formant differences as a function of stress were small (about 60 Hz for the women, and 40 Hz for the men). The /a/-like quality of the unstressed vowel in this utterance probably reflects coarticulation with the preceding and following /a/ vowels. There was also a general tendency across speakers for F1 and F2 to be highest in the loud condition, and lowest in the soft condition; this pattern was significant for F1 in the stressed vowel. Again, however, the overall formant differences were small (less than 75 Hz). Finally, in those cases where the formants were visible throughout the production of /h/, F1 and F2 generally showed little movement between the vowels. Thus, we are confident that variability in supraglottal postures as functions of stress and loudness was fairly minor in the modeled data. Moreover, we note that our objective was to fit the major variations in speakers' aerodynamic data, rather than details of token-to-token variability.

C. Male and female configurations

The following values were adopted for the male configuration of the vocal fold model (Flanagan, 1972; Ishizaka and Flanagan, 1972): $\hat{m}_{1M} = 0.125$ g, $\hat{m}_{2M} = 0.025$ g, $\hat{k}_{cM} = 25$ N/m, $\hat{k}_{1M} = 80$ N/m, $\hat{k}_{2M} = 8$ N/m, $\xi_{1M} = 0.1$, ξ_{2M}

$=0.6$, $l_{gM}=1.4$ cm, $d_{1M}=0.25$ cm, $d_{2M}=0.05$ cm, and for the vocal tract model (Flanagan, 1972; Goldstein, 1980): $S_{1M}=1\ \text{cm}^2$, $S_{2M}=7\ \text{cm}^2$, $L_{1M}=8.9$ cm, $L_{2M}=8.1$ cm. Here, subscript M denotes a male configuration.

To simulate female voices, McGowan *et al.* (1995) used the same male configuration of the model, as they were mainly interested in reproducing the gross dc airflow. Here, we intend to take a closer look at how onset and offset of the vocal fold vibration is controlled, and thus the size of the larynx and vocal tract might be an important factor. Female airways are smaller in cross-sectional area, and thus the pressure losses for viscosity are larger. The energy of the flow available for the oscillation is then smaller, and this fact may affect its control. Hence, we reduced the dimensions of the larynx and vocal tract to simulate female voices using scaling factors. According to Titze (1989), the size relation between male and female larynges is in the range 1.2–1.6, depending on the dimension. We adopted then a single scaling factor for the laryngeal dimensions, with a value $\beta_l=1.4$. Masses were accordingly computed dividing by β_l^3, to compensate for the volume reduction. For the tissue stiffness, we assumed a constant elasticity modulus for men and women. This assumption is again a simplification, since Titze (1989) reported a slightly stiffer tissue for females than for males, probably as a result of differences in tissue composition. Then, for a constant elasticity modulus, the stiffness coefficient is directly proportional to the cross-sectional area of the tissues, and inversely proportional to their length. The reduction of all dimensions by a factor β_l implies that stiffness is also reduced by this same factor. We assumed the same tissue damping ratios for men and women.

Denoting by subscript F the female configuration, the parameters of the vocal fold model are: $\hat{m}_{1F}=\hat{m}_{1M}/\beta_l^3=0.0456$ g, $\hat{m}_{2F}=\hat{m}_{2M}/\beta_l^3=0.0091$ g, $\hat{k}_{cF}=\hat{k}_{cM}/\beta_l=17.85$ N/m, $\hat{k}_{1F}=\hat{k}_{1M}/\beta_l=57.14$ N/m, $\hat{k}_{2F}=\hat{k}_{2M}/\beta_l=5.71$ N/m, $\xi_{1F}=\xi_{1M}=0.1$, $\xi_{2F}=\xi_{2M}=0.6$, $l_{gF}=l_{gM}/\beta_l=1$ cm, $d_{1F}=d_{1M}/\beta_l=0.179$ cm, $d_{2F}=d_{2M}/\beta_l=0.036$ cm.

For the female vocal tract we used data from Goldstein (1980): $L_{1F}=6.3$ cm, $L_{2F}=7.8$ cm. The cross-sectional areas were computed by scaling the male areas according to the total length relation of male to the female vocal tracts $\beta_v=(8.9+8.1)/(6.3+7.8)=1.21$. Thus, $S_{1F}=S_{1M}/\beta_v^2=0.688\ \text{cm}^2$, $S_{2F}=S_{2M}/\beta_v^2=4.816\ \text{cm}^2$.

IV. FIT TO COLLECTED SIGNALS

A. Control and target parameters

As control parameters for fitting the model to the data, we selected the subglottal pressure P_s, glottal half-width at the rest position of the vocal folds x_0, and Q factor. The glottal half-width characterizes the degree of abduction–adduction of the vocal folds, and should be the major control parameter for the flow increase during the production of /h/. The Q factor represents the degree of tension of the vocal folds, and is the main control for the fundamental frequency F_0 of the vocal fold oscillation. Subglottal pressure is a main control for the amplitude of the oscillation, and thus for the amplitude of the airflow ac component.

Vocal fold abduction and tension should be the main factors used by speakers to control phonation when producing /aha/ (Koenig, 2000; Löfqvist *et al.*, 1989; McGowan *et al.*, 1995). The value of subglottal pressure, on the other hand, should derive from the lung pressure, which should be roughly constant during the short duration of the utterance, as should the relative value between the resistances of the airways upstream and downstream the glottal entrance. As the vocal folds abduct for /h/ and glottal resistance decreases, subglottal pressure may decrease, and past experimental work has reported short term decreases in P_s when both the glottis and the supraglottal tract are open (Löfqvist, 1975; Ohala and Ohala, 1972). In the current work, we have chosen to use the subglottal pressure as a control parameter rather than explicitly controlling it during the abduction because the existant data do not permit precise quantification of how large its decrease should be: Löfqvist's (1975) measurements were made on a single male speaker of Swedish, producing voiceless obstruents, whereas Ohala and Ohala (1972) reported data on a single female speaker of Hindi. Our own past measurements (Löfqvist *et al.*, 1995; Koenig, 2000) further indicate that speakers may vary considerably in the degree of abduction they use during the consonant /h/ (cf. also Manuel and Stevens, 1989). Moreover, the well-known dimensional differences between adult male and female larynges (Titze, 1989; Kahane, 1982) imply that glottal areas during abduction may be significantly larger in men than women. Rather than speculate on the extent of group differences and/or intragroup variation, we approached this problem by allowing the subglottal pressure to vary as a fitting parameter.

As targets in the data, we chose the dc flow amplitude, and the rms and fundamental frequency of the ac component of the flow.

It is instructive to analyze briefly how the targets depend on the control parameters. Let us consider the flow dc component, ac component (rms value), and fundamental frequency F_0 as functions of the control parameters, in the form $(\text{dc},\text{ac},F_0)=\Gamma(P_s,x_0,Q)$, where Γ denotes a vector-valued function. We next compute their values when varying one of the control parameters, while keeping the other two fixed at some reference values. As reference, we used the standard values of the two-mass model $\hat{P}_s=800$ Pa, $\hat{x}_0=0.02$ cm, $\hat{Q}=1$ (Ishizaka and Flanagan, 1972).

Figure 6 shows results for the male configuration. The upper plot shows that both the ac and dc components increase almost linearly with the subglottal pressure. The fundamental frequency is almost constant, with a small negative slope. There is a threshold value for a subglottal pressure of 188 Pa (the threshold values were computed through additional simulations at a higher precision of the control parameters), which must be overcome for the start of the vocal fold oscillation. The middle plot shows that, when the glottal half-width is increased by abducting the vocal folds, the dc flow increases, since the glottal resistance decreases due to the larger glottal cross-sectional area. However, the vocal fold oscillation becomes more difficult (e.g., see Titze, 1988), so the ac amplitude decreases. There is a threshold value of 0.14 cm, above which the oscillation is no longer

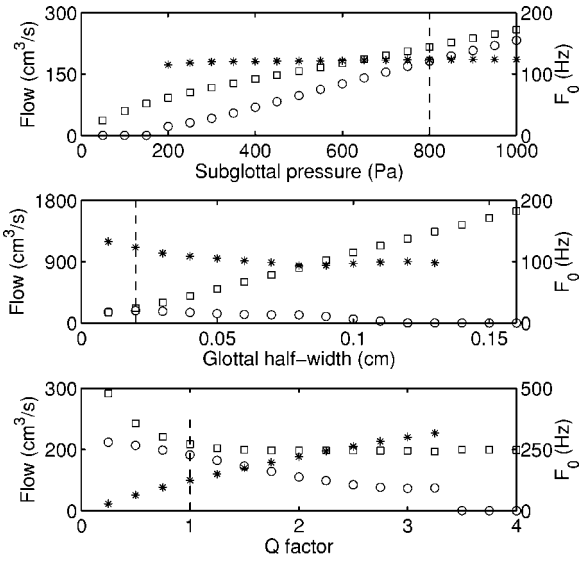


FIG. 6. Output targets as a function of the control parameters, for a male configuration of the model. Squares: dc component of oral airflow (scale at the left). Circles: rms value of the ac component of oral airflow (scale at the left). Asterisks: fundamental frequency of the ac component (the scale at the right). In the regions without asterisks, the AC component has a zero amplitude. The vertical dashed line marks the standard values of the model.

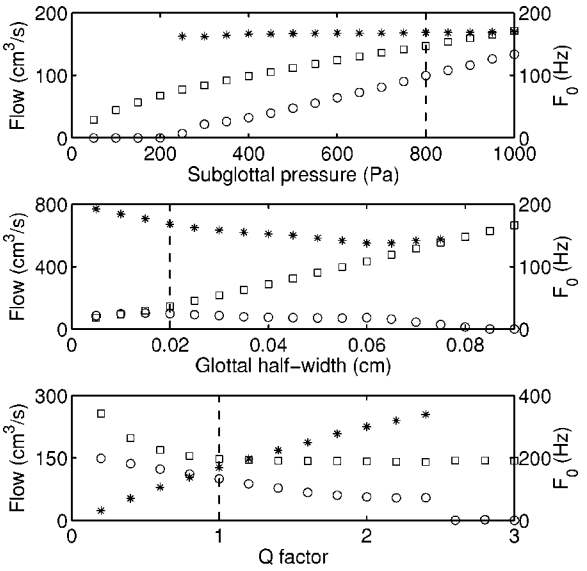


FIG. 7. Output targets as a function of the control parameters, for a female configuration of the model. Squares: dc component of oral airflow (scale at the left). Circles: rms value of the ac component of oral airflow (scale at the left). Asterisks: fundamental frequency of the ac component (scale at the right). In the regions without asterisks, the ac component has a zero amplitude. The vertical dashed line marks the standard values of the model.

possible. Considering, finally, variations of the Q factor, in the bottom plot, we see that its increase causes a linear increase of fundamental frequency. Let us recall that the Q factor is a linear scaling factor for the natural frequencies of the vocal folds. It also affects the dc and ac flows, causing them to decrease. As the Q factor increases, the vocal fold stiffness also increases, and so the oscillation is inhibited. Here too, there is a threshold value of 3.4, which sets an upper limit for the oscillation.

Figure 7 shows results for the female configuration. Results are qualitatively similar to the male case. The airflow values are lower than the male values, probably because of the increased resistance of the narrower airways. Frequency values are higher; although the stiffness coefficients are smaller than the male values by a factor of $1/\beta_l$, the masses are smaller by a factor of $1/\beta_l^3$, which results in higher natural frequencies. In fact, the fundamental frequency relations between the female and male cases for the same values of the control parameters are approximately equal to $\beta_l = 1.4$. The threshold value of the subglottal pressure is 240 Pa, larger than the male value. The difference is relatively small, 27% higher in women. The results seem in agreement with measures in men vs. women, which have found almost no gender differences in values of subglottal pressure (Koenig, 2000). Measured airflow values, on the other hand, are, in general, smaller in women, as found in the above simulations (for both dc and ac components).

The threshold values for the glottal half-width and Q factor are smaller than the male values, with values of 0.08 cm and 2.6, respectively. These results, together with the larger threshold pressure, imply that the adult female larynx has a more restricted region of vocal fold oscillation. Thus, voice would be harder to sustain in women than in men. In particular, the smaller threshold for the glottal width (43% reduction in women compared to men) would explain the higher incidence of devoicing during glottal abduction–adduction for /h/ in running speech in women as compared to men (Koenig, 2000). As the vocal folds abduct, they easily reach the oscillation offset threshold in women, whereas men would require more extreme degrees of abduction.

B. Algorithms

Fitting the model to the selected target parameters poses the difficulty that the dc flow is computed at each data point (time step) of the simulation results. The rms value and fundamental frequency, on the other hand, require a certain quantity of data points, since they are computed after each whole cycle. Thus, a dynamic algorithm that tracks the target signal and tunes the control parameters to get the correct output point by point (e.g., as in Pitermann and Munhall, 2001) would not work. Instead, the algorithm must track the target signals at larger steps, each of which must contain at least one ac cycle. However, the length of the ac cycles, and even the presence or absence of an ac component, depend on the same values of the control parameters that the algorithm must determine.

After trying several strategies, we found that the following one produced good results for all records. The target parameters were determined from the measured records at five instants of time: the two minima of the dc flow immediately before and after the /h/ peak (vertical solid lines in Figs. 1 and 2), the maximum at the /h/ peak, and the two midpoints between the peak and the two minima.

The algorithm worked then in two main stages. In the first stage, each set of target parameters for the five time points was considered separately. For each set, simulations of oral airflow were produced using constant control parameters. The dc component, and rms and fundamental frequency of the ac component were computed from the simu-

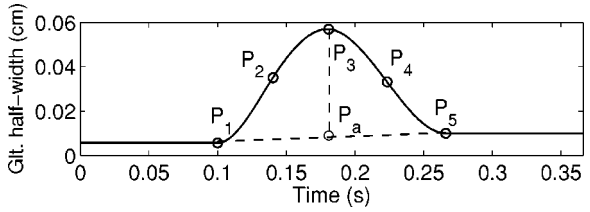


FIG. 8. Control time series for the glottal half-width. Points P_i are determined by the optimization algorithm described in the text. Before P_1 and after P_5, the parameter has a constant value, for a period of 0.1 ms. Between P_1 and P_5, intermediate points are determined using cubic spline interpolation. P_a is a weighted average of P_1 and P_5, used as a reference base value.

lated results and compared with the targets. In the devoiced regions (no ac component), fundamental frequency was not considered. As a cost function, a weighted sum of the square difference between these simulated parameters and the measured targets was computed,

$$C=\alpha_1(\mathrm{dc}^*-\mathrm{dc})^2+\alpha_2(\mathrm{ac}^*-\mathrm{ac})^2+\alpha_3(F_0^*-F_0)^2, \tag{14}$$

where the asterisk denotes the target values. The weights were given the values $\alpha_1=1/\overline{\mathrm{dc}}$, $\alpha_2=1/\overline{\mathrm{ac}}$, $\alpha_3=1/\overline{F_0}$, where the overbar denotes mean values over the target region. In addition, upper and lower limits were imposed on the control parameters, to avoid regions of unrealistic values. Optimal values of the control parameters were then determined using a Nelder–Mead simplex method (Kincaid and Cheney, 2002) implemented in MATLAB, to minimize the cost function. These values were used in the next stage as initial values for a refined optimization.

In the second stage, a time series for each control parameter was constructed, as illustrated in Fig. 8 for the case of the glottal half-width. There, points P_1 to P_5 represent the five points computed in the previous stage. The control time series were used to generate simulations of oral airflow, and the optimal values of the three control parameters at points P_2 to P_5 were recomputed sequentially. Values at P_1 were left as computed in the previous stage, because before this point the control parameters have constant values, as in the previous stage. After this point, they must be recomputed, because now the parameters change with time (according to the control time series). These computations were performed using the same simplex method and cost function as in the initial stage, and using the values determined in that stage as initial values. The control time series were updated each time the values of the control parameters were changed, and new simulations were generated using the updated time series.

The capability of this algorithm for finding correct control parameter values was verified using oral airflow records generated by the same models, in both male and female configurations. Using the default termination tolerances provided in the MATLAB scripts, the algorithm computed the control parameters with errors less than 0.5%, relative to their standard values.

V. RESULTS

Figures 9 and 10 show simulation results for the data in Figs. 1 and 2, respectively. The vertical solid lines mark the target region, as in Figs. 1 and 2. In the case of fully voiced /h/, as in Fig. 9, a good fit of the model was possible and the simulated signals reproduced well the targeted data. In the devoiced /h/, as in Fig. 10, there is some discrepancy in the restart of the oscillation. Comparing with Fig. 2, we may note that the ac amplitude grows slowly after the /h/ peak, causing a delay in the rebuilding of the ac component. As a consequence, the offset–onset hysteresis gap is even larger in the simulated results. This delay was typical of the simulations in the devoiced /h/ cases, and was also noted by McGowan *et al.* (1995) in their simulations. In their work, the restart of the oscillation was facilitated by setting an overshoot for the glottal width at the end of the vocal fold adduction. Here, none of the results for that parameter showed such an overshoot. The delay seems to be a shortcoming of the vocal fold model used, which does not seem capable of a quicker restart of the vocal fold oscillation. Let us also note that the onset of the oscillation depends not only on the values of the control parameters, but also on the velocity of variation of those parameters (Lucero, 1999b): The smoother the variation, the smaller the gap between offset and onset thresholds. In the present case, the simulation results suggest that the variation at the abduction–adduction gesture is somewhat drastic for the model, which results in a large hysteresis effect.

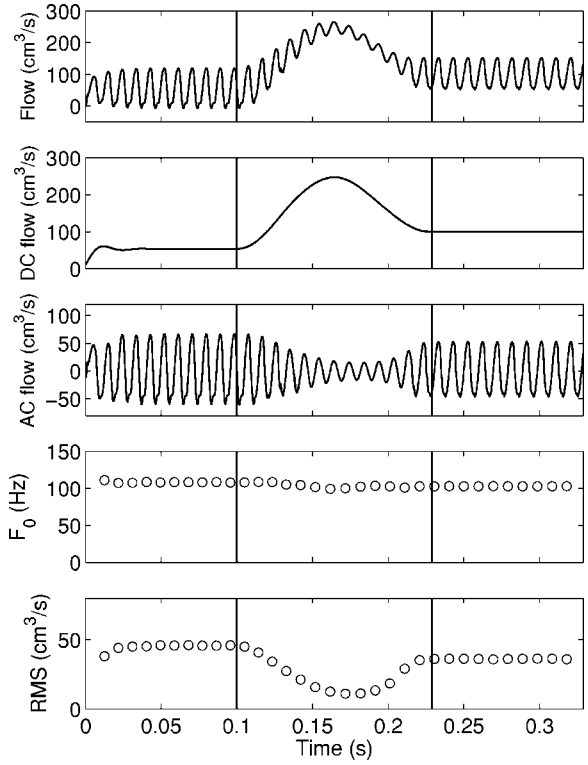


FIG. 9. Simulated airflow for data in Fig. 1. From top to bottom: airflow, dc component, ac component, fundamental frequency F_0, and the rms value of ac amplitude. The vertical lines mark the period that correspond to the simulation target.

Figure 11 shows the computed control parameters, for all the records of a male subject (the same subject as in Figs. 1 and 9). All other subjects presented qualitatively similar results, and so are not shown here. In general, all three parameters move in the direction of suppressing oscillation.

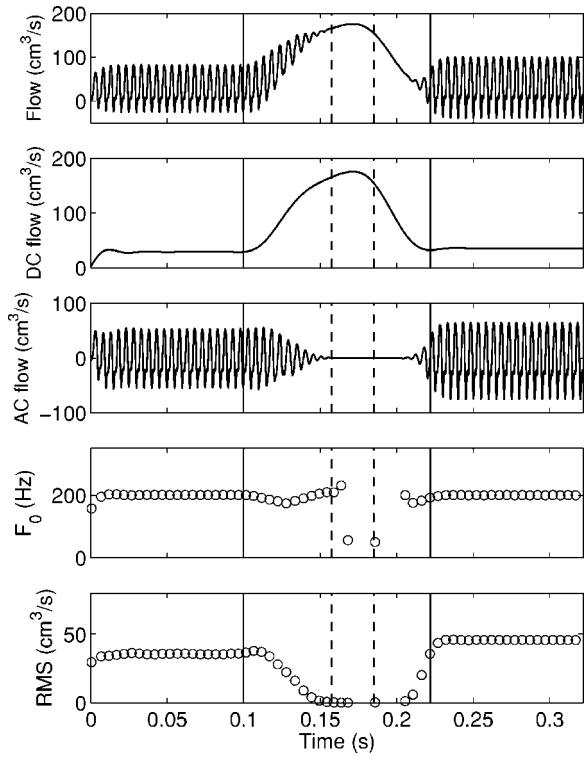


FIG. 10. Simulated airflow for data in Fig. 2. From top to bottom: airflow, dc component, ac component, fundamental frequency F_0, and rms value of ac amplitude. The vertical lines mark the period that correspond to the simulation target. The vertical dashed lines mark the devoiced period in Fig. 2.

The glottal half-width x_0 grows and then decreases, following a glottal abduction–adduction curve. At the same time, the subglottal pressure varies in the inverse direction. As discussed above (Sec. IV A), this reduction of glottal pressure at glottal abduction has been observed experimentally (Löfqvist, 1975; Ohala and Ohala, 1972), and is probably caused by a reduction of glottal resistance during abduction. The correlation of the increase in glottal half-width, computed as the difference between the peak abduction value and the base value, and the decrease of subglottal pressure, computed analogously, is 0.901 for this subject ($p=0.0038$), and 0.800 ($p<\text{Eps}=2.2\times10^{-16}$) for all subjects. There is also an increase of the Q factor. An increase in the Q factor may correspond both to active (muscularly controlled) changes in vocal-fold characteristics and to passive changes in the vocal-fold tissues, e.g., as they are stretched during abduction. Although we cannot separate these two effects based on the current data, the higher Q factor is at least consistent with past studies, which have argued that the cricothyroid muscle may be activated around abduction movements to help suppress the vocal fold oscillation (Löfqvist *et al.*, 1989; McGowan *et al.*, 1995). In the cases of devoicing (broken lines), a stronger movement of parameters may be seen so that the threshold level is reached (smaller subglottal pressure, larger glottal half-width, and/or larger Q value).

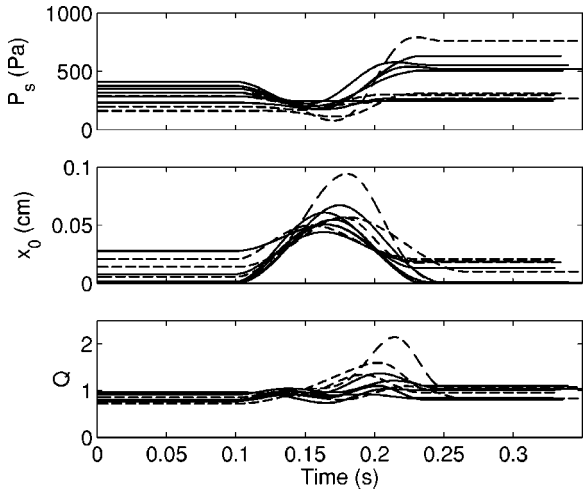


FIG. 11. Optimal control parameters for subject MJB (male). Top: subglottal pressure P_s. Middle: glottal half-width x_0. Bottom: Q factor. The full curves correspond to cases of fully voiced /h/, and the dashed curves to cases of devoiced /h/.

TABLE II. Summary of fitting results. The table shows mean values of airflow parameters, grouped in fully voiced /h/ and devoiced /h/ for each subject, computed from the simulated airflow as in Table I.

				DC (cm^3/s)	
Subject		F_0 (Hz)	AC (cm^3/s)	base	peak
		Female subjects			
FEM	voiced	207	42.5	77.8	203.5
	devoiced	194	26.5	74.4	232.5
FGR	voiced	189	35.8	57.5	159.4
	devoiced	200	40.5	64.2	346.9
FJH	voiced	200	26.8	33.4	123.1
	devoiced	200	32.7	30.7	153.3
FRS	voiced	236	96.3	203.0	493.9
	devoiced	249	111.6	151.9	726.2
		Male subjects			
MJB	voiced	120	50.7	64.8	291.5
	devoiced	115	34.4	77.0	288.0
MJW	voiced	144	148.2	150.8	824.9
	devoiced	143	102.4	89.7	571.9
MSFA	voiced	123	61.1	72.4	341.5
	devoiced	103	43.8	71.7	362.1
MSO	voiced	149	19.8	68.4	230.8
	devoiced	172.6	14.9	71.3	277.3

Table II summarizes fitting results for F_0, ac rms, and dc flow, for a comparison with the respective values in the recorded data shown in Table I. In general, a good fit could be achieved for F_0 and peak dc flow for all subjects. The fit was a bit poorer for ac rms flow and base dc flow. The worst case is the simulation for male subject MJW, which, according to Table I, combines large ac flow with comparatively low dc values at the base level. The large error of the results in Table II reveals a difficulty of the model in producing this particular combination of ac and dc flow.

Table III shows the computed control parameters. In general, the estimated subglottal pressures were higher for female subjects than for male subjects. The only exception is again the simulation for subject MJW, which has high values of subglottal pressures. The combination of large ac flow values with low dc values at the base level demands an almost zero glottal width at the base level, as seen in Table III, to obtain a small dc flow, combined with a large pressure, to obtain a large ac flow. Female subject FRS also has large subglottal pressures compared to the others. They are also a result of large flow values, as seen in Table I. Note that in

TABLE III. Summary of computed control parameters. The table shows mean values, grouped in fully voiced /h/ and devoiced /h/ for each subject.

		P_s (Pa)		x_0 (cm)		Q	
Subject		base	peak	base	peak	base	peak
		Female subjects					
FEM	voiced	593	450	0.010	0.037	1.15	1.39
	devoiced	429	294	0.013	0.053	1.11	1.93
FGR	voiced	432	374	0.010	0.031	1.03	1.12
	devoiced	499	416	0.010	0.065	1.13	1.09
FJH	voiced	419	291	0.004	0.028	1.05	1.16
	devoiced	491	282	0.002	0.035	1.02	0.99
FRS	voiced	1227	716	0.020	0.071	1.50	1.91
	devoiced	1239	608	0.013	0.120	1.50	1.86
		Male subjects					
MJB	voiced	393	228	0.005	0.054	0.89	0.89
	devoiced	337	177	0.011	0.062	0.91	1.15
MJW	voiced	1123	623	0.000	0.091	1.10	0.95
	devoiced	774	393	0.000	0.079	1.03	1.13
MSFA	voiced	423	342	0.006	0.051	0.91	0.88
	devoiced	273	211	0.012	0.069	0.78	0.64
MSO	voiced	287	249	0.011	0.041	1.20	1.13
	devoiced	355	221	0.010	0.054	1.39	1.44

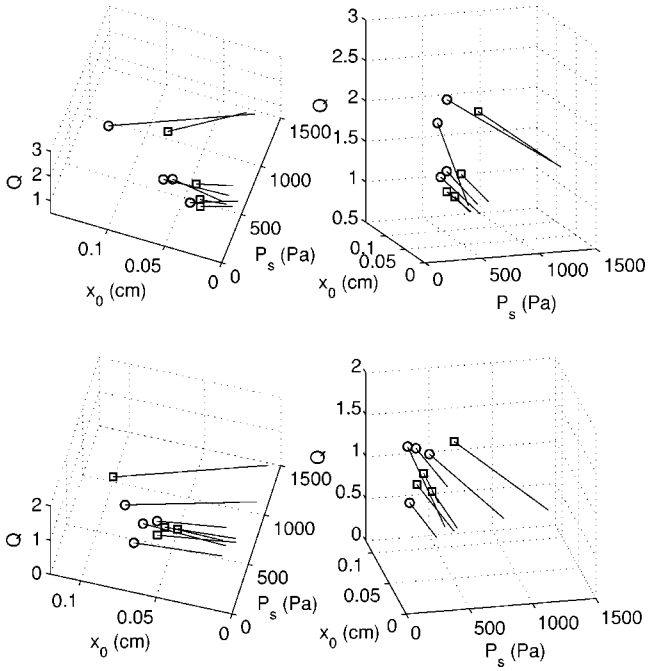


FIG. 12. Plots of mean values of control parameters for each subject, taken from Table III. The lines start at the average between starting and ending values of the parameters (point P_a in Fig. 8), and end at the values at the peak glottal abduction (point P_3 in Fig. 8). The circles indicate peak values for devoiced /h/, and the squares indicate peak values for voiced /h/. The top two plots are different views of the female results, and the bottom plots are different views of the male results.

this case the ac and dc (base) flow values are both large. Thus, the glottal width is also large, to reproduce that combination of flow values. Comparing peak glottal widths, men have, in general, larger values than women, with the exception of FRS, showing a larger vocal fold abduction. Q values are a bit higher in women, which, combined with a higher natural frequency, results in the greater fundamental frequency values for the female data in Table I.

Figure 12 show variations of the control parameters, from their values in Table III. The lines start at their base values and end at the peak abduction values. In general, the trajectories start at the regions of small x_0, high P_s, and small Q, and move toward high x_0, small P_s, and high Q. Trajectories that correspond to devoiced cases are longer than the voiced cases.

VI. CONCLUSION

In this paper we have shown a technique to fit a model of the larynx and vocal tract to speech data, which allows us to infer how the larynx is dynamically controlled in running speech. As noted in the Introduction, this is a more demanding test for the model, and might reveal aspects not seen in simulations using static configurations. In fact, it has shown the need for a nonlinear damping characteristic, to reproduce the ac flow pattern at glottal abduction.

The results show that the two-mass model plus two-tube approximation of the vocal tract, in spite of their simplicity, are generally capable of reproducing the speech data, with control parameters of physiologically realistic values and dynamical behavior. One aspect that could not be fitted with a good approximation was the restart of the oscillation after its offset, where the model had a slow behavior. The slower oscillation onset resulted in longer unvoiced periods and a larger hysteresis effect than observed in the data. Also, the model failed to reproduce airflow records combining large ac flow with low dc values.

The results also show that in both male and female speakers, devoicing at the abduction–adduction gesture for /h/ is achieved by the combined action of the three parameters: vocal fold abduction, decrease of subglottal pressure (which is an aerodynamic consequence of the glottal width increase), and an increase of vocal fold tension. Each of these actions has the effect of inhibiting the vocal fold oscillation, and even suppressing it when reaching an offset threshold. The simulation results also suggest variability across speakers in the factors that are manipulated during /h/; for example, some speakers show substantial variation in the Q factor during /h/, whereas others show a minimal change. This is consistent with our recent measurements, which indicate that individual speakers have unique strategies for achieving voicing and devoicing in running speech (Koenig and Lucero, 2004). Women have, in general, more restricted conditions for the vocal fold oscillation, which would explain the larger occurrence of devoicing in glottal abduction–adduction gestures, compared to men (Koenig, 2000). This restriction might be consequence of a smaller medial surface of the vocal folds, where energy is transferred from the flow to the fold tissues (Lucero, 2004). It also might be related to larger losses for air viscosity in the glottal airway, due to the smaller cross-sectional area, which would leave less energy to fuel the oscillation.

In general, fitting models to speech data is an important technique for making quantitative predictions of speech biomechanics, and has been used in previous studies (Döllinger *et al.*, 2002; Mergell *et al.*, 2000; Tao *et al.*, 2004; Trevisan *et al.*, 2001). Our technique may be extended to other instances of running speech to study voicing control under

different conditions. Such studies should also be accompanied by a deeper analysis of the stability and convergence of the numerical algorithm, in order to determine is the range of application to more general speech signals. Since the fitting algorithm uses ac rms and F_0 of individual ac cycles as target parameters, it may be inappropriate for cases of disordered voices that have extreme degrees of cycle-to-cycle variability. In such cases, algorithms with other target parameters should be used (e.g., Mergell *et al.*, 2000).

Even if we limit ourselves to speakers with normal voices, this work has an applicability to other research questions. Oral airflow signals are relatively easy to obtain, and they provide a noninvasive means of assessing laryngeal behavior. This is particularly important when studying special populations such as children. Modeling can refine our estimates of parameters that are difficult to measure in living speakers and tissues. Two examples are the subglottal pressure decrease during abducted conditions and the tissue damping. In future work, we plan to extend these methods to investigate the development of the larynx and phonatory control.

ACKNOWLEDGMENTS

This work was supported by Grant No. DC-00865 from the National Institute on Deafness and Other Communication Disorders of the National Institutes of Health (USA), CNPq and Finatec (Brazil).

Alipour-Haghighi, F., and Titze, I. R. (**1985**). ''Viscoelastic modeling of canine vocalis muscle in relaxation,'' J. Acoust. Soc. Am. **78**, 1939–1943.

Bilston, L. E., Liu, Z. Z., and Phan-Thien, N. (**2001**). ''Large strain behavior of brain tissue in shear: Some experimental data and differential constitutive model,'' Biorheology **38**, 335–345.

Chan, R. W., and Titze, I. R. (**2000**). ''Viscoelastic shear properties of human vocal fold mucosa: Theoretical characterization based on constitutive modeling,'' J. Acoust. Soc. Am. **107**, 565–580.

de Vries, M. P., Schutte, H. K., Veldman, A. E. P., and Verkerke, G. J. (**2002**). ''Glottal flow through a two-mass model: Comparison of Navier Stokes solutions with simplified models,'' J. Acoust. Soc. Am. **111**, 1847–1853.

Döllinger, M., Hoppe, U., Hettlich, F., Lohscheller, J., Schuberth, S., and Eysholdt, U. (**2002**). ''Vibration parameter extraction from endoscopic image series of the vocal folds,'' IEEE Trans. Biomed. Eng. **49**, 773–781.

Flanagan, J. L. (**1972**). *Speech Analysis, Synthesis, and Perception* (Springer-Verlag, New York).

Gielen, C. C. A. M., and Houk, J. C. (**1984**). ''Nonlinear viscosity of human wrist,'' J. Neurophysiol. **52**, 553–569.

Goldstein, U. (**1980**). ''An articulatory model for the vocal tracts of growing children,'' Doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Ishizaka, K., and Flanagan, J. L. (**1972**). ''Synthesis of voiced sounds from a two-mass model of the vocal folds,'' Bell Syst. Tech. J. **51**, 1233–1268.

Kahane, J. C. (**1982**). ''Growth of the human prepubertal and pubertal larynx,'' J. Speech Hear. Res. **25**, 446–455.

Kincaid, D., and Cheney, W. (**2002**). *Numerical Analysis: Mathematics of Scientific Computing*, 3rd ed. (Brooks/Cole, Pacific Grove, CA), pp. 722–723.

Koenig, L. L. (**2000**). ''Laryngeal factors in voiceless consonant production in men, women, and 5-year-olds,'' J. Speech Lang. Hear. Res. **43**, 1211–1228.

Koenig, L. L., and Lucero, J. C. (**2002**). ''Oral–laryngeal control patterns for fricatives in 5-year olds and adults,'' *Proceedings of the 7th International Conference on Spoken Language Processing*, pp. 49–52.

Koenig, L. L., and Lucero, J. C. (**2004**). ''Measurements of voicing offset and onset in men and women: further data,'' *Abstracts of the 4th International Conference on Voice Physiology and Biomechanics*, p. 93.

Koizumi, T., Taniguchi, S., and Hiromitsu, S. (**1987**). ''Two-mass models of the vocal cords for natural sounding voice synthesis,'' J. Acoust. Soc. Am. **82**, 1179–1192.

Lin, D. C., and Rymer, W. Z. (**2001**). ''Damping actions of the neuromuscular system with inertial loads: Human flexor pollicis longus muscle,'' J. Neurophysiol. **85**, 1059–1066.

Löfqvist, A. (**1975**). ''A study of subglottal pressure during the production of Swedish stops,'' J. Phonetics **3**, 175–189.

Löfqvist, A., Baer, T., McGarr, N. S., and Seider Story, R. (**1989**). ''The cricothyroid muscle in voicing control,'' J. Acoust. Soc. Am. **85**, 1314–1321.

Löfqvist, A., Koenig, L. L., and McGowan, R. S. (**1995**). ''Vocal tract aerodynamics in /aCa/ utterances: Measurements,'' Speech Commun. **16**, 49–66.

Lous, N. J. C., Hofmans, G. C. J., Veldhuis, R. N. J., and Hirschberg, A. (**1998**). ''A symmetrical two-mass vocal-fold model coupled to vocal tract and trachea, with application to prosthesis design,'' Acta Acust. (Beijing) **84**, 1135–1150.

Lucero, J. C. (**1999a**). ''A theoretical study of the hysteresis phenomenon at vocal fold oscillation onset–offset,'' J. Acoust. Soc. Am. **105**, 423–431.

Lucero, J. C. (**1999b**). ''Bifurcations at voice onset–offset,'' J. Acoust. Soc. Am. **105**, 1161.

Lucero, J. C. (**2004**). ''Dynamics of a two-mass model of the vocal folds for men, women, and children,'' *Abstracts of the 4th International Conference on Voice Physiology and Biomechanics*, pp. 191–195.

Manuel, S. Y., and Stevens, K. N. (**1989**). ''Acoustic properties of /h/,'' J. Acoust. Soc. Am. **86**, S49.

McGowan, R. S., Koenig, L. L., and Löfqvist, A. (**1995**). ''Vocal tract aerodynamics in /aCa/ utterances: Simulations,'' Speech Commun. **16**, 67–88.

Mergell, P., Herzel, H., and Titze, I. R. (**2000**). ''Irregular vocal fold vibration—High-speed observation and modeling,'' J. Acoust. Soc. Am. **108**, 2996–3002.

Nasseri, S., Bilston, L. E., and Phan-Thien, N. (**2002**). ''Viscoelastic properties of pig kidney in shear, experimental results and modeling,'' Rheol. Acta **41**, 180–192.

Ohala, M., and Ohala, J. (**1972**). ''The problem of aspiration in Hindi phonetics,'' Annual Bulletin, Research Institute of Logopedics and Phoniatrics, University of Tokyo, Vol. 6, pp. 39–46.

Pelorson, X., Hirschberg, A., van Hassel, R. R., Wijnands, A. P. J., and Auregan, Y. (**1994**). ''Theoretical and experimental study of quasisteady-flow separation within the glottis during phonation. Application to a modified two-mass model,'' J. Acoust. Soc. Am. **96**, 3416–3431.

Pelorson, X., Hirschberg, A., Wijnands, A. P. J., and Bailliet, H. (**1995**). ''Description of the flow through in-vitro models of the glottis during phonation,'' Acta Acust. (Beijing) **3**, 191–202.

Pitermann, M., and Munhall, K. G. (**2001**). ''An inverse dynamics approach to face animation,'' J. Acoust. Soc. Am. **110**, 1570–1580.

Tao, C., Zhang, Y., Du, G., and Jiang, J. J. (**2004**). ''Estimating model parameters by chaos synchronization,'' Phys. Rev. E **69**, 036204.

Thompson, J. M. T., and Stewart, H. B. (**1986**). *Nonlinear Dynamics and Chaos* (Wiley, New York), pp. 108–131.

Titze, I. R. (**1988**). ''The physics of small-amplitude oscillation of the vocal folds,'' J. Acoust. Soc. Am. **83**, 1536–1552.

Titze, I. R. (**1989**). ''Physiologic and acoustic differences between male and female voices,'' J. Acoust. Soc. Am. **85**, 1699–1707.

Titze, I. R. (**1994**). *Principles of Voice Production* (Prentice–Hall, Englewood Cliffs, NJ).

Titze, I. R., and Liang, H. (**1993**). ''Comparison of F0 extraction methods for high-precision voice perturbation measurements,'' J. Speech Hear. Res. **36**, 1120–1133.

Trevisan, M. A., Eguia, M. C., and Mindlin, G. B. (**2001**). ''Nonlinear aspects of analysis and synthesis of speech time series data,'' Phys. Rev. E **63**, 026216.

Measurements of glottal structure dynamics

John F. Holzrichter[a)]
Lawrence Livermore National Laboratory, Livermore, Californiaand Department of Applied Science, UC Davis, Davis, California

Lawrence C. Ng, Gerry J. Burke, Nathan J. Champagne II, Jeffrey S. Kallman, and Robert M. Sharpe
Lawrence Livermore National Laboratory, Livermore, California

James B. Kobler and Robert E. Hillman
Voice and Speech Laboratory, Massachusetts Eye and Ear Infirmary Boston, Massachusetts

John J. Rosowski
Easton Peabody Laboratory, Massachusetts Eye and Ear Infirmary, Boston, Massachusetts

(Received 6 September 2004; revised 22 October 2004; accepted 29 October 2004)

Low power, radarlike electromagnetic (EM) wave sensors, operating in a homodyne interferometric mode, are being used to measure tissue motions in the human vocal tract during speech. However, when these and similar sensors are used in front of the laryngeal region during voiced speech, there remains an uncertainty regarding the contributions to the sensor signal from vocal fold movements versus those from pressure induced trachea-wall movements. Several signal-source hypotheses are tested by performing experiments with a subject who had undergone tracheostomy, and who still was able to phonate when her stoma was covered (e.g., with a plastic plate). Laser-doppler motion-measurements of the subject's posterior trachea show small tissue movements, about 15 microns, that do not contribute significantly to signals from presently used EM sensors. However, signals from the anterior wall do contribute. EM sensor and air-pressure measurements, together with 3-D EM wave simulations, show that EM sensors measure movements of the vocal folds very well. The simulations show a surprisingly effective guiding of EM waves across the vocal fold membrane, which, upon glottal opening, are interrupted and reflected. These measurements are important for EM sensor applications to speech signal de-noising, vocoding, speech recognition, and diagnostics. [DOI: 10.1121/1.1842775]

PACS numbers: 43.70.Aj, 43.70.Jt [AL] Pages: 1373–1385

I. INTRODUCTION

The use of electromagnetic (EM) waves for the measurement of human organ positions and motions has been limited because of their coarse transverse resolution which is approximately the wavelength in human tissue, e.g., 2–3 cm for 2-Ghz waves. In addition, there is often a longitudinal positional ambiguity regarding the reflection location due to the many-wave long EM wave trains transmitted by interferometric sensors. Impulse EM sensors are being developed to resolve some of these problems, but they are expensive, complicated, and not considered in this paper. However, for identified interfaces, requiring noninvasive data, interferometric sensors measure the longitudinal positions to parts in 10^4 giving unparalleled spatial resolution (i.e., microns) and temporal information (i.e., 10^{-4} s).

Because radarlike EM sensors are becoming so economical and so compact, they are being considered for special applications such as real time monitoring of heart motions, breathing, and vocal articulators (Holzrichter *et al.*, 1998). In this paper, it is shown that vocal fold and vocal tract wall movements can be easily measured, leading to many potentially useful applications in speech signal processing, especially for cellular phones and other portable systems.

Radarlike sensors transmitting electromagnetic (EM) waves have been used to measure properties of the human vocal system during speech production (Holzrichter, 1995; Holzrichter *et al.*, 1998). These sensors function by transmitting very low power (<0.3 mW) EM waves, at high frequencies, i.e., 2.3 Ghz for the experiments herein [based on measurements by D. Poland and E. T. Rosenbury, Lawrence Livermore Lab (LLNL)]. A multi-cycle wave train is directed into the neck or head toward the vocal articulators whose movements are being monitored (see Fig. 1). The absorption coefficients at 1–4 GHz EM waves in human tissue are 5 to 10 cm^{-1}, typically allowing 10-cm penetration into the body and back to the sensor [Gabriel *et al.* (1996) and animal tissue measurements by L. Haddad, LLNL]. EM waves reflect from all dielectric and conductivity discontinuities, associated with tissue-air or tissue-tissue interfaces, and which are in the path of the propagating EM waves. An antenna and receiver detect the reflected EM waves using a homodyne (i.e., interferometric) technique (Skolnik, 1990), upon which electronic circuits average, high pass filter (e.g., AC couple), and amplify the detected signal for subsequent analysis. Sensor signals, associated with the movement of vocal folds and related air pressure induced tissue motions, occur at pitch

a)Electronic mail: holzrichter1@llnl.gov

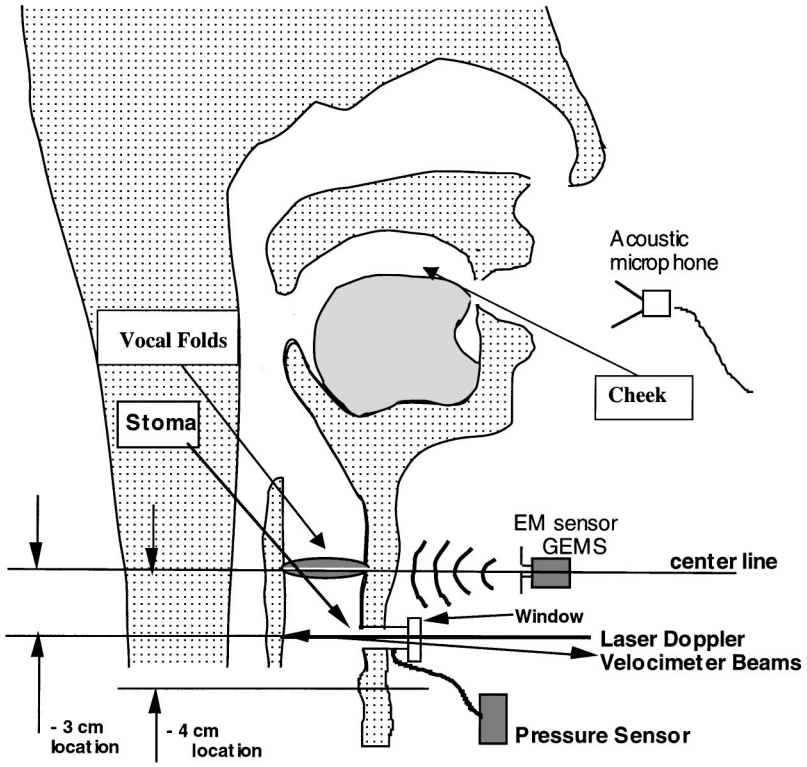


FIG. 1. Experimental layout showing position of the laser, the GEMS experimental location, pressure sensor, and acoustic microphone. EGG electrodes are not shown but are mounted normally. Three locations for EM sensor placements are noted as prominence, subglottal, and cheek. These locations are outside the body, against the neck skin surface.

frequencies nominally between 70 and >250 Hz. These are easily distinguishable from signals returning from stationary tissue interfaces by employing suitable frequency band-pass filters. The sensor used for the experiments herein is called a ''GEMS sensor'' (McEwan, 1994; Burnett, 1999; Holzrichter and Burnett, 1999) because it is optimized for glottal electromagnetic sensing.

The signal amplitude of a homodyne sensor is proportional both to a change in distance of the targeted tissue interface from the sensor and to the reflecting locations on the tissue interface (i.e., the scattering cross-sections, see Skolnik, 1990). Hence there has been an ambiguity in separating targets with large frontal areas but small position changes from those targets that have small frontal areas, but relatively large positional changes. In the case of the larynx region, measurement ambiguities occur if subglottal tracheal walls (with areas of cm^2) were moving 50–100 μ caused by air pressure changes as the vocal folds open and close. In contrast, the area of the frontal view of the glottal opening is small (with an area of mm^2), but the opening distance can be cm in length, perhaps together causing a signal similar to that from the tracheal walls. In addition, when the GEMS sensor is placed in front of the larynx as voiced speech is produced, its signals resemble those from an electro-glottograph (EGG) sensor (see Figs. 2 and 3) (Titze *et al.*, 2000).

The purpose of this paper is twofold: first to identify the

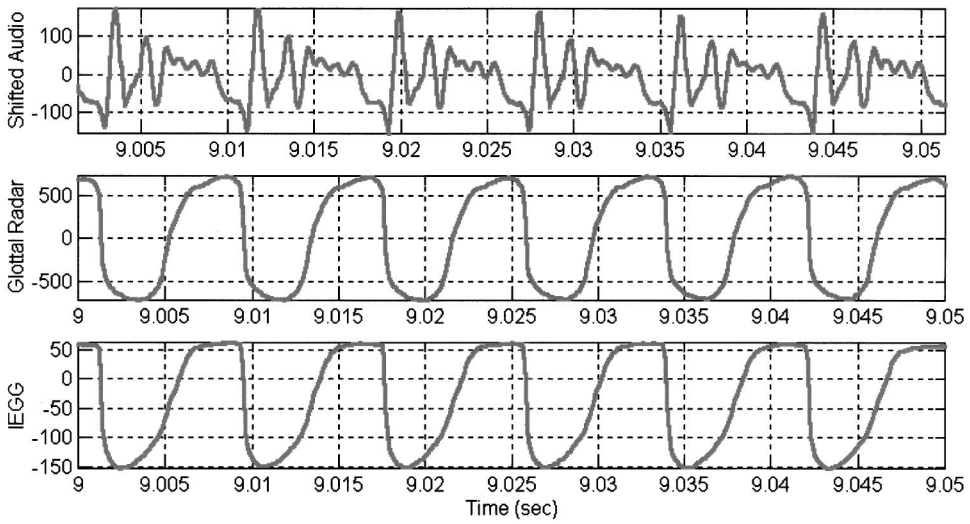


FIG. 2. Typical data showing simultaneous acoustic, GEMS sensor (i.e., glottal radar trace), and inverse EGG signals for a male speaker. Closure times are aligned. The EM sensor signals show strong similarities to those from the EGG sensors.

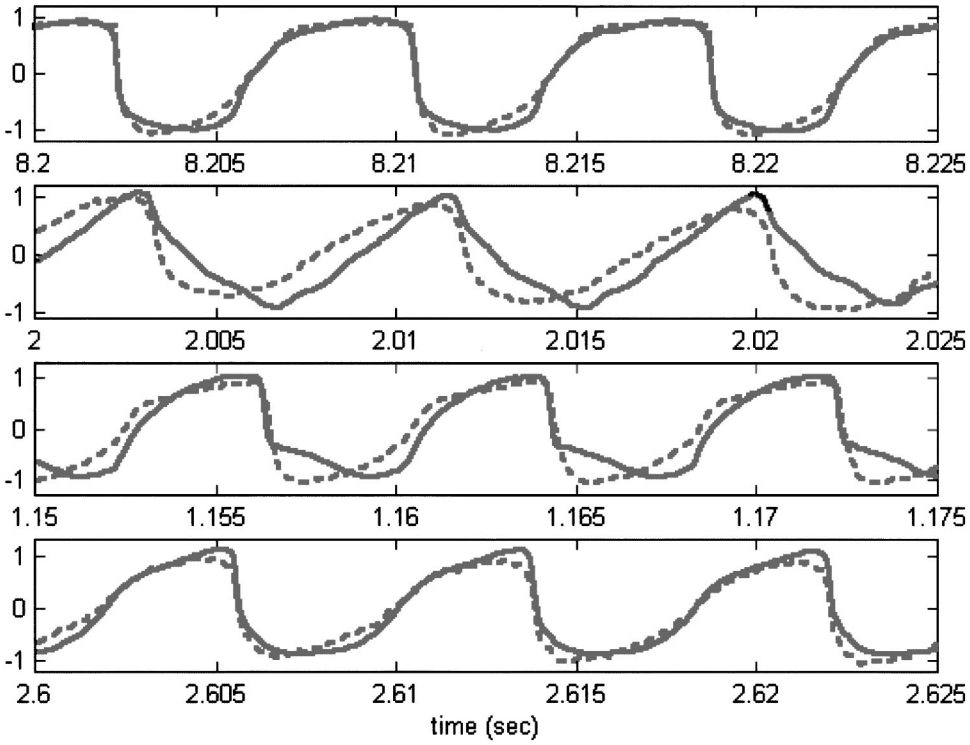


FIG. 3. Data from four separate male subjects comparing GEMS sensor data (solid) to EGG data (dashed). Closure times are not aligned. The idiosyncrasies between subjects and the differences between the EGG and the EM sensors are also apparent.

sources of EM sensor signals when used in the laryngeal area, and second to explain the shape of these EM sensor signals and their similarities to EGG signals. It will be shown that signals from the larynx area are due to vocal fold reflections, but signals from trachea and oral cavity walls can also be easily measured and used. Three sources of reflected EM waves from the laryngeal region are considered: (1) reflections from the changing shape of the frontal view of glottal opening, (2) reflections from tracheal-wall air-interface motions as they ''balloon'' due to air pressure excursions, and (3) reflections from changes in conductivity as the folds separate (i.e., EGG-like). The tests used herein to reach this paper's conclusions are that simulated EM wave reflections from these three targeted tissues must simulate an EM sensor signal whose intensity versus time resembles the peak amplitude and the shape versus time of the measured EM sensor signals, as seen in Figs. 2–5.

The data and interpretations in this paper provide a great deal of information on EM wave propagation and reflection from tissues internal to the neck and head, and by extension from other articulators in and connected to the vocal tract. This understanding is important because of the potential for GEMS and other EM sensors to provide a wide range of useful, real-time information on speech type and onset, on pitch and pitch periods, and for estimates of voiced excitation functions by associating the glottal signal with air flow excitation information. These are being applied to several applications in denoising (Ng *et al.*, 2000; Aliph, 2000), for narrow band vocoding, and other applications.

II. EXPERIMENTS AND DISCUSSION

The fortunate identification of an unusual subject at the Massachusetts Eye and Ear Infirmary in Boston enables subglottal pressure and posterior tracheal wall motion measurements to be made that are instrumental in resolving EM wave reflection ambiguities in the larynx region. The subject is a 58-year-old female who had undergone a tracheostomy 5 years ago as a consequence of treatment for laryngeal paresis following thyroplasty with implantation of a silicone prosthesis. Her paresis was partially resolved at the time of the recording, and the subject's voice, when occluding the stoma, is only slightly breathy. The experiments described below are approved by the human studies IRBs of the Massachusetts Eye and Ear Infirmary and the Lawrence Livermore National Laboratory. While the unique subject enabled unique posterior tracheal wall measurements, EM sensor and microphone signals are consistent among several dozen subjects tested to date including the subject described above.

A. Experimental approach

A laser Doppler velocimeter (HLV-1000, Polytec PI, CA) is used to measure the movements of the subject's posterior tracheal wall by transmitting a laser beam through a transparent plastic cover, located over her stoma valve opening. Simultaneously the GEMS EM sensor is placed in one of two locations, depending upon the experiment. To test vocal-fold reflectivity, it is placed slightly below the thyroid-prominence on the mid-sagittal plane. To measure tracheal wall movements, it is placed below the stoma opening, about 4 cm below the prominence, also in the mid-sagittal plane. The skin is marked to provide repeatable locations for sequences of measurements. In all experiments, the EM sensor's antenna-cover (1 mm plastic) is placed against the neck tissue to prevent skin vibration. It is then gently pushed inward or outward to maximize the signal. In addition, data from an EGG sensor (Glottal Enterprises MC2-1), a subglot-

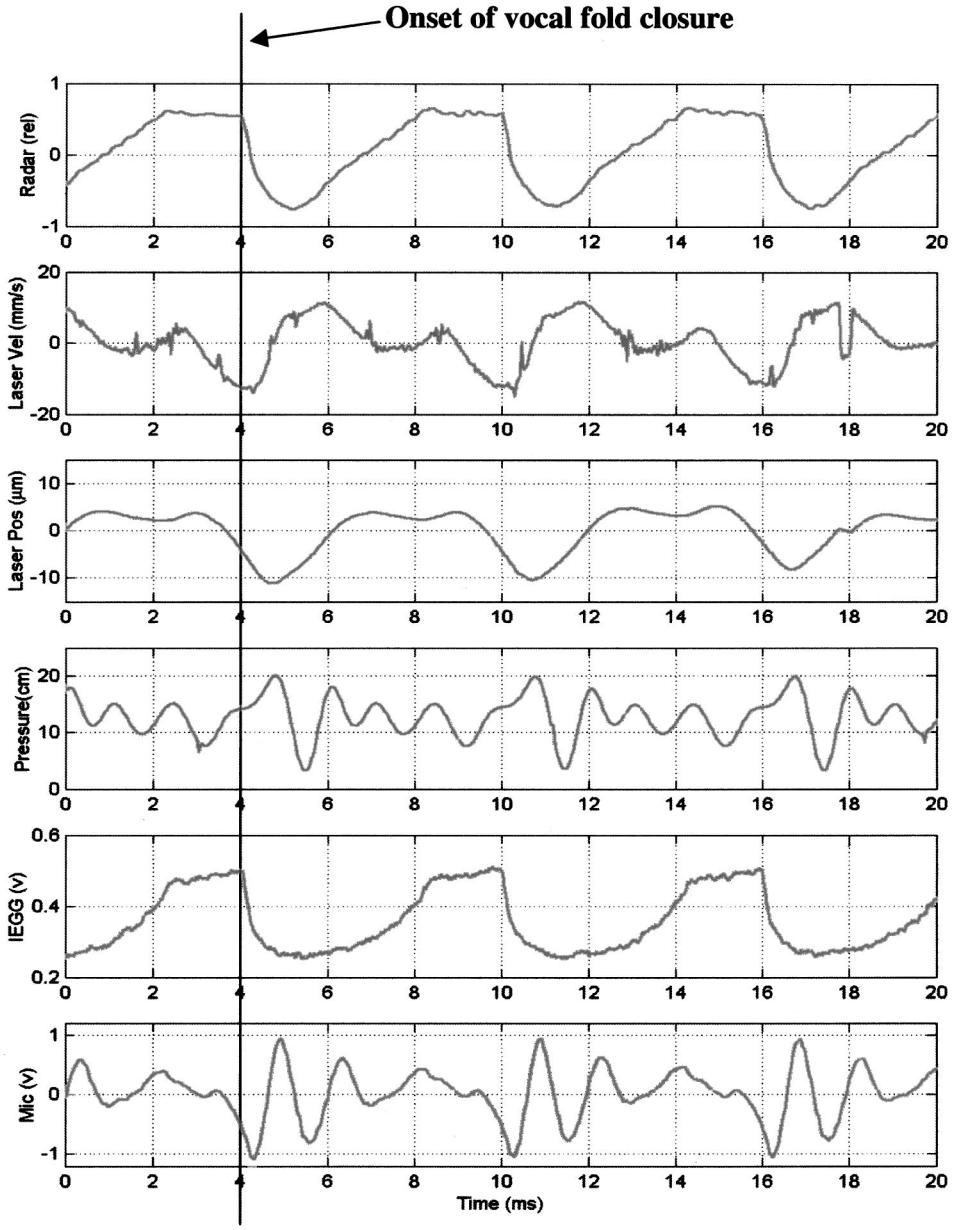


FIG. 4. Data set from subject with partial paresis but phonating ''ah'' very clearly. The vertical line, across all five panels, shows the time of onset of vocal fold closure. (a) Top panel: GEMS sensor signal, positive value indicating increased vocal structure reflectivity. (b) Second panel: Laser velocity signal from posterior tracheal wall, positive signal for tissue movement away from laser (i.e., posterior direction). (c) Third panel: Position versus time of posterior wall, positive signal indicating ''ballooning'' as pressure increases. (d) Fourth panel: Subglottal pressure versus time in cm H_2O. (e) Inverted EGG signal. (f) Fifth panel: Acoustic signal from microphone 10 cm in front of subject's mouth.

tal pressure sensor, and an acoustic microphone (placed about 10 cm in front of the lips) are also recorded, as described below. Details and calibrations are discussed in more detail elsewhere (Holzrichter *et al.*, 2003).

The subject habitually wears a size 6 tracheostomy tube with a Montgomery Tracheostomy Speaking Valve. For the experimental measurements, the speaking valve membrane was removed and a side opening in the valve body was drilled, to which a 6-cm-long side tube, 3.0-mm i.d., was attached. The tube was connected to an SMI-5552-015-D pressure transducer (Silicon Microstructures Inc., Fremont, CA). To enable phonation, the open end of the modified speaking valve is covered with a transparent plastic window through which the laser was aimed at the posterior wall of the trachea. The plastic window is heated by immersion in warm water prior to placements to minimize fogging. Most of the measurements are made with the subject reclining in a supine position.

The laser instrument has been used for many physiological structure measurements in the past, including making measurements through windows and through narrow openings (Rosowski *et al.*, 1999). Its internal velocity calibration is quite accurate and using a given calibration of mm/s/v, the data is converted to velocity in mm/s, an example of which is presented in Fig. 4(b). The velocity data is numerically integrated to obtain the change in position of the targeted tissue versus time [see Figs. 4(c) and 5(d)]. For example, in a 1-ms

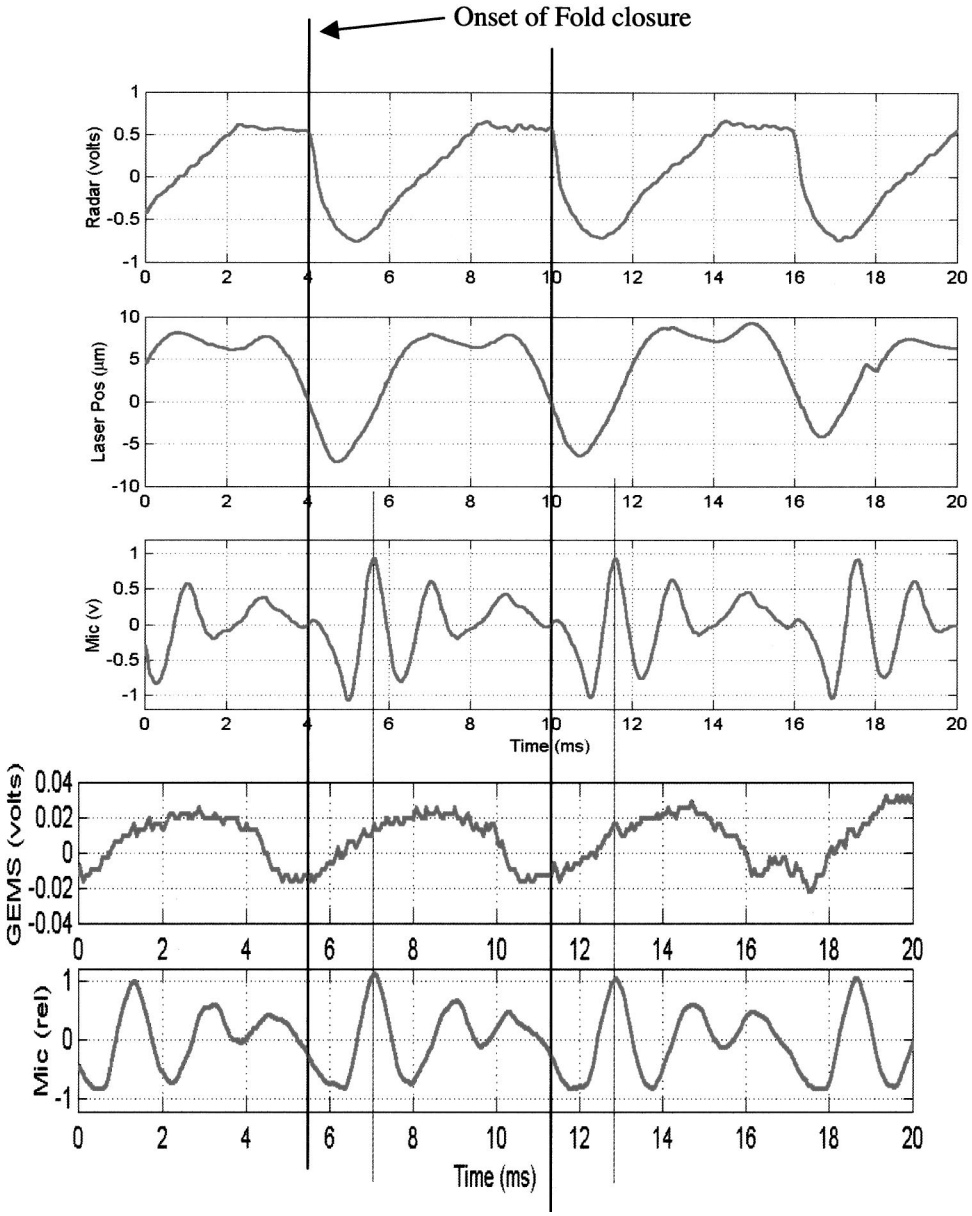


FIG. 5. Composite GEMS sensor data for the same speaker and same sound ''ah.'' Traces 4(a)–(c) are from the laryngeal area, as shown in the previous figure. Lower trace, 5(d), shows GEMS signals from the subglottal area, when it was located 4 cm below the prominence. Trace 5(e) shows simultaneous acoustic data. These two sets of data, taken on different dates, were time aligned using the acoustic signals.

time interval, a velocity of 5 mm/s leads to a movement of 5 μm.

The GEMS sensor's response is calibrated using an aluminum target, 1 cm wide by 2 cm high, affixed to a mechanical vibration generator and moved at 200 Hz (Bruel and Kjaer vibration generator with exciter type 4802 and head type 4817). The sensor response is found to be consistent with data obtained earlier by Burnett (1999). These procedures give a GEMS signal calibration in air of about 5 mV/μ for a target vibrating about 4 cm from the sensor.

In the case of a vibrating air/tissue interface, such as tissue vibrations from a surface located about 1 cm inside the skin (equivalent to about 5 cm of air path) the calibration is about 2.5 mV/μm. This calibration is obtained by measuring the motion of the inside surface of a subject's cheek with the laser-Doppler instrument, as a subject phonates and pronounces the sound ''m.'' This signal is then compared to the corresponding EM sensor, located 1 cm away, on the cheek skin surface. Extending the cheek data to the similar geometry of the anterior air-wall interface of the trachea results in a similar calibration of 2.5 mV/μm, or about 35 mV, for expected 15 μm tissue excursions.

B. Experimental discussion

Measurements were taken on four separate occasions over the period of $1\frac{1}{2}$ years. A very large amount of data was taken, with Fig. 4 illustrating one of the clearer and more complete sets of data obtained. In this experiment, five sensors are used while the subject phonates the sound ''ah'' as in ''father.'' The data traces, from top to bottom in Fig. 4, are labeled as 4(a)–(f), and are discussed below:

Figure 4(a) shows the raw GEMS sensor output at a location just below the laryngeal prominence. This data is

consistent with previous GEMS and EGG data such as that shown in Figs. 2 and 3, and elsewhere, and with the inverted EGG data (i.e., IEGG) shown below in Fig. 4(e). Positive signal direction indicates increased EM wave reflectivity. As discussed in Sec. III, this signal is caused by an increased amplitude of reflection, mostly from the front edges of the glottal opening, as it opens from a completely closed condition. In addition, there is a reduction in the effective EM wave path and corresponding phase from the sensor to the point of reflection and back (i.e., the point of EM wave scattering moves closer to the sensor as the glottis opens).

Figure 4(b) shows the velocity of the subject's posterior tracheal wall, measured 3 cm below the laryngeal prominence. At the 3-ms time, the velocity becomes negative as the glottis opens and the pressure drops. At the time of vocal fold closure, at the 4-ms time, the negative wall velocity stops increasing (in the negative direction toward the laser) and begins to slow to zero. At the 5-ms time, the posterior wall begins to move away from the laser, i.e., it ''balloons,'' and it expands as the glottis closes and the subglottal air pressure rises. Nominal velocity noise, lasting typically 0.1 ms, indicates short periods of scattered laser light, due to scintillation.

Figure 4(c) shows the numerical integration of the tracheal wall velocity trace [from Fig. 4(b)] gives the posterior tracheal wall position versus time. A positive value indicates that the posterior wall is ''ballooning'' in a posterior direction away from the laser sensor. The amplitude of the movement is approximately 12 μm, which is much lower than earlier estimated values used to identify the source of GEMS signals [Burnett (1999) based on intra-oral, tongue data by Svirsky *et al.* (1997)], but consistent with other estimates discussed below. This tissue movement is insufficient to generate the nominal 1 V GEMS signals shown in Fig. 4(a). Also, the signal amplitude and shape, corresponding to the wall movement, show little correlation with the signals shown in traces 4(a) and (e), and thus it cannot be the dominant source of the GEMS signal from the laryngeal location.

Figure 4(d) shows the subglottal air pressure versus time shows pressure variations of 7 to 20 cm of H_2O versus time, with peak pressure excursions ranging from 5 to 15 cm H_2O. This results in a time averaged pressure change that moves the trachea walls of about 5 cm H_2O. The signal also illustrates distinct subglottal resonances (Ishizaka *et al.*, 1976, Cranen and Boves, 1985; Fredberg and Hoenig, 1978). The data from trace 4(c) indicates that the tracheal wall is responding slowly and out of phase to the rapid subglottal pressure signal resonances, with about a 1.5–2-ms time constant. This is expected from a membrane being driven above its natural resonances (about 30–40 Hz).

Figure 4(e) shows the inverted EGG signal is especially useful for timing the onset of vocal fold closure and providing information on the quality of the vocal fold contact area versus time. The EGG signal is inverted to make it consistent with GEMS data from previous papers, e.g., Figs. 2, 3, and 11. This IEGG data follows the GEMS data closely for these and previous experiments (Titze *et al.*, 2000), and is used to inspect data sets to identify those with normal vocal fold closure (e.g., Fig. 4). In cases when this subject (and others) phonated with incomplete closure, the IEGG signal vanishes, yet the GEMS signal shows oscillatory signals associated with oscillating, but noncontacting, folds.

Figure 4(f) shows the acoustic microphone signal shows the characteristic initial negative pressure signal as the folds close. The acoustic signal is advanced by 0.7 ms to correct for the slower traveling acoustic wave, from the vocal folds to the microphone located 10 cm from the lips.

Another set of experiments [see Figs. 5(d) and (e)] are presented to illustrate the differences between GEMS measured lower trachea wall movement and data from the larynx location in Fig. 4. These are taken with the GEMS positioned 4 cm below the laryngeal prominence (about 1 cm below the stoma, see Fig. 1). Because of the spatial ''congestion'' of instruments within 1 cm of the stoma, and due to some rf-interference, it is not possible to arrange for all of the instruments to be operative during the subglottal experiments. The data from two separate sessions are time aligned by aligning the acoustic-peak signatures. There are three important differences between the subglottal data shown in Figs. 5(d) and (e) and the laryngeal prominence data of Fig. 4. The first is that the sub-stoma GEMS signal is about 30-fold lower in amplitude, about 40 mV peak to peak in Fig. 5(d), than the 1.2-V prominence signal in Fig. 5(a). Using the GEMS calibration described above (2.5 mV/μm), the 40-mV signal indicates that the source is likely to be the anterior wall of the trachea, vibrating with a 10–20-μm amplitude. Second, the shape is consistent with a pressure-induced expansion and contraction, and shows no distinct ''sharp'' cutoff of the type associated with vocal fold closure. Third, the timing is consistent with the laser-measured posterior-wall data of Fig. 5(b), except that it precedes it in time by about 1 ms. This timing offset may be due to the quite different types of trachea-wall tissues being measured by the two instruments. Similar subglottal experiments have been conducted on male subjects at the Livermore Laboratory over the past years (in accord with earlier IRBs), which also show similar timing of the subglottal signal relative to the vocal fold closure signature. They also show similar relative signal amplitudes and show similar signal shape. In summary, the GEMS subglottal data are completely consistent with EM wave reflections from the anterior tracheal wall ''ballooning'' due to changes in subglottal pressure.

III. NUMERICAL SIMULATIONS

Two- and three-dimensional numerical simulations were used to estimate the reflections of EM waves from the posterior trachea wall and from the vocal fold structures. It became clear that while 2-D calculations were useful for the investigation of EM waves interacting with cylindrical geometry, they did not enable an estimate of the effect of the vocal fold membrane and the glottal opening. For the 3-D simulation, see Fig. 6(a), several ''snapshots'' of relatively simple geometries are chosen for the tracheal tube and the glottal membranes. The sum of reflections of the incoming EM wave, as it scatters from all the parts of the vocal fold machinery, is illustrated in Fig. 7. A simplified circular glottal hole is chosen to enable relatively simple zoning. The simplified, smooth dielectric structures, with zero conductiv-

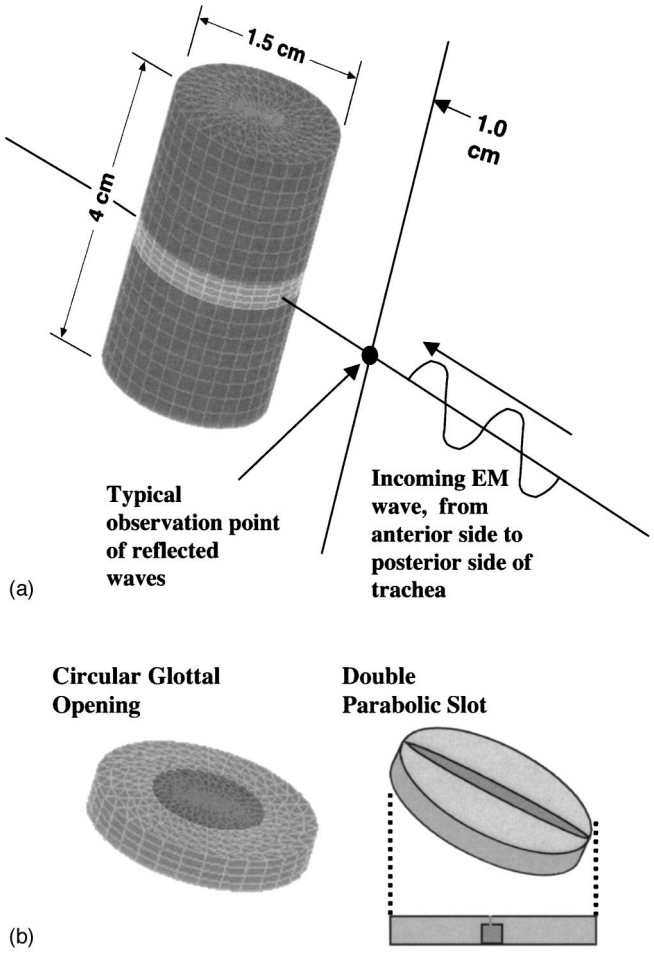


FIG. 6. (a) Zoning of a 4-mm-thick vocal fold membrane (shown as darker ring in the center of tube) and the 1.5-cm-diam trachea tube for computer simulations of EM waves propagating into the vocal fold structures. The vocal fold membrane zones are simulated as a 4-mm-thick dielectric disk, $\epsilon=25$ with no conductivity, in the mid-transverse plane of a 4-cm-long, ''can-shaped'' air tube. The entire ''can'' structure is surrounded by an infinite dielectric material, $\epsilon=25$ and zero conductivity. The simulation is conducted with a time-stationary, surface element code based on the method of moments. Time behavior is calculated by computing ''snap shots'' of reflections as the glottal shapes are changed. (b) Two types of modeled glottal shapes are shown: a circular opening in the center and a double-parabolic slot that extends from the anterior to the posterior sides of the membrane. A side view of the parabolic slot also shows a zoned side view of an incompletely opened glottis, as a phonation cycle begins. Data presented in figures in this paper were simulated using fully developed glottal openings, extending smoothly from the inferior to the superior sides of the vocal fold membrane.

ity, fixed glottal structures, and time integrated reflections, enable the use of a surface element code based on the ''method-of-moments'' (Miller *et al.*, 1992). For this survey, it is easier to use than using other electro-magnetic codes based on finite elements or finite differences, which are superior for more complex geometries and for time-dependent behavior.

Once a few circular openings are understood, a few cases where the glottis is slot-shaped are examined [see also Fig. 6(b)]. These slot-shaped openings require more complicated zoning, especially at the closure points. The total reflections are given in Fig. 10, with amplitude and phase information provided in more detail elsewhere (Holzrichter *et al.*, 2003). For improved simulations, the simulated slots should be shorter in length and the zoning should capture detailed time behavior of one or more observed modes of vocal fold opening and closing, as the margins open from the inferior side, ultimately to the superior margins, and then close.

A. 3-D simulation approach

A ''method of moments'' code, called EIGER (Sharpe *et al.*, 1997), was chosen for the survey calculations because of its availability, its reliability, and the presence of computers able to deal with the relatively large number of computations needed for applications of this procedure (Miller *et al.*, 1992). This code was used to analyze the propagation of an EM wave internal to an infinite block of neck tissue. The electrical properties of the layers of muscle and cartilage were approximated by an average dielectric constant, $\epsilon_{\text{average}}=25$ and zero conductivity (no loss in the tissues). Figure 6 illustrates the shape of the modeled larynx region, coded using 1500 nodes. (In some cases, such as slot calculations, more extensive zoning is used.) For the method of moments, the trachea air tube requires proper boundary conditions which are satisfied by using ''end caps'' to keep the simulation volume finite in extent, and by extending the surrounding dielectric material out to infinity. Contributions to the E field at the point of measurement, from EM wavelets scattering from the caps, were tested to be sure that they did not contribute significant signals at the measuring point. In addition, the zoning of the membrane, with its various holes and slots, was varied to be certain that the amplitudes and phases of the simulated EM wave reflections were not caused by numerical artifacts.

A 2.3-GHz, EM plane-wave-train, y-polarized (in the plane of the membrane), with an amplitude of 75 V/M was ''launched'' from one side of the simulation (i.e., from within the dielectric coming from the right side in Fig. 6). The wavelength in the dielectric is $\lambda=2.6$ cm. The EM wave propagated in the dielectric medium to the front surface of the 1.5-cm-diam trachea (i.e., from the lower right of Fig. 6 into the trachea section, which is 1.5 cm in diameter by 4 cm long). The EM wave then propagated (i.e., evanesced) around the tube. Normally an EM wave cannot enter an air cavity with a dimension smaller than the wavelength in the surrounding high dielectric-constant media [Fig. 7(a)]. However, as the vocal folds are adducted into the trachea tube, the tissue membrane ''waveguides'' the EM wave across the air tube, causing the EM wave to follow the high dielectric-constant material [Figs. 7(b)–(d)]. At the same time the EM wave was propagating forward into the tube, it was also being reflected by the front and back walls of the tube, and by glottal holes or slots in the membrane. Figure 7 shows four examples of the scattered component of the EM wave, where it is being forward-scattered across the membrane and back-scattered to the GEMS sensor. The numerical simulations also enable an estimate of the reflection of an incoming EM wave from a nominal glottal opening back to the sensor to be about 0.5%.

The Eiger code's numerical diagnostics enable simulations of both amplitude and phase of the scattered incoming EM wave at a desired observing point. This point is located

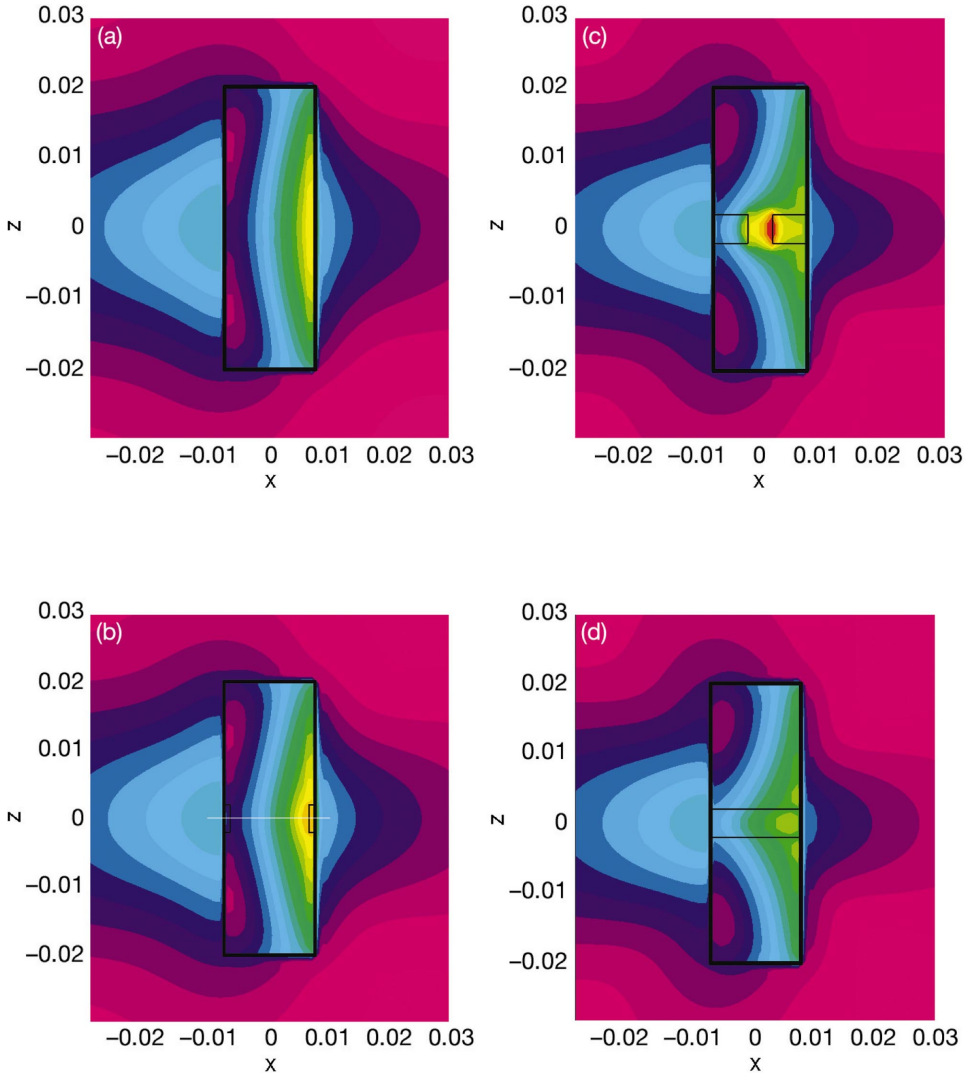


FIG. 7. Four examples of simulations showing the amplitudes of reflection of incoming EM waves (from right to left) onto four different tracheal tube sections, each with a different diameter circular glottis. The four cases, (a)–(d), have corresponding glottal diameters of 1.5-, 1.2-, 0.4-, and 0.0-cm diameter, respectively. In cases (b)–(d) the outline of the cross section of vocal fold membrane is in black. The white line in insert (b) shows the axis of ''line-out'' data showing scattered field-amplitude and field-phase versus position (shown in following figures). From ''line-out'' data the color code follows: red at the edge of the hole in (c) corresponds to a field strength of 180 V/M, yellow 150 V/M, green 130 V/M, light blue 75 V/M, and the dark-purple color, on the right of the tubes, 40 V/M.

1 cm in front of the anterior wall of the trachea in the mid-transverse plane (on the right side of the images, on the side of the incoming EM wave). This is commonly the place in the experimental configuration where the EM sensor receiver antenna would be located on the skin on the laryngeal prominence. While this diagnostic point is useful for survey work, a real antenna extends several cm laterally, measuring many locations where the signal can change significantly from the on-axis value. The numerical data was used primarily in the differential mode, where the input wave is held constant and changes in vocal fold configurations cause phase and/or amplitude differences in the scattered EM waves, which are compared, one to another.

B. Simulation results

Many numerical and empirical experiments show that the sensitivity of the EM sensor signal, corresponding to positional and directional variations of ± 1 cm and $\pm 10°$ respectively, is small. This phenomenon occurs because the EM wave is about 2–3 cm wide inside the neck tissue, allowing it to be easily ''captured'' by the vocal fold membrane and guided across the trachea. The images shown in Fig. 7 also show color contours of scattering EM wave amplitudes versus distance in the mid-sagittal plane. Figure 7(a) illustrates EM scattering from the tube with completely retracted folds, showing how the EM wave weakly penetrates the anterior tube surface, but strongly reflects backward from it. In the intermediate cases, Figs. 7(b) and (c), red and yellow colors, indicate that the EM wave is ''carried into'' the tracheal tube along the membrane, and scatter from the posterior edges of the glottal opening. Figure 7(d) illustrates how the EM wave is carried across the tube by the 4-mm-thick dielectric membrane with a low level of reflection.

Figures 8 and 9 show the amplitudes and phases of the reflected EM wave at a continuum of measurement locations along a path from the trachea center to beyond the measure-

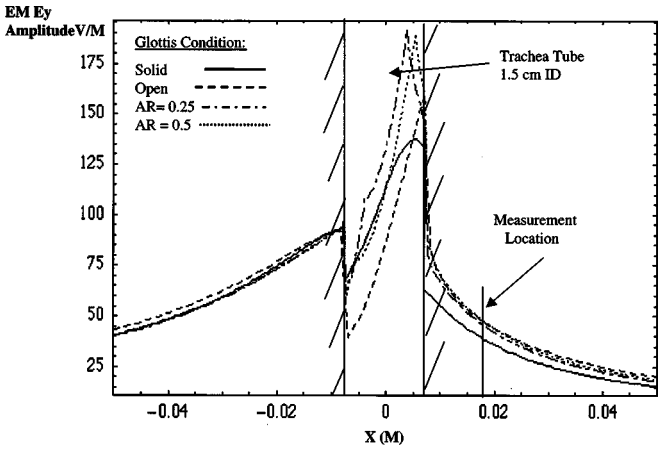


FIG. 8. Amplitude of the forward and backward scattered EM waves, both inside the trachea tube and outside the tube, for four different circular glottal openings. These four curves follow the scattered EM field amplitude along the mid-sagittal center line of the calculations. They give the data ''line-out'' along the white line in the previous figure [in Fig. 7(b)]. At the noted ''measurement location,'' the intensity of the reflected EM wave depends upon the structure of the reflections in the glottal membrane, as well as surrounding structural elements. The solid line is for a closed glottis in a 4-mm-thick membrane, the dot-dashed line is for a 0.75-cm hole (area ratio $=0.25$), and the dotted line is for a hole diameter of 1.06 cm (area ratio $=0.5$), dashed line is completely open. Signals are interpreted as differences between changing and stationary reflections.

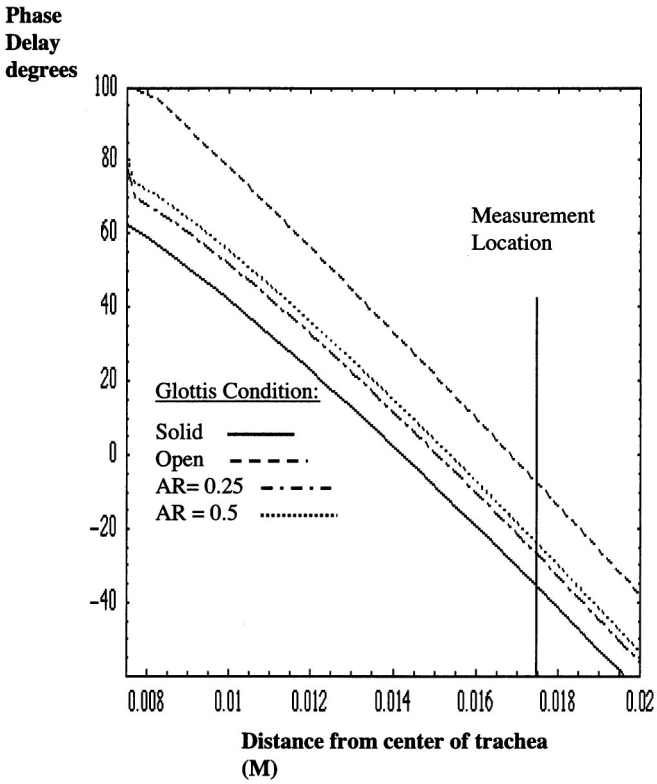


FIG. 9. Phase of the forward and backward scattered EM waves, in degrees, versus distance and versus glottal condition, measured from the trachea tube center to the ''measuring point.'' The left ordinate line is positioned at the anterior trachea wall. The phase increases (becomes more negative) when the glottis closes and ''draws'' the EM wave further inside the trachea tube.

ment point of the hypothetical experiments [see white line in Fig. 7(b) illustrating the path]. This point is located 1 cm in front of the tube (at the 0.0175 M location) and is designated by an arrow on the graphs. The glottal model, used to simulate the results reported here, consists of a simple expanding and contracting circular hole in a 4-mm-thick membrane ($e_{\mathrm{Ave}}=25$), starting with a solid membrane with area ratio$=0$. The area ratio, AR, is given by the ratio of glottal-area/trachea-area$=$area ratio. Next a hole with relative area ratio AR$=\frac{1}{4}$ is simulated, and finally a hole with an area ratio AR$=\frac{1}{2}$. The condition noted with the dashed line is the empty tube. The circular glottis model also enables simulations of empty-trachea wall reflections, which occur when the glottal diameter becomes equal to the trachea diameter (i.e., 1.5 cm in these examples). In the model, as the glottal opening is increased in the simulation, the glottal membrane is thickened to represent the volume of tissue that moves to the sides as the glottal space opens. This approach conserves tissue molecules in the path of the EM wave.

Figure 8 shows that the reflected field amplitude, on axis, ranges from 42 V/M for closed folds, 45 V/M with AR$=0.25$, 47 V/M with AR$=0.5$, to 48 V/M for an open tube trachea. For this example, at the measurement location, the reflection amplitudes change by 12%–15% from the closed glottal condition to the open condition. The background amplitude of 42 V/M is the scattered amplitude from all the modeled reflecting structures, such as the tracheal walls, the end caps, and the membrane-wall interfaces. Since these are constant, they do not show up in high-pass filtered signals.

Figure 9 shows the reflected field phase versus distance, due to glottal scattering condition changes. The measurement location is noted. The convention in the ''Eiger'' code is that the phase is increasingly negative as the distance increases from the effective source of reflection to the measuring point. For example, at the noted measurement location, the phase is increasingly negative as the glottis closes and the wave continues to the far side of the trachea tube. This means that a fraction of the EM wave is carried deeper into the trachea tube as the glottis closes, thus increasing the path of the wave. The phase changes for the four glottal conditions shown are $-36°$, $-27°$, $-25°$, and $-8°$ for the empty tube, giving a maximum phase change of 28°, or a 33% phase change within a 90° quadrant of phase. As is summarized in Fig. 10, the phase change from a slot opening behaves differently than that of an expanding circle because the front edge of the slot scatters most of the energy and defines the reflection location, thus fixing the phase. The phase changes little as the slot widens because waves reflected from the sides do not come back to the measuring point.

C. Comparison of sensor signals to glottal area

The GEMS sensor signal is estimated by taking into account the output of the signal from a homodyne receiver detector (Skolnik, 1990) as shown in Eq. (1):

$$S_{\mathrm{mixer}}=\mathrm{const}*A*\cos\phi. \tag{1}$$

A is the received signal amplitude and ϕ is the phase change as the tissues moves. For maximum sensitivity, this phase is usually referenced to the tissue rest position, at which point the phase is adjusted to be $\pi/2$ (Burnett, 1999). Hence the ''AC-coupled'' sensor signal, which is function of both the

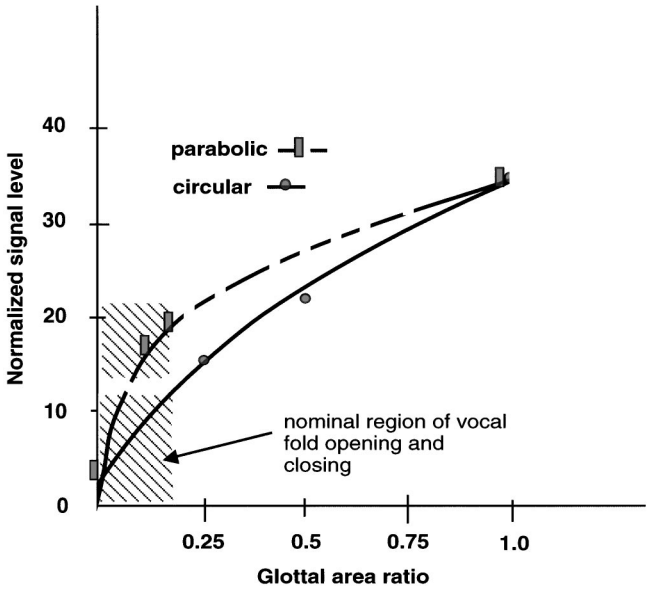


FIG. 10. Smoothed and normalized EM sensor signal data from simulations of reflected EM waves from circular glottal openings and from parabolic openings versus AR. AR is the area ratio showing the relative area of the glottal opening to the tracheal cross-sectional area. The data takes into account both the field amplitude and phase change as the glottis opens and closes. The zero signal point corresponds to the reflected signal of 42 V/M, from all reflections in the simulation. Similar, stationary reflections occur when an EM sensor is placed against the neck or other feature of a user.

phase change and the amplitude of reflection, is estimated by comparing the signal from a final configuration to that of an initial configuration.

Figure 10 shows relative, on axis sensor signal levels for two types of glottal openings, a parabolic slot and a circular shape [see Fig. 6(b) for the geometries]. The data is plotted as signal versus area ratio (i.e., where the area ratio, AR, is the area of glottal opening divided by tracheal tube area). This presentation is intended to show to what degree glottal air flow, proportional to glottal area (Stevens, 2000, p. 65), can be related to the EM sensor signal. This approximation, that the vocal tract excitation source is proportional to the EM sensor signal which is, in turn, proportional to glottal area, is being used for signal processing applications (Ng *et al.*, 2000). The signals measured by the EM sensor are the modulated signals from the moving tissues, which pass the filter bands, and which are superimposed on top of a non-modulated background (e.g., 42 V/M in the example above). The circular glottal opening begins as a small hole in the middle of the membrane, scattering a small amount of the incoming EM wave. As it increases in diameter the reflection increases, and, at AR=0.2, its signal modulation is (52−42)/52=0.21.

For a slot, the opening extends across the diameter of the tube [see Fig. 6(b)] from the anterior to the posterior walls of the trachea, and its area is related to the width of the parabolic edges. Simulations show that upon the slot opening, the EM wave reflects strongly from the anterior point of the slot shaped opening, showing an amplitude and phase jump. As the slot widens further, Fig. 10 shows that the rate of reflectivity increase versus area is slower than for the circular opening. At an AR value of 0.2, the parabolic slot modulates the reflected EM sensor by over 30% [e.g., signal modulation is (62−42)/62=0.33]. In the parabolic case for the small glottal area regime (e.g., $0<AR<0.1$), a signal proportional to glottal area, and thus to glottal air flow (Stevens, 2000, p. 65), could be constructed from data in Fig. 10. However, more detailed calculations are needed.

Actual glottal shapes resemble a combination of the two model structures discussed here, but with several additional complexities. For most persons, the glottis is shorter in length and does not extend across the tracheal tube. Upon opening, the slot evolves to an elliptical shape (e.g., see Fig. 11). In addition, the glottis opens as the inferior margins of the vocal folds slowly open, and then close relatively rapidly, mucosal motions take place that also reflect EM waves, incomplete posterior closure is common, and many other phenomena occur (Rothenberg, 1981; Titze, 1984, 1990). These types of varying signal sources will be discussed below in conjunction with Fig. 11.

IV. DISCUSSION OF DATA AND SIMULATIONS

The simulations in Figs. 8–10 enable a calibration of the GEMS sensor's signal as a function of the modeled changes in phase and amplitude of the reflecting EM wave from reflecting tissues. First, it is estimated that for each unit of EM wave amplitude transmitted toward the glottis from the anterior skin surface of the neck, about 0.005 (i.e., 0.5%) is reflected back to the sensor's antenna, or in other words the total received power is about −42 dB. Typically 1%–10% of this reflected signal is modulated by moving tissue interfaces. Next, in order to compare calculations to measured signals, on a relative basis, consider the hypothetical sensor, located just below the laryngeal prominence at the ''measuring point'' [see Fig. 6(a)]. For a nominal glottal area ratio change, e.g., from 0.0 to 0.20 in Fig. 10, the EM sensor signal is about 1 V [see Fig. 4(a)]. This signal is caused mostly by a phase change of about 14°, causing a 25% signal change [as shown in Eq. (1)], but it is also due to a small reflected amplitude change of about 8% (circular model, Figs. 8 and 9). For comparison to calculated signals from tracheal wall tissues whose signal is mostly due to small phase changes, the GEMS sensor phase sensitivity is estimated by Eq. (2):

$$C_{\text{glottal}} = 1 \text{ V}/14^\circ \text{ phase change} = 0.07 \text{ V}/^\circ. \quad (2)$$

A. Tracheal wall results

EM waves reflected from both the posterior wall and the anterior wall are considered. These are important for applications because they represent other vocal tract wall movements, including the pharynx, oral cavity, and sinuses, that can be measured and used for signal processing applications. The direct measurement of the subglottal, posterior tracheal wall tissue with the laser velocimeter shows 12 μm movements (see Fig. 4). The opposing anterior tracheal wall movement was not initially considered (Burnett, 1999) because of its expected stiffness. However, a direct examination of a human trachea, excised from a cadaver, showed that the stiff ''C'' shaped trachea cartilage structural elements occupy about 30% of the anterior tracheal wall area. The re-

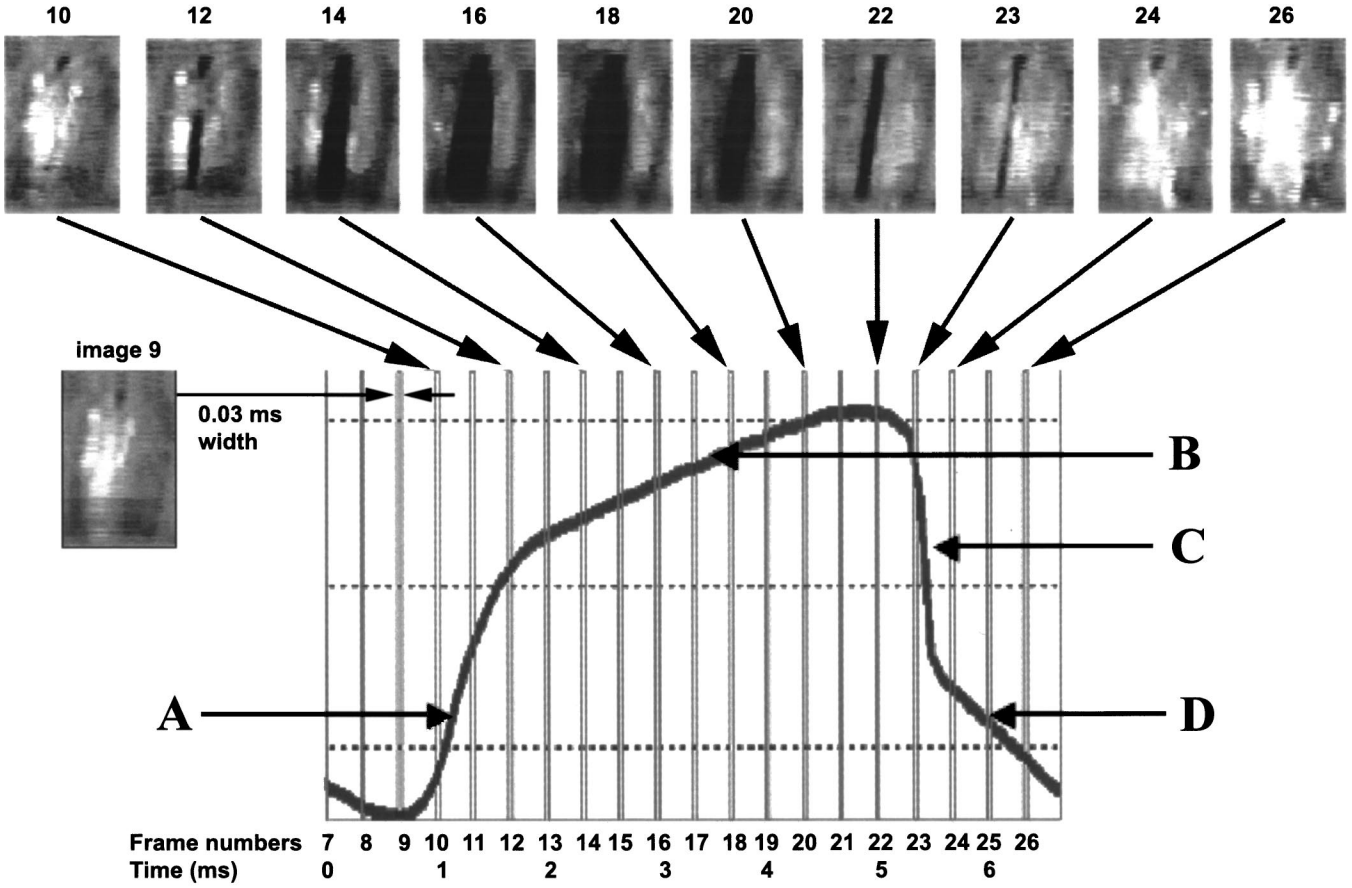


FIG. 11. Typical GEMS signal from a male subject with corresponding vocal fold images. Location A shows rapid sensor signal increase as the inferior edges of the fold open increasing the EM wave reflectivity more rapidly than linear from a paraboliclike slot; location B shows signal increasing nominally with area increase as in an elliptical-circular shape; location C shows rapid signal drop, probably due to strong anterior fold contact; and location D shows continued sensor signal drop, after the visual image shows fold-closure, indicating continued vocal fold tissue rearrangement in preparation for the next cycle.

maining 70% of the wall area, between the cartilage elements, was found to be similar to other vocal tract wall tissue, and would move similarly to other wall tissues, as the subglottal pressure oscillates.

1. Anterior trachea wall

Using the calculations shown in Fig. 9, for an open tube, an estimate of the tracheal wall phase delays due to small wall movements can be made by moving the numerical observation point by a small distance (see Holzrichter *et al.*, 2003, for details). The anterior tracheal wall calibration is shown in Eq. (3):

$$C_{\text{anterior}} = (0.44°/10\ \mu\text{m})*(1\ \text{V}/14°) = 0.030\ \text{V}/10\ \mu\text{m}. \quad (3)$$

The GEMS signal from 4 cm below the prominence, Fig. 5(d), is 35 mV, which upon using the calibration of 30 mV/10 μ [Eq. (3)], yields a simulated movement of 12 μm.

Earlier, the calibration of a signal for a similar geometry was discussed, that of the inner cheek wall, which was measured using a laser velocimeter. This calibration gave 2.5 mV/μm of inner cheek surface motion for a 5 cm H_2O intraoral pressure change. Using the measured 30–40 mV GEMS subglottal signal [see Figs. 5(d) and (e)], and the cheek-based calibration of 2.5 mV/μm, an anterior trachea inner-wall surface movement of 14 μm was measured. This agreement between the simulated anterior wall signal of 12 μm and the similar cheek motion of 14 μm is sufficiently close that it is concluded that a GEMS-like sensor, when placed at the −4-cm subglottal location as in Figs. 5(d) and (e), measures movement of the anterior tracheal wall.

2. Posterior tracheal wall

The posterior wall induced changes in the GEMS signal can be simulated in a fashion similar to those just used to estimate the anterior wall signal, and can be compared to directly measured signals as shown in Fig. 4(c). Equation (4) shows its calibration:

$$C_{\text{posterior}} = 0.018°/10\ \mu m*1\ \text{V}/14° = 0.0013\ \text{V}/10\ \mu m. \quad (4)$$

Using the laser-doppler system, 13 μm of posterior wall motion were measured [Fig. 4(c)]. Using the calibration just above in Eq. (4), the GEMS signal from posterior tracheal wall motion would be about 1.7 mV. This signal is substantially smaller than the GEMS measured signals from both the laryngeal and subglottal region, indicating that the posterior trachea wall motion does not contribute significantly to the GEMS signals in Figs. 1–5.

Because of the small measured motions of the posterior tracheal wall, which conflicts with some earlier estimates of air pressure induced intraoral tissue measurements that were extended to tracheal tissue responses (Burnett, 1999), some additional estimates of the tracheal wall motions using mechanics formulas are made. A standard thick-walled pipe formula (e.g., Roark and Young, 1975) was used with an inner radius $a=0.75$ cm, and an approximate outer radius $b=2.5$ cm. Then the change in the inner radius due to pressurization is given in Eq. (5):

$$\Delta a = (\Delta \text{ pressure}) * (a/E) * [(b^2 + a^2)/(b^2 - a^2) + \upsilon]. \tag{5}$$

The radius a is 0.75 cm, the Δ pressure is 5 cm H_2O, Young's modulus E for smooth muscle is estimated to be about 190 psi (Fung, 1993), and Poisson's ratio is $\upsilon=0.2$. This calculation leads to a cyclic wall deflection of about 5 μm, due to air pressure excursions of about 5 cm H_2O. A second calculation, modeling the posterior trachea wall as a compressible plate, using a unit strain formula of $\epsilon=\sigma/E$, yields a movement of approximately 10 μm per 2-cm-thick wall for a 5 cm H_2O pressure. In addition, a tissue compliance (i.e., spring like ''K'' values) can be estimated from the posterior tracheal wall measurements in Fig. 4 (i.e., 13 μm per 5 cm H_2O pressure change) which gives a value of about 30×10^5 dynes/cm. This measured value is, however, substantially stiffer than exterior side-neck tissue compliance data from Ishizaka *et al.*, 5×10^5 dynes/cm (1975). This Ishizaka neck data is, in turn, about tenfold stiffer than tongue tissue data presented by Svirsky *et al.* (1997), 0.5×10^5 dynes/cm, upon which early tracheal wall motions were estimated. In summary, the static stiffness estimates above, the computer simulations, Ishizaka's data, and the experiments in this paper show that the posterior tracheal wall movement is approximately 5–15 μm, which is difficult to detect using presently configured GEMS-like sensors.

V. COMPARISON OF EM SENSOR SIGNALS TO LARYNGOGRAPH IMAGES

It is useful to compare the superior view of the vocal fold surface tissues (Burnett, 1999; Burnett and Leonard, 1999) to the corresponding EM sensor signals. See Fig. 11 for an example of data. The descriptions are aided by vocal fold cycles as illustrated in Stevens (2000). It is interesting to note that at location ''A'' the EM signal is rising relatively rapidly as the inferior margins of the vocal folds ''peel'' open from the inferior side of the membrane [see Fig. 6(b) slot picture for a model example]. The increase in vertical slot opening causes a more rapid than linear increase in signal, estimated to be at least proportional to the square of the slot thickness (Born and Wolf, 1965), because EM waves reflect poorly from structures that are small compared to their wavelength and reflect strongly as the dimensions approach $\frac{1}{4}$ wavelength and larger. In addition, numerical simulations show that tissue conductivity is not needed to generate these rapidly changing signals, although it should be considered in future calculations. The signal in region B increases less rapidly than that in region A because the slot is already formed, and the targeted opening has become more circlelike (see Fig. 10). However, the images also show that closing begins at the 4-ms time marker, yet the EM sensor signal continues to rise. This remains puzzling, except that the EM sensor is measuring the changing configuration of all of the tissues in the laryngeal region, including the vocal fold mucosa, processes, and ligaments. [However, in the case of the subject whose EM sensor data is shown in Fig. 4(a), this pattern is not seen.] The rapid closure signal at location C is associated with vocal-fold contact, most likely with initial contact in the anterior region, where the contact location and extent most strongly influences the phase and amplitude of the EM wave reflection. The rapid closure signal in location C is often used for signal processing applications, because it provides accurate timing and because it empirically behaves well as an excitation function, generating most of the spectral energy in the voiced speech signal. The signal may be proportional to air-flow cut-off as discussed in Sec. III above, and thus is an adequate representation of a voiced excitation. The more slowly relaxing signal in region D indicates that for this subject (also see subject 3 in Fig. 3), complete closure of the folds and their rearrangement for the next cycle takes more time than normal. This example illustrates the sensitivity of radarlike EM sensors to many idiosyncratic factors of a subject's phonation, which can be used for speaker verification parameters (Gable, 2000). In addition, these EM sensors easily measure the state of folds during falsetto, breathy speech, and many other conditions where vocal fold contact does not take place. The EM sensor is sensitive to the size, shape, dielectric constant, modes of motion, and the time-dependent arrangements of the vocal folds and surrounding tissues, and not strongly sensitive to vocal fold contact conductivity.

VI. CONCLUSION

Based upon the experiments and simulations described herein, homodyne EM sensors can measure real-time motions of the vocal folds and vocal tract walls, such as those of the trachea, cheek, and other locations. However, the signals from these two general types of tissues are very different in both shape and amplitude. EM sensor signals, measured just below the laryngeal prominence, are due to reflections from an EM wave guided by the vocal fold membrane, which samples the changing configurations of the vocal folds. In contrast, EM sensor measurements of air pressure induced subglottal, tracheal wall movements, and intraoral cheek walls, are shown experimentally and analytically to be associated with a nominal air-pressure-induced 10–20 μm movements of the anterior tracheal walls.

Data from the EM sensors of the type illustrated here are usually combined with corresponding acoustic data for speech processing applications. For example, they enable robust, extremely accurate onset of voiced speech, pitch detection ($<$1 Hz accuracy), and an estimate of a voiced excitation function by using the spectral information associated with rapid fold closure (e.g., signal segment C in Fig. 11). They enable acoustic signal denoising of at least -20-dB reduction, depending on type of noise (Ng *et al.*, 2000; Aliph, 2000). They also appear to enable narrow bandwidth speech compression (200–400-Hz bandwidth), and experiments indicate that speaker verification is possible with error

rates less than 1:1000 (Gable, 2000). These relatively new EM sensors are becoming increasingly economical, use very little power, and can be built into a variety of instruments and commercial devices.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Greg Burnett for many early suggestions and measurements. They would also like to thank Dr. Rebecca Leonard, Dr. Wayne Lea, Professor Ingo Titze and Professor Brad Story for their continued interest and support. The work at LLNL was supported by the US Department of Energy, DARPA, and the National Science Foundation. Work at the Massachusetts Eye and Ear Infirmary was supported by NIH/NICDC.

Aliph (**2000**). ''Applications of EM sensor/acoustic speech technologies,'' Aliphcom Incorporated, San Francisco, CA. See examples at www.aliphcom/sound (best accessed using Microsoft Explorer).

Born, M., and Wolf, E. (**1965**). *Principles of Optics* (Pergamon, Oxford).

Burnett, G. C. (**1999**). ''The Physiological Basis of Glottal Electromagnetic Micropower Sensors (GEMS) and Their Use in Defining an Excitation Function for the Human Vocal Tract,'' thesis, UC Davis, 15 January 1999, available through ProQuest Digital Dissertations, document number 9925723.

Burnett, G. C., and Leonard, R. (**1999**). ''Use of Kodak Ektapro High-Speed DigitalCameras in Laryngoscopy,'' Phonoscope **2**(1), 33.

Cranen, B., and Boves, L. (**1985**). ''Pressure measurements during speech production using semiconductor miniature pressure transducers: Impact on models for speech production,'' J. Acoust. Soc. Am. **77**, 1543.

Fredberg, J. J., and Hoenig, A. (**1978**). ''Mechanical Responses of the Lungs at High Frequencies,'' J. Biomech. Eng. **100**, 57.

Fung, Y. C. (**1993**). *Biomechanics—Mechanical Properties of Living Tissues*, 2nd ed. (Springer, New York), p. 474.

Gable, T. J. (**2000**). ''Speaker Verification Using Acoustic and Glottal Electromagnetic Micro-power Sensor (GEMS) Data,'' thesis, December 2000, University of California at Davis, available through ProQuest Digital Dissertations, index number 9997362.

Gabriel, S., Lau, R. W., and Gabriel, C. (**1996**). ''The dielectric properties of biological tissues: III. Parametric models for the dielectric spectrum of tissues,'' Phys. Med. Biol. **41**, 2271–2293.

Holzrichter, J. F. (**1995**). ''New Ideas for Speech Recognition and Related Technologies,'' Lawrence Livermore National Laboratory Report, UCRL-UR-120310. Available from Lawrence Livermore National Laboratory library, or from NTIS in Springfield, VA at http://www.ntis.gov/ordering.htm

Holzrichter, J. F., and Burnett, G. C. (**1999**). ''Human Speech Articulator measurements using low power, 2 GHz Homodyne Sensors,'' 24th International Conference on Infrared and Millimeter Waves, edited by L. A. Lombardo, and C. A. Monterey, 5 September (Kluwer, Dordrecht); see also Lawrence Livermore Laboratory report UCRL-JC-134775.

Holzrichter, J. F., Burnett, G. C., Ng, L. C., and Lea, W. A. (**1998**). ''Speech articulator measurements using low power EM wave sensor,'' J. Acoust. Soc. Am. **103**, 622. Also see the website http://speech.llnl.gov/

Holzrichter, J. F., Kobler, J. B., Rosowski, J. J., and Burke, G. J. (**2003**). ''EM wave simulations and measurements of glottal structure dynamics—a Technical Report,'' UC Report UCRL-JC-147775, 2003. Available on the Website http://speech.llnl.gov/ or from Lawrence Livermore National Laboratory library, or from NTIS in Springfield, VA at http://www.ntis.gov/ordering.htm

Ishizaka, K., French, J. C., and Flanagan, J. L. (**1975**). ''Direct Determination of Vocal Tract Wall Impedance,'' IEEE Trans. Acoust., Speech, Signal Process. **ASSP-11**(4), 370.

Ishizaka, K., Matsudaira, M., and Kaneko, T. (**1976**). ''Input acoustic-impedance measurement of the subglottal system,'' J. Acoust. Soc. Am. **60**, 190.

McEwan, T. E. (**1994**). U.S. Patent Nos. 5,345,471, 5,361,070, and 5,573,012.

Miller, E. K., Medgyesi-Mitschang, L., and Newman, E. H. (eds.) (**1992**). *Computational Electromagnetics-Frequency-Domain Method of Moments* (IEEE, New York).

Ng, L. C., Burnett, G. C., Holzrichter, J. F., and Gable, T. J. (**2000**). ''Denoising of Human Speech Using Combined Acoustic and EM Sensor Signal Processing,'' Icassp-2000, Istanbul, Turkey, 6 June.

Roark, R. J., and Young, W. C. (**1975**). *Formulas for Stress and Strain*, 5th ed. (McGraw-Hill, New York).

Rosowski, J. J., Ravicz, M. E., Teoh, S. W., and Flandermeyer, D. (**1999**). ''Measurements of Middle-Ear Function in the Mongolian Gerbil, a Specialized Mammalian Ear,'' Audiol. Neuro-Otol. **4**, 129–136.

Rothenberg, M. (**1981**). ''Some relations between glottal flow and vocal fold contact area,'' ASHA Rep. **11**, 88–96.

Sharpe, R. M., Grant, J. B., Champagne, N. J., Johnson, W. A., Jorgenson, R. E., Wilton, D. R., Brown, W. J., and Rockway, J. W. (**1997**). ''EIGER: Electromagnetic Interactions GeneRalized,'' IEEE AP-S International Symposium and North American URSI Radio Science Meeting, Montreal, Canada, July, pp. 2366–2369.

Skolnik, M. (**1990**). ''Radar Handbook,'' 2nd ed. (McGraw–Hill, New York).

Stevens, K. N. (**2000**). Acoustic Phonetics (MIT, Cambridge, MA).

Svirsky, M. A., Stevens, K. N., Matthies, M. L., Manzella, J., Perkell, J. S., and Wilhelms-Tricarico, R. (**1997**). ''Tongue surface displacement during bilabial stops,'' J. Acoust. Soc. Am. **102**, 562.

Titze, I. R. (**1984**). ''Parameterization of the glottal area, glottal flow, and vocal fold area,'' J. Acoust. Soc. Am. **74**, 570–580.

Titze, I. R. (**1990**). ''Interpretation of the Electroglottographic Signal,'' J. Voice **4**(1), 1–9.

Titze, I. R. (**1994**). *Principles of Voice Production* (Prentice Hall, Englewood Cliffs, NJ).

Titze, I. R., Story, B. H., Burnett, G. C., Holzrichter, J. F., Ng, L. C., and Lea, W. A. (**2000**). ''Comparison between electroglottography and electromagnetic glottography,'' J. Acoust. Soc. Am. **107**, 581.

Estimation of sound pressure levels of voiced speech from skin vibration of the neck

Jan G. Švec[a)]
National Center for Voice and Speech, the Denver Center for the Performing Arts, 1245 Champa Street, Denver, Colorado 80204, and Groningen Voice Research Lab, Department of Biomedical Engineering, University of Groningen Medical Center, A. Deusinglaan 1, NL 9713 AV Groningen, The Netherlands

Ingo R. Titze[b)] and Peter S. Popolo[c)]
National Center for Voice and Speech, the Denver Center for the Performing Arts, 1245 Champa Street, Denver, Colorado 80204, and Department of Speech Pathology and Audiology, The University of Iowa, 330-WJSHC, Iowa City, Iowa 52242

(Received 9 February 2004; revised 27 July 2004; accepted 26 November 2004)

How accurately can sound pressure levels (SPLs) of speech be estimated from skin vibration of the neck? Measurements using a small accelerometer were carried out in 27 subjects (10 males and 17 females) who read Rainbow and Marvin Williams passages in soft, comfortable, and loud voice, while skin acceleration levels (SALs) and SPLs were simultaneously registered and analyzed every 30 ms. The results indicate that the mean SPL of voiced speech can be estimated with accuracy better than ± 2.8 dB in 95% of the cases when the subjects are individually calibrated. This makes the accelerometer an interesting sensor for SPL measurement of speech when microphones are problematic to use (e.g., noisy environments or in voice dosimetry). The estimates of equivalent SPL, which is the logarithm of averaged relative energy of voiced speech, were found to be up to 1.5 dB less accurate than the mean SPL. The estimation accuracy for instantaneous SPLs was worse than for the mean and equivalent SPLs (on average ± 6 and ± 5 dB for males and females, respectively). [DOI: 10.1121/1.1850074]

PACS numbers: 43.70.Jt [AL] Pages: 1386–1394

I. INTRODUCTION

During voice production, vibrations originating at the level of the vocal folds are transmitted through the air as well as through the tissues of the speaking subject. The vibration of the body tissues can be registered with an accelerometer (Stevens, Kalikow, and Willemain, 1975; Sundberg, 1983, 1992; Sakakura and Takahashi, 1988; Pešák, 1990). An accelerometer is a sensor frequently used in industry for measurement of mechanical vibrations of solid bodies. In contrast to the microphone, which picks up air pressure oscillations, the accelerometer registers only accelerations of the surface to which it is attached and is almost insensitive to the surrounding air pressure oscillations.

Accelerometers can be very small and lightweight. While not used frequently, miniature accelerometers have been applied to voice and speech measurement, e.g., for detecting the glottal vibrations and nasalization (e.g., Stevens *et al.*, 1975; Lippmann, 1981; Horii, 1983). The accelerometer signal has been found useful also for the extraction of fundamental frequency of voice and frequency perturbation measurements (e.g., Horii, 1982; for review, see Hess, 1983). When attached to the body of a speaking subject, the accelerometer is sensitive mainly to vibrations generated by that speaker and practically ignores the surrounding acoustic noise as well as speech of other subjects (Coleman, 1988), which makes it potentially well suited for applications dealing with voicing detection such as in voice accumulation/dosimetry (Hillman and Cheyne 2003a; 2003b; Švec, Titze, and Popolo, 2003b; Cheyne *et al.*, 2003).

A convenient location for the attachment of the accelerometer on speaking subjects was found to be the jugular notch located at the lower frontal part of the neck between the sternum and larynx. Some of the reasons for the jugular notch being the location of choice are (1) signals of maximal amplitude can be expected here (Stevens *et al.*, 1975); (2) exact positioning of the accelerometer around this location was reported not to be critical (Stevens *et al.*, 1975); and (3) this location is relatively inconspicuous and quite comfortable for long-term attachment of the accelerometer (Cheyne *et al.*, 2003). The last two reasons make the jugular notch a more convenient location for placement of the accelerometer than the skin over the lamina of the thyroid cartilage.

In situations when a microphone is problematic or inconvenient to use, such as in measurement of speech in noisy environment, or in long-term monitoring of voice, it can be advantageous to use an accelerometer for estimation of sound pressure level (SPL) of voice. So far, it has not been very clear whether this can be done with an accuracy approaching that of microphone measurements. Hypothetically, it can be expected that the more intense the voice is, the more intense the vibration of the skin of the neck will be. The SPL of voice signal is, however, influenced not only by the voice source signal but also by the radiation characteristics of the mouth and vocal-tract formants, whereas the ac-

[a)]Electronic mail: svecjan@vol.cz
[b)]Electronic mail: ititze@dcpa.org
[c)]Electronic mail: peter-popolo@uiowa.edu

celeration amplitudes of the skin at the jugular notch are influenced by the skin transmission properties and subglottal formants. A major step toward relating the skin vibration to the vocal output was made by Cheyne (2002), who created a phonation model which included all these characteristics. Using this model, he was able to estimate the sound pressure levels of sustained vowels from the accelerometer signal within the 95% confidence interval of ca. ±7 dB.

The present study has been driven by the need to estimate the sound pressure levels from accelerometer signals for the purpose of voice dosimetry in public school teachers. A low-cost small accelerometer was chosen as the voice sensor for a pocket-PC-based portable device called a ''voice dosimeter.'' The dosimeter was developed by the authors, enabling the skin acceleration level (SAL) and the fundamental frequency of voice to be recorded in 30-ms intervals for the entire day (Švec *et al.*, 2003b). The overall long-term goal of the authors is to measure the daily vocal doses in speaking subjects (Titze, Švec, and Popolo, 2003; Švec, Popolo, and Titze, 2003a) including their SPLs. For this goal, running speech was analyzed to obtain the best-fit relationship between speaking SPLs and SALs for all speech sounds over a large intensity range. The goals of the current study are (1) to describe a method for relating skin acceleration levels (SAL) to sound pressure levels (SPL) in voiced speech and (2) to determine how accurately the SPL can be predicted from the skin vibration using this method.

II. METHOD

A. Subjects and equipment setup

Laboratory measurements were carried out on 27 volunteer subjects (17 females and 10 males), mostly high-school teachers participating in the vocal dosimetry project plus some vocally healthy employees of the National Center for Voice and Speech (NCVS). Teachers with pathologies were accepted as subjects if they were able to carry out their profession. First, the accelerometer (Knowles Electronics BU-7135) was attached to the neck of the subject at the jugular notch, i.e., at the anterior part of the neck, below the larynx, on the softest place between the cricoid cartilage and the sternum. The accelerometer was attached using a surgical adhesive (Mastisol®, manufactured by Ferndale Laboratories) applied between the skin and the accelerometer, and a Suture-Strip (TS-3101, manufactured by Derma Sciences) was applied over the surface of the accelerometer, further securing it to the skin (for more detail, see Popolo *et al.*, 2004). This attachment was the same as used by the subject when wearing the pocket-PC-based voice dosimeter during field measurements.

The equipment setup is shown in Fig. 1. The subject with the accelerometer attached to the neck was seated in a sound booth (single-wall IAC isolation booth, 2.3 m deep, 2.2 m wide, and 2.3 m high) wearing a head-mounted microphone (omnidirectional B3 Lavalier microphone by Countryman Associates) placed slightly to the side of the mouth and out of the air stream. The microphone was mounted on a plastic frame, which was worn like a pair of glasses and was about 5 cm from the mouth. The exact distance of the head-

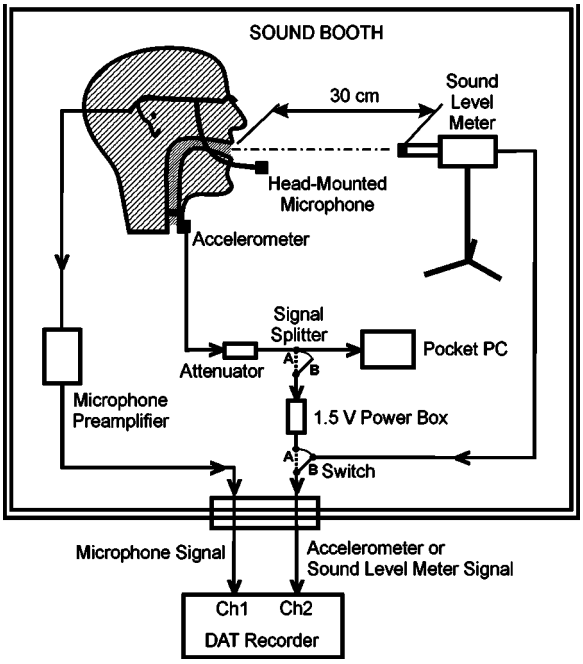


FIG. 1. The setup for relating the SPL of speech to the skin vibration of the neck as used in pocket-PC-based voice dosimetry.

mounted microphone from the mouth was not critical as long as it stayed constant during the experiment because the microphone was calibrated to yield, for the same phonation, identical sound pressure levels as the sound level meter (Brüel&Kjaer 2238, set to C-frequency weighting) positioned 30 cm from the mouth (see also the two-step calibration procedure in Sec. II B). The head-mounted microphone signal was phantom powered with the AudioBuddy Dual Mic Preamp (M-Audio) and connected to channel 1 of the DAT recorder (Technics SV-DA10, sampling rate 48 kHz). The sound level meter microphone signal was connected to channel 2 of the DAT recorder. The gains were adjusted so that the peak level of the maximally loud voice [ca. 115 dB(C) SPL@30 cm] was ca. 6 dB below the clipping level of the DAT recorder and the preamplifier was not overloaded. The ambient noise level in the booth was about 55 dB(C) (sound-level meter value) which corresponded to ca. 45-dB(unweighted) SPL background noise level of the head-mounted-microphone recording (lowering of the noise level was achieved by setting the SPL at the head-mounted microphone to correspond to the SPL at 30-cm distance).

The NCVS portable voice dosimeter, i.e., a pocket PC (Compaq iPAQ, model 3765) in which the internal microphone input was rewired into an external audio input (Popolo *et al.*, 2004; Švec *et al.*, 2003b), was used to record the vibrations of the neck registered by the accelerometer signal. The dynamic range of the pocket PC was rather small (about 45 dB) and it was necessary to use an attenuator (6–10 dB) to prevent the accelerometer signal from clipping and to match the accelerometer signal range to the pocket PC range. The combined frequency response of the accelerometer, attenuator, and pocket PC sound board was checked experimentally by comparing it to a known response of a calibration accelerometer. It showed less than ±1.5-dB variation in the 50–1000-Hz frequency range (i.e., the range containing

most of the skin acceleration energy which determines the speaking SAL values). The pocket PC standard voice-recording software was used to record the accelerometer signal, which was saved as a *WAV* audio file (Švec *et al.*, 2004). The specifications of the pocket PC recording format were selected as follows: PCM, mono, 16-bit resolution, and 11 025-Hz sampling frequency. The automatic gain control was disabled.

To ensure synchronization of the DAT and pocket PC recordings, a signal splitter, together with a 1.5-V accelerometer power box (custom-made to match the input impedance of the accelerometer as specified by the manufacturer), was used during selected parts of the procedure (see below) to record the accelerometer signal simultaneously on the pocket PC and on the DAT recorder. By changing the switch positions from B to A (Fig. 1), the accelerometer signal replaced the sound level meter microphone signal in channel 2 of the DAT recorder.

B. Recording procedure

An experimental protocol was designed to allow the calibration to be done within a short time frame (total duration of the recording session was between 5–10 min). It was comprised of tasks that were easy for the participating subjects to perform. Two recordings were obtained. In the first recording, the head-mounted microphone signal was calibrated for absolute SPL measurement using a two-step calibration procedure (Švec *et al.*, 2003a): the switches in Fig. 1 were set to B positions, the DAT recorder was started and the Brüel&Kjaer 4231 calibrator was used to produce a stable tone of 94 dB SPL *re* 20 μPa registered by the sound level meter. The dB level was verified visually on the sound level meter display. The calibrator was removed and the calibration dB value was announced in the microphone. After that, the subject produced sustained /a/ at comfortable pitch and loudness which was registered simultaneously by the head-mounted microphone and the sound level meter at 30-cm distance. These calibration signals were later used in the analysis phase to adjust the level of the head-mounted microphone to be identical to that measured by the sound level meter at 30 cm. The DAT recorder was stopped.

The second recording was done simultaneously on the DAT recorder and the pocket PC starting with the switches in Fig. 1 set to A positions. After the recordings started, the accelerometer was mechanically excited (finger-flicked) several times to produce sharp peaks, which were later used for exact synchronization of the beginning of the DAT and pocket PC recordings. While both the recordings were running, the splitter was removed from the accelerometer signal path (the switch positions in Fig. 1 were changed from A to B) and the subject read the standard first six sentences of the Rainbow Passage (Fairbanks, 1960). The passage was read successively in comfortable, soft, and loud voice. Then, the subject read the all-voiced Marvin Williams Passage (Sataloff 1997, p. 828) in comfortable, soft, and loud voice. After that the subject produced a sequence of soft phonations: sustained /i/ on a comfortable pitch as softly as possible, glide from low to high pitch as softly as possible on /i/ vowel, high-pitched extremely soft repetitions /hee-hee-hee-hee-hee/ and high-pitched soft singing of the first bars of ''Happy Birthday.'' These utterances, which were previously found to be useful in perceptually detecting vocal-fold swelling (Bastian, Keidar, and Verdolini-Marston, 1990), were used here also for determining the voicing threshold. At the end of the recording, the splitter was again put in the accelerometer signal path (switch positions in Fig. 1 changed back to A) and the accelerometer was flicked several times to allow synchronization of the ends of the DAT and pocket PC recordings. For repeatability measurements, six of the 27 subjects (three male and three female, all of them vocally healthy) repeated the second recording three times.

C. Data processing

The *WAV* audio file recorded on the pocket PC was downloaded to a PC through a USB cable using Microsoft ACTIVESYNC software (part of the standard pocket PC software). The DAT recordings were digitally transferred to a PC using the CSL 4400 (Kay Elemetrics) DAT interface and saved as stereo *WAV* audio files.

MATLAB (v.6.5) was used for signal analysis. The soft voiced phonations were first analyzed to determine the subject's softest voiced SALs (in relative units of dB) and to verify that these levels were at least 5 dB above the noise level of the recording. This condition was satisfied in all the subjects. The voicing threshold was manually set to be 5 dB lower than the softest sustained SALs produced unless these levels were less than 10 dB above the noise level of the accelerometer recording. If so, the voicing threshold was set at the average dB level between the softest voiced SAL and the noise level.

The two-step calibration recording described earlier was used to adjust the SPL values of the head-mounted microphone to be equal to the absolute sound levels given by the sound level meter (see Švec *et al.*, 2003a, for details). Then, the files from the pocket PC and DAT recordings of the Rainbow and Marvin Williams passages were synchronized using the impulses at the beginning and end of the recordings. The microphone and accelerometer signals were divided into synchronized successive frames of 30-ms duration (adjacent frames with no overlap). In order to compensate for a time shift caused by about 17-cm difference between the accelerometer (ca. 5 cm below the vocal folds) and the head-mounted microphone (ca. 5-cm distant from the mouth plus ca. 17-cm vocal-tract length), the microphone signal was shifted 0.5 ms forward in time. The sound pressure levels (SPL) and the skin acceleration levels (SAL) were calculated in each frame as $20\log_{10}[\mathrm{rms}(y-\overline{y})]-C$, where y denotes the microphone or accelerometer signal, $\overline{y}$ is the mean signal value within the frame, rms is the root mean square value over the frame, and C is the calibration constant.

The accelerometer signal was used to distinguish between the voiced and unvoiced segments (in 30-ms frames) of speech. For this, the SAL was compared to the voicing threshold determined earlier. The frame was considered voiced if its SAL was above the voicing threshold, and if also the accelerometer spectral energy within the ''voicing range'' was greater than the energy outside this range. The voicing range was defined as the range between the lowest

expected fundamental frequency of the subject's speech and 1000 Hz (this is the range which is expected to contain most of the skin acceleration energy in voiced speech). A 1024-point FFT of the Hanning-windowed signal was used for the spectral analysis. More details on the calibration, recording procedure, and data processing can be found in Švec *et al.* (2004).

D. Calculation of mean and equivalent SPL of speech

Mean SPL values of voiced speech were calculated as arithmetic means of the SPLs from the voiced frames

$$\mathrm{SPL}_{\mathrm{mean}} = \sum_{n=1}^{N} [k_v(n)\mathrm{SPL}(n)] \Big/ \sum_{n=1}^{N} k_v(n), \tag{1}$$

where n is the frame index, N is the total number of frames in the analyzed segment of speech, k_v is the voicing unit step function (1 for voiced and 0 for unvoiced frame), and the $\mathrm{SPL}(n)$ is the SPL value within the frame n.

Equivalent SPL values of voiced speech were calculated using the equation

$$\mathrm{SPL}_{\mathrm{eq}} = 10\log_{10}\left(\frac{1}{N}\sum_{n=1}^{N} [k_v(n)10^{\mathrm{SPL}(n)/10}]\right). \tag{2}$$

This equation is consistent with the ANSI definition for the equivalent continuous sound level (ANSI S1.1-1994), the only difference being the inclusion of the k_v factor which sets the energy in the unvoiced frames to zero. Such a modification was found useful here because the SPL of unvoiced segments can hardly be estimated from SAL. As the SPLs of the unvoiced frames are considerably smaller than those of the voiced frames, and the factor $10^{\mathrm{SPL}(n)/10}$ (i.e., frame energy rather than frame SPL) is used in Eq. (2), the replacement of the true energy of the unvoiced frames with zero shall have only minor influence on the resulting equivalent SPLs for continuous speech.

E. Estimation of SPL from the skin vibration—The best fits

The best-fit relationship between the SPL and SAL values was obtained for each subject using standard MATLAB best-fit routines (the least-squares method, function *polyfit*). These routines were applied to all (i.e., over 1000) pairs of corresponding SAL and SPL values (30-ms frames) measured in the combined soft, comfortable and loud speech of that subject. Two kinds of best-fit relationships were investigated—the first-order (linear) fit

$$\mathrm{SPL}_{\mathrm{FIT}} = L_1\mathrm{SAL} + L_0, \tag{3}$$

and the second-order (quadratic) fit

$$\mathrm{SPL}_{\mathrm{FIT}} = Q_2\mathrm{SAL}^2 + Q_1\mathrm{SAL} + Q_0, \tag{4}$$

where L_1, L_0 and Q_2, Q_1, Q_0 are polynomial coefficients of the two fits. Once the polynomial coefficients were found for a given individual, the estimated long-term SPLs ($\mathrm{SPL}_{\mathrm{mean\,FIT}}$ and $\mathrm{SPL}_{\mathrm{eq\,FIT}}$) were calculated by obtaining the predicted individual frame values $\mathrm{SPL}_{\mathrm{FIT}}(n)$ by means of Eqs. (3) and (4) and using these instead of the true $\mathrm{SPL}(n)$ values in Eqs. (1) and (2).

The true and estimated values for all the subjects were imported into MS EXCEL 2000. Accuracy of the fits for estimation of the $\mathrm{SPL}_{\mathrm{mean}}$ and $\mathrm{SPL}_{\mathrm{eq}}$ was quantified by an error function expressed in dB difference

$$\epsilon_{\mathrm{SPL}_x} = \mathrm{SPL}_{x\ \mathrm{FIT}} - \mathrm{SPL}_x, \tag{5}$$

between the estimated and true SPL values, where the abbreviation SPL_x replaces $\mathrm{SPL}_{\mathrm{mean}}$ or $\mathrm{SPL}_{\mathrm{eq}}$. The uncertainty of the dB differences was calculated as standard deviation

$$\mathrm{SD}(\epsilon_{\mathrm{SPL}_x}) = \sqrt{\frac{\sum_{j=1}^{M}(\epsilon_{\mathrm{SPL}_x\ j} - \overline{\epsilon}_{\mathrm{SPL}_x})^2}{M-1}}, \tag{6}$$

where j is the subject index and M is the total number of the measured subjects. The symbol $\overline{\epsilon}_{\mathrm{SPL}_x}$ denotes the average values of the $\epsilon_{\mathrm{SPL\,mean}}$ and $\epsilon_{\mathrm{SPL\,eq}}$ across subjects; they represent the biases of the fits that should be expected to be close to zero if the fits estimate the $\mathrm{SPL}_{\mathrm{mean}}$ and $\mathrm{SPL}_{\mathrm{eq}}$ accurately.

F. Fit repeatability

To investigate the repeatability of the best fit within a subject, an average best fit for a subject was calculated by averaging the individual polynomial coefficients L and Q of the best fits from three different trials. The mean and equivalent SPLs of the speech estimated from the individual and the average best fits $\mathrm{SPL}_{\mathrm{mean\,FIT}}$, $\mathrm{SPL}_{\mathrm{mean\,AVG.FIT}}$, $\mathrm{SPL}_{\mathrm{eq\,FIT}}$, $\mathrm{SPL}_{\mathrm{eq\,AVG.FIT}}$ were calculated by means of Eqs. (1)–(4). Then, the differences

$$\epsilon_{\mathrm{SPL}_x\ \mathrm{FIT}} = \mathrm{SPL}_{x\ \mathrm{FIT}} - \mathrm{SPL}_{x\ \mathrm{AVG.FIT}}, \tag{7}$$

were obtained for $\mathrm{SPL}_x = \mathrm{SPL}_{\mathrm{mean}}$ and $\mathrm{SPL}_x = \mathrm{SPL}_{\mathrm{eq}}$ and their standard deviation was calculated for each subject. The average standard deviation for all the subjects was obtained by averaging the individual standard deviations using the formula

$$\mathrm{SD}(\epsilon_{\mathrm{SPL}_x\ \mathrm{FIT}}) = \sqrt{\frac{1}{K}\sum_{j=1}^{K}[\mathrm{SD}_j(\epsilon_{\mathrm{SPL}_x\ \mathrm{FIT}})]^2}, \tag{8}$$

where j is the subject index and K is the total number of subjects involved in the repeatability experiments. Equation (8) quantifies the uncertainty of the $\mathrm{SPL}_{\mathrm{mean}}$ and $\mathrm{SPL}_{\mathrm{eq}}$ estimation from skin vibration caused by the variation of the best fits.

III. RESULTS

A representative example of the relationship between the levels of the accelerometer signal and the microphone signal is shown in Fig. 2. The figure was obtained by plotting the corresponding pairs of the SAL and SPL values from the voiced 30-ms frames of the soft, comfortable, and loud readings of the Rainbow passage by a female subject. First and last frames of each voicing segment were excluded to reduce the number of outlier points (which is generally recommended when performing the best fit of the empirical data). Higher skin acceleration levels generally correspond to higher sound pressure levels.

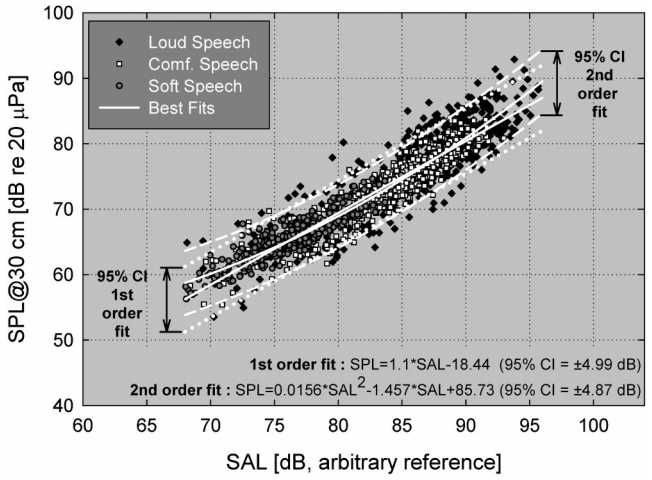


FIG. 2. The relationship between the skin acceleration levels (SAL) and sound pressure levels (SPL) obtained from the combined soft (gray), comfortable (white), and loud (black) readings of the Rainbow passage by a female subject. The solid white lines show the first-order (linear) and second-order (quadratic) best fits of the data; the dotted and dashed white lines mark the 95% confidence intervals for the linear and the quadratic fits, respectively.

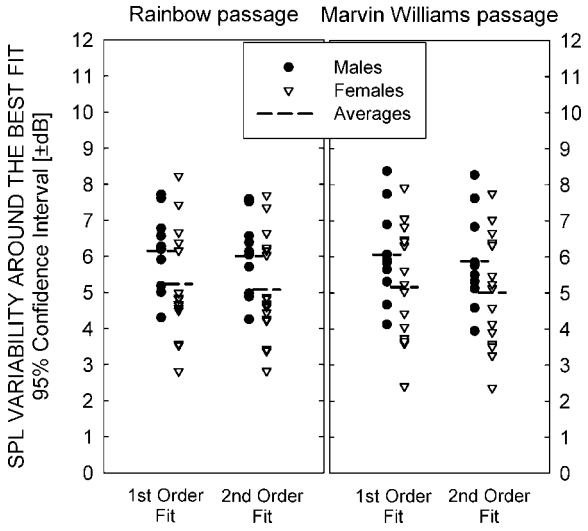


FIG. 3. SPL variability around the best SAL/SPL fit in running speech measured in 10 male and 17 female subjects. The 95% confidence interval is between ±2 and ±8.5 dB for the best and worst cases, respectively. The average 95% confidence interval is ca. ±6 dB for men and ca ±5 dB for women. Only little differences can be seen between the first-order and second-order fits and between the Rainbow and Marvin Williams passages. The results are obtained for SPL estimations in 30-ms windows.

The solid white lines show the best first-order (linear) and second-order (quadratic) fits of the data for this subject. The equations of these fits are given at the bottom of the figure. The confidence intervals (marked in Fig. 2 by the white dotted and dashed lines) reveal that 95% of all the measured SPLs lie approximately within ±5 dB around the best fits, which suggests that the running SPL values can be predicted with the uncertainty of ca. ±5 dB in this subject.

Notice that the loud speech values (black quadrangles in Fig. 2) are found not only towards the loud boundaries (upper-right quadrant) as expected, but are distributed along the whole dynamic range of speech, reflecting the fact that loud speech is not only created by high SPLs but also by SPLs varying rapidly between high and low levels. Also notice that the loud speech is responsible for great majority of the outliers, i.e., the values outside the 95% confidence interval. On the basis of this finding it can be hypothesized that the worsened SAL-to-SPL correspondence might be linked to differences between the transients in the accelerometer and microphone signals (i.e., the skin vibration and the acoustic signal seem to respond differently to the rapid dynamic changes occurring in the loud speech). Confirmation of this hypothesis would require a more detailed study of these phenomena and is outside the scope of the present study, however.

Figure 3 shows the 95% confidence intervals in dB for all the subjects. The Rainbow and Marvin Williams passages, first-order and second-order fits, and male and female subjects are compared here. The female subjects (unfilled triangles) had, on average, about 1 dB smaller confidence intervals, i.e., less SPL variability around the fits, than the male subjects (filled circles). The average 95% interval was found, in all the cases shown in Fig. 3, to be ca. ±6 dB in males and ca. ±5 dB in females, which are the average uncertainties one can expect when predicting the running SPL values from skin acceleration in speech.

The second-order fits did not perform much better than the first-order fits, as suggested by less than 0.25-dB difference between the averages of the 95% confidence intervals seen in Fig. 3. Also, we expected the readings of the Marvin Williams passage to give better fits than the Rainbow passage (due to less variation between the voiced and unvoiced segments), i.e., yielding smaller 95% confidence intervals. The results from the 27 subjects shown in Fig. 3 do not support this hypothesis, however (suggesting the transitions between vowels and voiced consonants pose a problem similar to the voiced–voiceless transitions when relating the SALs to SPLs in running speech). In both the Marvin Williams (right plot) as well as the Rainbow passages (left plot), the 95% confidence intervals are between ±2 dB (the best cases) and ±8.5 dB (the worst cases). On the basis of these results, only the Rainbow passage fits will be considered further.

In many speech investigations, rather than measuring the instantaneous SPL values, the goal is to determine the mean SPL (SPL_{mean}) or equivalent SPL (SPL_{eq}) of speech over longer time periods. In order to find out how accurately these long-term SPL values can be estimated from skin acceleration levels, the true SPL_{mean} and SPL_{eq} for the softly, comfortably, and loudly read Rainbow passages were compared to those estimated from the SALs using the best-fit relationships. Figure 4 shows the relationship between the true and estimated mean SPLs for all the subjects when using the individually determined first-order best-fit relationship. In all the cases the estimated values are no more than 2 dB different from the true values.

The estimated values displayed in Fig. 4 were obtained from the calibration readings, which were identical to those from which the best fits were obtained. In practice, however, the calibration readings for the best fits are done separately from the measurement itself. Therefore, in order to find out

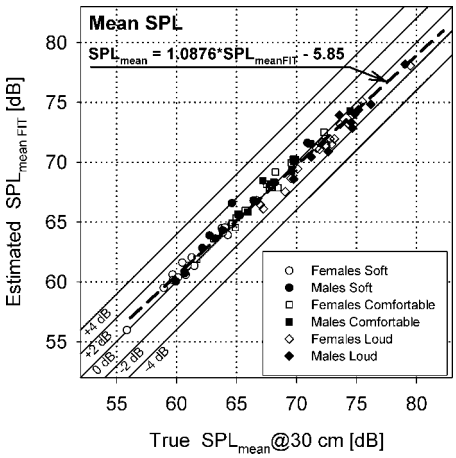


FIG. 4. The relationship between the true mean SPL of the Rainbow passage and the estimated mean SPL derived from the first-order best fit from skin vibration. Values for soft, comfortable, and loud readings by 10 male and 17 female subjects are shown in the plot. The solid lines mark equality and under- and overestimates in 2-dB intervals. The data show that the accuracy of the first-order best fit is better than ±2 dB for all the cases. The thick dashed line indicates the regression relationship between the true and estimated mean SPLs.

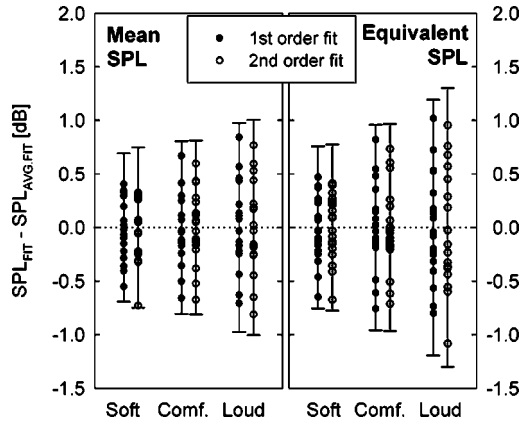


FIG. 6. Variability of the best fits around the average fit when repetitively reading the Rainbow passage. Results of three repetitive readings by six subjects are shown; the average fit was obtained for each of the subjects separately. The error bars show *two* standard deviations, which correspond to 95% confidence interval in Gaussian distribution.

what the accuracy of the method is, it is important to know how repeatable is the best fit. Figure 5 shows three fits obtained from three subsequent readings of the Rainbow passage by the same male subject. There are slight differences between the fits.

The differences among the fits can be quantified by obtaining an average fit from the three fits and measuring the differences between the SPL values estimated by the individual fits and those estimated by the average fit [Eqs. (7)–(8)]. Figure 6 shows these differences in six subjects (three males and three females) who read the Rainbow passage three times. The smallest variation was found in soft voice, the largest in loud voice. Overall, for the estimation of the SPL_{mean} (left graph) the variation is less than 1 dB and the variation for estimation of the SPL_{eq} (right graph) is only slightly larger.

The difference between the speech produced during calibration and that produced during the measurement adds uncertainty to the overall SPL estimates. The total uncertainty of the SPL estimation from skin vibration shall take into account (a) the SD_{cal} uncertainty of the best fit in estimating the SPL_{mean} and SPL_{eq}; (b) the SD_{fit} uncertainty of the calibration fit with respect to average fit for that person; and (c) the SD_{speech} uncertainty of the measured speech (i.e., its best fit) with respect to average speech of that person. As these uncertainties can be considered independent of each other, the total uncertainty SD_{total} can be calculated from the three standard deviations as (Adams, 2002)

$$SD_{total} = \sqrt{SD_{cal}^2 + SD_{fit}^2 + SD_{speech}^2}. \quad (9)$$

Here, SD_{cal} is the standard deviation of the residual between the estimated and true SPL [i.e., the values $SD(\epsilon_{SPL\ mean})$, $SD(\epsilon_{SPL\ eq})$ calculated by Eq. (6)], SD_{fit} is the standard deviation of the fit of the repeated calibration speech [i.e., the values $SD(\epsilon_{SPL\ mean\ FIT})$, $SD(\epsilon_{SPL\ eq\ FIT})$ calculated by Eq. (8)] and the SD_{speech} is the standard deviation of the fit of the repeated measured speech. The SD_{speech} is identical to SD_{fit} when the same speech material (e.g., the Rainbow passage) is used for both the calibration and the measurement.

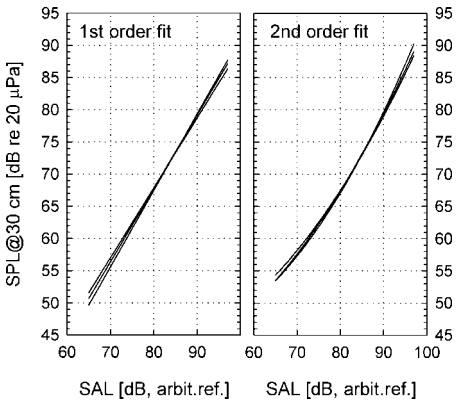


FIG. 5. The repeatability of the best fit: three best fits obtained from three consecutive readings of the Rainbow passage by the same male subject.

The final accuracy of the long-term SPL estimation for the Rainbow passage based on the measurements of the 27 subjects is displayed in Fig. 7. The estimation accuracy for the SPL_{mean} is on the left, the right graph is for the SPL_{eq}. The mean values reveal that soft speech tends to be slightly overestimated (on average by ca. 0.5 dB) while loud speech tends to be slightly underestimated (by 1–2 dB on average) for both fits. The second-order fit is, on average, slightly more accurate (i.e., less biased) for the loud speech than the first-order fit, but not by much (by about 0.25 dB for the SPL_{mean} and 0.5 dB for the SPL_{eq}).

The error bars in Fig. 7 show the interval of *two standard deviations* from the mean, which correspond to 95% confidence interval in Gaussian distribution. The shorter error bars correspond to the calibration uncertainty [SD_{cal} in Eq. (9)], i.e., the variation of the error in SPL estimation caused by the limited ability of the fit to express the true SAL/SPL relationship. The longer error bars show the total uncertainty, which considers also the limited repeatability of the best fit and of the measured speech [SD_{total} in Eq. (9)]. The 95% confidence interval is larger for the equivalent

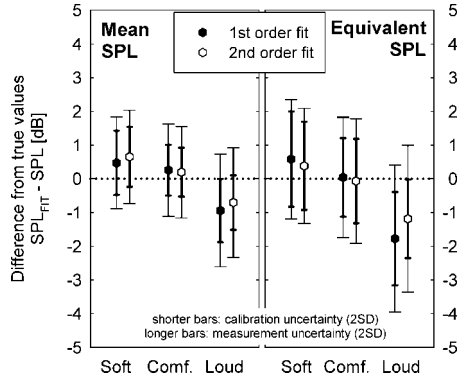


FIG. 7. Accuracy of the first- and second-order best fits for estimating the mean and equivalent SPLs for the soft, comfortable, and loud readings of the Rainbow passage. The mean (bias) values (diamond symbols) reveal that the soft values are on average slightly overestimated, whereas the loud values are slightly underestimated by the fits. The error bars show *two* standard deviations (95% confidence interval in Gaussian distribution). The results for the first-order-fit-derived mean SPLs (black diamonds in the left graph with the shorter error bars) correspond to all the individual values shown in Fig. 4.

SPLs (right plot) than for the mean SPLs (left plot), which suggests the estimation of the SPL_{eq} from the skin vibration is less reliable than the estimation of the SPL_{mean}. While the estimated SPL_{eq} of the loud speech is, in 95% of the cases, less than 4 dB different from the true level, no more than 2.8-dB difference can be expected for the SPL_{mean} in 95% of the cases. The first-order (empty diamonds) and the second-order fits (black diamonds) show very similar 95% confidence intervals.

The next question is, how good is the fit from the Rainbow passage when applied to general speech? To address this question, the best-fit relationships from the Rainbow passage were applied to estimate the SPL_{mean} and SPL_{eq} of the Marvin Williams passage. Figure 8 shows the SPL_{eq} estimation results for the first-order fit. When compared to Fig. 4, note the larger spread of the values. Whereas in Fig. 4 the SPL_{mean} values lay within the ± 2-dB interval from the identity line,

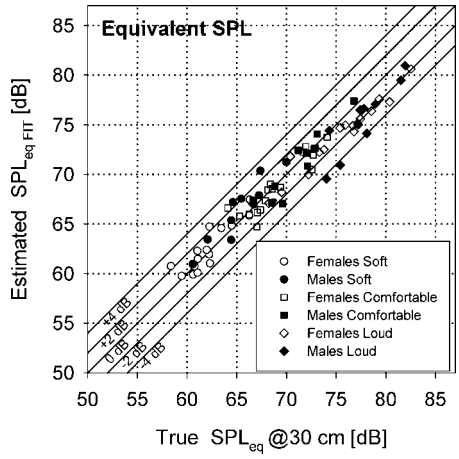


FIG. 8. Similar to Fig. 4 but the Marvin Williams levels are estimated here by the Rainbow-passage best fits and the results are plotted for equivalent SPLs instead of the mean SPLs. Note the larger spread of the values than in Fig. 4. In the worst cases the skin-derived equivalent SPLs are ca. 4.5 dB different from the true values.

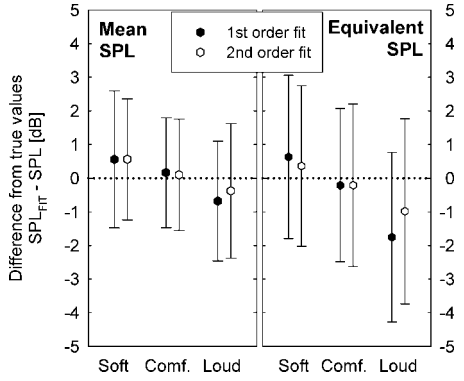


FIG. 9. Similar to Fig. 7 but here the SPLs of the Marvin Williams passage are estimated by the Rainbow-passage best fits. The biases and the 95% confidence intervals (the error bars) are not much different from those of the Rainbow passage (the longer error bars in Fig. 7). The mean and error distribution values for the first-order-fit-derived equivalent SPLs (black diamonds with the error bars in the right graph) correspond to all the individual values shown in Fig. 8.

in Fig. 8 the worst-case estimated SPL_{eq} values are about 4.5 dB different from the true values.

The accuracies of the Marvin Williams estimates are shown in Fig. 9, again comparing the first- and second-order fits, soft, comfortable, and loud speech, and mean and equivalent SPLs as done in Fig. 7. Overall, the estimates are only up to 0.8 dB worse than those for the repeated Rainbow passage shown in Fig. 7. The estimated SPL_{mean} is, with 95% probability, less than 2.6, 1.8, or 2.5 dB different from the true level for the soft, comfortable, or loud speech, respectively. The estimated SPL_{eq} is, with 95% probability, less than 3.1, 2.7, or 4.3 dB different from the true level for the soft, comfortable, or loud speech, respectively.

IV. DISCUSSION

A. SPL_{mean} versus SPL_{eq}

The results for the long-term average values indicate that the estimation of the mean SPL of the voiced speech is more accurate than the estimation of the equivalent SPL. The equivalent levels have been, however, often preferred for speech measurements—for instance, there exists an ISO standard [ISO 9921-1:1996(E)], which specifies the equivalent sound levels for acceptable and unacceptable vocal effort for occupational purposes. What is the difference between the SPL_{mean} and SPL_{eq}? The SPL_{mean} is obtained by averaging the individual SPLs over the *voiced frames only* [Eq. (1)]. The SPL_{eq}, on the other hand, averages the *voiced energy* over *all the frames, including the unvoiced ones* [Eq. (2)].

Both the SPL_{mean} and SPL_{eq} measures have advantages and disadvantages with respect to each other. The SPL_{eq} values are dependent on the duration of voicing pauses: the SPL_{eq} of speech produced with constant effort theoretically decreases by 6 dB when the voicing pauses are doubled, while its SPL_{mean} stays constant. The SPL_{mean} is, on the other hand, more sensitive to the method of elimination of the unvoiced speech segments than the SPL_{eq}. Different SPL_{mean} values can theoretically be obtained for the same speech when different frame durations are used (i.e., 30 ms, 125 ms,

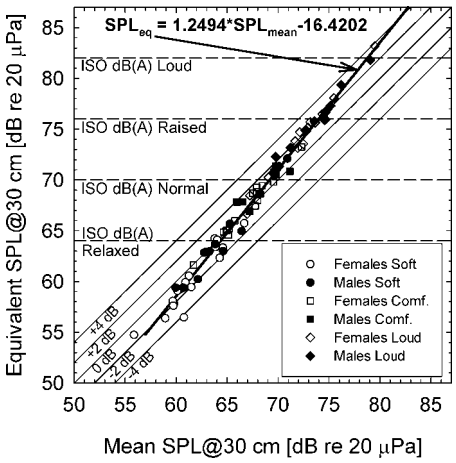


FIG. 10. The relationship between the true mean SPLs and true equivalent SPLs obtained from 27 subjects (10 male, 17 female, Rainbow passage readings, 30-ms frames used for SPL averaging). Note that for the soft voice the equivalent SPLs are slightly lower than the mean SPLs, but for the loud voice the equivalent SPLs are higher than the mean SPLs. The ISO values for standard vocal effort are shown in the figure for reference purposes.

1 s, etc.), as the speech pauses and voiceless consonants are eliminated less accurately with longer frames. The SPL_{mean} value approaches the SPL_{eq} value when the analysis frames are being prolonged [identity of these two is warranted when the duration of the analysis frame is equal to the duration of the whole speech signal analyzed, i.e., as N in Eqs. (1) and (2) approaches the value of 1].

A relatively short analysis frame of 30 ms was chosen here to distinguish between the voiced and unvoiced segments with greater accuracy. For such short frames, however, differences between the SPL_{mean} and SPL_{eq} can be expected. It is therefore useful to know the relationship between the two SPLs for this particular frame duration. To the knowledge of the authors, such a relationship has also not been studied before. The relationship obtained from the analysis of the Rainbow passage read by the 27 subjects is shown in Fig. 10. The true airborne SPLs (not the SAL-derived estimates) are related here.

The figure shows that the mean SPLs are not much different from the equivalent SPLs of voiced speech, especially for the comfortable speech levels. When speaking softly the SPL_{eq} tends to get slightly smaller than SPL_{mean}. In loud speech, on the other hand, the SPL_{eq} values become slightly higher than the SPL_{mean} values. The figure also shows the equivalent SPLs of speech for various efforts in accordance with the ISO 9921-1 standard [ISO 9921-1:1996(E), also Sala *et al.*, 2002]. (The effort SPL_{eq} values were given by ISO for 1-m distance; the ISO values in Fig. 10 are 10 dB higher to relate them to the 30-cm distance used here.) While the standard ISO values were determined for the dB(A) levels, and thus comparisons with our linearly weighted dB levels (which are better suited for vocal dose calculations, Švec *et al.*, 2003a) shall be taken with caution, it is interesting to note that the soft speech values of our subjects (circles) fall mostly in the relaxed ISO category, the comfortable speech values (squares) are placed mainly in the normal effort ISO category, and the loud speech values are distributed within the raised and loud ISO categories, as expected. Values belonging to the very loud [88 dB(A)], shout [94 dB(A)], and maximum shout [100 dB(A)@30 cm] categories recognized by ISO were not reached in this study.

The data indicate strong correlation between the SPL_{mean} and the SPL_{eq} (multiple correlation coefficient $R=0.993$, calculated by SIGMA PLOT 8.0). The relationship can be expressed by the formula

$$SPL_{eq} = 1.25\ SPL_{mean} - 16.42. \tag{10}$$

The formula is applicable for a *monologue* speech signal segmented into 30-ms frames. For frame durations different than 30 ms and for a speech with considerable pauses (dialogue, etc.), the formula would likely need some modification. Despite its limitations, the formula can be helpful in some occupational studies, e.g., in approximately relating the mean SPLs estimated by voice dosimeters to vocal effort categories.

B. Overall discussion, conclusions

The results reveal that (a) the skin acceleration levels generally correlate with sound pressure levels, i.e., increase in SPL is causing an increase in SAL (as indicated by the best-fit relationship), but simultaneously (b) there is considerable variability of the values around the best-fit relationship (Fig. 2). The best-fit relationship can be used to estimate the SPL of voice from SAL. The variation around the best fit, however, limits the accuracy of the estimation of the instantaneous SPLs to, on average, about ± 6 dB in males and ± 5 dB in females (Fig. 3).

The estimations of the mean and equivalent SPLs of speech are more accurate than the estimations of the instantaneous SPLs because the acoustic differences among speech sounds (vowels, voiced consonants, varying fundamental frequency) average out. The most accurate are the estimates of SPL_{mean} (better than ± 2.8 dB with 95% confidence, Fig. 7, left). The estimates of the SPL_{eq} were found to be up to 1.5 dB less accurate than the SPL_{mean} (Fig. 7, right). The definitions of the mean and equivalent SPLs indicate that the SPL_{eq} is more sensitive to isolated SPL peaks occurring in speech than the SPL_{mean}. The decreased accuracy in estimating the SPL_{eq} can thus hypothetically also be attributed to the limitations of the accelerometer signal to reflect the sudden dynamic changes in the acoustic signal (recall the loud-speech outliers in Fig. 2). The SPL_{mean} and SPL_{eq} estimation accuracy worsened only by up to 0.8 dB when the Marvin Williams passage levels were derived from the Rainbow passage best fits, which suggests that the speech content is not critical here (Fig. 9 vs Fig. 7).

First- and second-order SAL/SPL relationships were studied and compared in terms of their accuracy. Whereas the second-order relationship appeared to provide a slightly better fit visually (Fig. 2), not much difference was found between the accuracy of the two fits quantitatively—only about 0.25 and 0.5 dB smaller bias of the second-order fit over the first-order fit for the SPL_{mean} and SPL_{eq}, respectively, in loud speech (Fig. 7). Both the fits were on average slightly overestimating the SPLs of the soft speech while slightly underestimating the loud SPLs in the majority of the subjects, which suggests that a better alternative relationship

between SAL and SPL may be found in the future. The biases can, in principle, be reduced and the estimation accuracy improved using the regression relationship shown by the dashed line in Fig. 4. The corrected estimated SPLs would be calculated as

$$\mathrm{SPL}_{\mathrm{est}} = 1.088\ \mathrm{SPL}_{\mathrm{FIT}} - 5.9, \tag{11}$$

where $\mathrm{SPL}_{\mathrm{FIT}}$ are the values obtained from Eqs. (3) or (4). The regression coefficients in Eq. (11) were found to be very close for the first- vs second-order fits (1.0876 vs 1.0884 and 5.85 vs 5.93).

In traditional SPL measurements of voice with a sound-level meter positioned at the distance of 30 cm (as recommended by UEP, Schutte and Seidner, 1983) ca. ±2-dB variation can be expected when a subject moves 5 cm towards and away from the sound level meter. On the basis of this finding, accuracy of ±2 to ±3 dB can be considered sufficient for SPL measurement of voice and speech (Schutte, 2000). The accuracy of ±2.8 dB observed here for the mean SPL levels (Figs. 7 and 9) compares well to the accuracy of traditional SPL measurements. Therefore, in situations when the use of a microphone is inconvenient, an accelerometer can be considered as a viable alternative sensor for determining long-term average SPL values of speech. Individual calibration of each subject is, however, necessary in order to achieve the accuracy reported here, as differences in skin transmission characteristics can exist among different subjects (Cheyne *et al.*, 2003). In field measurements other factors (such as, e.g., long-term stability and repeatability of the accelerometer attachment) must also be considered, but these will be treated in another study.

ACKNOWLEDGMENTS

The work was supported by the National Institutes of Health, Grant Number DC RO1 04224-01. The authors appreciate the help of Andrew Starr, Hana Švecová, and Eric Hunter in data collection and processing, the consultation of Karen Kadafar of Department of Mathematics, University of Colorado at Denver on statistical analysis of the data, the advice of Robert Hillman of MEEI in Boston on the choice of the accelerometer, and the comments of three anonymous reviewers and the associate editor A. Löfqvist on the manuscript. The paper was presented at the International Conference on Voice Physiology and Biomechanics (ICVPB) in Marseille, France, August 18–20, 2004.

Adams, T. M. (**2002**). *A2LA guide for the Estimation of Measurement Uncertainty in Testing* (The American Association for Laboratory Accreditation, Frederick, MD). Available online from http://www.a2la2.net/guidance/est_mu_testing.pdf

ANSI (**1994**). S1.1-1994. "Americal National Standard: Acoustical Terminology" (American National Standards Institute, New York).

Bastian, R. W., Keidar, A., and Verdolini-Marston, K. (**1990**). "Simple vocal tasks for detecting vocal fold swelling," J. Voice **4**(2), 172–183.

Cheyne, H. A. (**2002**). "Estimating glottal voicing source characteristics by measuring and modeling the acceleration of the skin on the neck," Doctoral dissertation. Massachusetts Institute of Technology.

Cheyne, H. A., Hanson, H. M., Genereux, R. P., Stevens, K. N., and Hillman, R. E. (**2003**). "Development and testing of a portable vocal accumulator," J. Speech Lang. Hear. Res. **46**, 1457–1467.

Coleman, R. F. (**1988**). "Comparison of microphone and neck-mounted accelerometer monitoring of the performing voice," J. Voice **2**(3), 200–205.

Fairbanks, G. (**1960**). *Voice and Articulation Drill Book* (Harper and Row, New York).

Hess, W. (**1983**). *Pitch Determination of Speech Signals: Algorithms and Devices* (Springer, Berlin-Heidelberg-New York-Tokyo).

Hillman, R., and Cheyne, H. (**2003a**). "A portable vocal accumulator with biofeedback capability," in *AQL 2003 Hamburg: Proceeding Papers for the Conference Advances in Quantitative Laryngology, Voice and Speech Research*, edited by G. Schade, F. Müller, T. Wittenberg, and M. Hess (IRB, Stuttgart, Germany), available online from http://www.uke.uni-hamburg.de/AQL2003

Hillman, R., and Cheyne, H. (**2003b**). "Development of a portable voice monitor with biofeedback capability," Perspect. Voice Voice Disord. **13**, 23–25.

Horii, Y. (**1982**). "Jitter and shimmer differences among sustained vowel phonations," J. Speech Hear. Res. **25**(1), 12–14.

Horii, Y. (**1983**). "An accelerometric measure as a physical correlate of perceived hypernasality in speech," J. Speech Hear. Res. **26**(3), 476–480.

ISO (**1996**). 9921-1:1996(E) international standard. Ergonomic assessment of speech communication. Part 1: Speech interference level and communication distances for persons with normal hearing capacity in direct communication (SIL method). (International Organization for Standardization, Geneva), pp. 1–10.

Lippmann, R. P. (**1981**). "Detecting nasalization using a low-cost miniature accelerometer," J. Speech Hear. Res. **24**, 314–317.

Pešák, J. (**1990**). "Differentiated laryngeal phonation," Folia Phoniatr. **42**, 296–301.

Popolo, P. S., Švec, J. G., Hunter, E. J., Rogge-Miller, K., Nix, J., and Titze, I. R. (**2004**). "Teacher's Guide to Voice Dosimetry," NCVS Online Technical Memo, No. 4, 1-32. Available online from http://www.ncvs.org

Sakakura, A., and Takahashi, H. (**1988**). "Body wall vibration in trained and untrained voices," in *Vocal Physiology: Voice Production, Mechanisms, and Functions*, edited by O. Fujimura (Raven, New York), pp. 391–401.

Sala, E., Airo, E., Olkinuora, P., Simberg, S., Ström, U., Laine, A., Pentti, J., and Suonpää, J. (**2002**). "Vocal loading among day care center teachers," Logoped. Phoniatr. Vocol. **27**(1), 21–28.

Sataloff, R. T. (ed.) (**1997**). *Professional Voice: The Science and Art of Clinical Care*, 2nd ed. (Singular, San Diego).

Schutte, H. K. (**2000**). "The phonetogram: measurement and interpretation," in *Diseases of the Larynx*, edited by A. Ferlito (Arnold/Oxford University Press, London, Great Britain/New York, NY), pp. 145–149.

Schutte, H. K., and Seidner, W. (**1983**). "Recommendation by the Union of European Phoniatricians (UEP): Standardizing voice area measurement/phonetography," Folia Phoniatr. **35**, 286–288.

Stevens, K., Kalikow, D., and Willemain, T. (**1975**). "A miniature accelerometer for detecting glottal waveforms and nasalization," J. Speech Hear. Res. **18**, 594–599.

Sundberg, J. (**1983**). "Chest wall vibrations in singers," J. Speech Hear. Res. **26**, 329–340.

Sundberg, J. (**1992**). "Phonatory vibration in singers: A critical review," Music Percept. **9**, 361–382.

Švec, J. G., Popolo, P. S., and Titze, I. R. (**2003a**). "Measurement of vocal doses in speech: Experimental procedure and signal processing," Logoped. Phoniatr. Vocol. **28**, 181–192.

Švec, J. G., Titze, I. R., and Popolo, P. S. (**2003b**). "Vocal dosimetry: Theoretical and practical issues," in *AQL 2003 Hamburg: Proceeding Papers for the Conference Advances in Quantitative Laryngology, Voice and Speech Research*, edited by G. Schade, F. Müller, T. Wittenberg, and M. Hess (IRB, Stuttgart, Germany), available online from http://www.uke.uni-hamburg.de/AQL2003

Švec, J. G., Hunter, E. J., Popolo, P. S., Rogge-Miller, K., and Titze, I. R. (**2004**). "The calibration and setup of the NCVS dosimeter," NCVS Online Technical Memo, No. 2, April 2004, 1–52. Available online from http://www.ncvs.org/ncvs/library/tech

Titze, I. R., Švec, J. G., and Popolo, P. S. (**2003**). "Vocal dose measures: Quantifying accumulated vibration exposure in vocal fold tissues," J. Speech Lang. Hear. Res. **46**(4), 922–935.

The relative importance of spectral tilt in monophthongs and diphthongs

Michael Kiefte[a)]
School of Human Communication Disorders, Dalhousie University, Halifax, Nova Scotia B3H 1R2, Canada

Keith R. Kluender
Department of Psychology, University of Wisconsin–Madison, Madison, Wisconsin 53706

(Received 11 April 2003; revised 8 December 2004; accepted 20 December 2004)

Ito *et al.* [J. Acoust. Soc. Am. **110**, 1141–1149 (2001)] demonstrated that listeners can reliably identify vowel stimuli on the basis of relative formant amplitude in the absence of, or in spite of, F_2 peak frequency. In the present study, formant frequencies and global spectral tilt are manipulated independently in synthetic steady-state vowels. Listeners' identification of these sounds demonstrate strong perceptual effects for both local (formant frequency) and global (spectral tilt) acoustic characteristics. Subsequent experiments reveal that effects of spectral tilt are attenuated in synthetic stimuli for which formant center frequencies change continuously. When formant peaks are kinematic, perceptual salience of the relative amplitudes of low- and high-frequency formants (as determined by spectral tilt) is mitigated. Because naturally produced English vowels are rarely spectrally static, one may conclude that gross spectral properties may play only a limited role in perception of fluently produced vowel sounds.
[DOI: 10.1121/1.1861158]

PACS numbers: 43.71.Es, 43.71.An [PFA] Pages: 1395–1404

I. INTRODUCTION

An enduring, but not dominant, perspective concerning vowel perception emphasizes the putative efficacy of gross spectral properties versus local spectral prominences such as formant peaks (e.g., Bladon and Lindblom, 1981; Bladon, 1982). By contrast, most contemporary accounts of vowel perception emphasize almost exclusively the role of formants, and there is a large body of evidence in support of the ''formant hypothesis'' of vowel perception (see Rosner and Pickering, 1994, for a review).

Rekindling prospects for an explanation based upon global spectral properties, Ito *et al.* (2001) demonstrated that spectral tilt, or the relative amplitudes of high- to low-frequency components, can be at least as effective as second formant (F_2) frequency in signaling vowel quality. Ito *et al.* asked native speakers of Japanese to identify isolated, synthetic steady-state vowels from which the F_2 peak was excised while amplitudes of remaining formant peaks were preserved. Their finding that subjects' identifications of these stimuli were not markedly different from identifications of similar stimuli, in which all formant peaks were retained, demonstrated that a spectral prominence corresponding to F_2 is not necessary for vowel identification. Ito *et al.* also showed that listeners were able to identify the full set of five Japanese vowels when the F_2 peak was present, but fixed in frequency across a stimulus set for which only F_1 frequency and spectral tilt were manipulated.

The only spectral properties manipulated in the vowels used by Ito *et al.* (2001) were the amplitudes of higher formants which are normally determined by formant bandwidth and the frequencies of other formants in cascade synthesis (Klatt, 1980). In particular, lowering or raising the frequency of one formant lowers or raises the amplitudes of all higher formants, respectively. Therefore, the original frequency of the excised formant peak influenced the relative spectral balance between low- (e.g., F_1) and high-frequency spectral energy (e.g., F_3).

These results lend support to the whole-spectrum hypothesis of Bladon and Lindblom (1981), who suggest that gross spectral properties such as auditory spectral density contain more phonetically relevant information for vowel quality than formant frequencies alone. However, the results of Ito *et al.* also run counter to previous observations regarding the importance of formant amplitude in perception. For example, Assmann (1991) found that manipulation of F_1 or F_2 amplitude has only a small or nonsignificant effect on perception. Additionally, Klatt (1982) has shown that changes in spectral tilt or formant bandwidth do not produce a change in perceived vowel identity, but are instead heard as a change in speaker or transmission channel characteristics. Even theories that suggest that formant amplitude does have an effect on the perceptual integration of formants within 3–3.5 bark, such as the F_2-Prime hypothesis (e.g., Chistovich, 1985; Beddor and Hawkins, 1990) depend critically on the *presence* of spectral prominences corresponding to resonant frequencies of the vocal tract. Such theories are therefore unable to explain the importance of global spectral tilt in the absence of F_2 peak as found in the study by Ito *et al.*

Although Ito *et al.* (2001) show that listeners can identify vowels reliably on the basis of spectral slope for stimuli without F_2 peak, it is known that the converse is also true. Subjects are able to identify vowels from formant peaks without relative amplitude information (e.g., Pattern Playback speech, Delattre *et al.*, 1952), or even when vowel for-

a)Electronic mail: mkiefte@dal.ca

mants are represented by nonharmonically related sinusoids (e.g., Hose *et al.*, 1983). In addition, one important factor may have affected the validity of the results of Ito *et al.*: only four subjects were recruited for the study and each subject participated in all of the experiments described. Moreover, the result may not be generalizable to another language with a vowel system with a larger number of vowel contrasts, such as English.

Based on the above-cited experiments, both gross spectral shape and spectral prominences such as formant peaks likely play important roles in vowel perception especially when one cue is presented in the absence of another. In this paper, we present results from a *conflicting-cue* experiment for vowel identification in which both formant frequency and global spectral tilt are manipulated independently. These experiments were conducted to evaluate the relative importance of detailed (e.g., formant peaks) and gross spectral cues such as spectral tilt (i.e., relative spectral balance between low- and high-frequency energy).

Listeners' identifications in the first experiment reveal a complex relationship between perception of detailed (formant frequencies) and global (spectral tilt) attributes of the vowel spectrum in synthetic steady-state vowels. Because naturally produced vowels are typically characterized by spectral change over time (Hillenbrand *et al.*, 1995) and this kinematic information is known to be important to the perception of vowel sounds (Hillenbrand and Nearey, 1999), experiments 2 and 3 investigate the perception of diphthongs. Results from these experiments show that perceptual effects of spectral tilt are mitigated in stimuli in which formant frequencies change over time.

II. EXPERIMENT 1

The purpose of this experiment is to examine the relative contribution of formant frequencies and spectral tilt in a simple two-dimensional vowel matrix in which these two properties are manipulated independently.

A. Method

Subjects were asked to identify isolated, synthetic steady-state vowels drawn from a matrix of stimuli varying perceptually from [i] to [u] and varying acoustically in formant frequencies and gross spectral tilt.

1. Subjects

Fourteen native speakers of Midwestern American English were recruited from the Department of Psychology at the University of Wisconsin–Madison. None reported any hearing impairment. Subjects received course credit for their participation.

2. Stimuli

A seven-step series ranging from [i] to [u] was generated using the cascade branch of a flexible implementation of Klatt's 1980 speech synthesizer in MATLAB (Klatt, 1980; Kiefte *et al.*, 2002). Vowel stimuli were 245 ms in duration. Formant values for series' end points are similar to those for adult native speakers of American English as reported by Hillenbrand *et al.* (1995) and each end point was judged to be a reasonable approximation of either [i] or [u]. F_2 ranged from 1000 to 2200 Hz in 200-Hz steps, while F_3 varied as a linear function of F_2 between 2400 and 3000 Hz in 100-Hz steps to encourage subjective naturalness. F_1 was fixed at 300 Hz while F_4 and F_5 were held constant at 3500 and 4500 Hz, respectively. Fundamental frequency (f_0) was set at 100 Hz so formant peaks corresponded to harmonics. Voicing amplitude (AV) was set at 60 dB and was reduced to 0 dB for the 5-ms frame starting at 240 ms. All remaining synthesis parameters were maintained at default values (Klatt, 1980). Figures 1(b) and (c) show the [i] and [u] end points of this continuum, respectively.

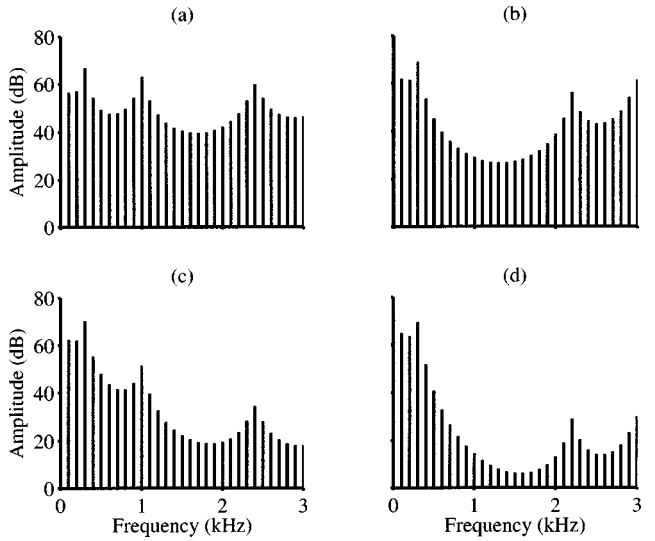


FIG. 1. Four extrema of the stimulus spectra. (a) [u] with [i] spectral tilt; (b) [i] with natural spectral tilt; (c) [u] with natural spectral tilt; (d) [i] with [u] spectral tilt.

Spectral tilt was fully crossed with formant frequencies. The spectral tilt for each vowel was adjusted to match that of every other vowel in the series for a total of $7\times7=49$ stimuli. For the purposes of this study, spectral tilt was defined as the dB/octave difference between F_1 and F_3 peak amplitude. Amplitudes were determined analytically based on the source and cascade synthesis parameters for the peak harmonic for both F_1 and F_3. The dB/octave difference was imposed on the target stimulus via a filter similar to the one-pole filter used by the KLSYN88 speech synthesizer to adjust source tilt (Klatt and Klatt, 1990). A single-zero filter was used instead of a pole in those cases where spectral tilt needed to be *increased* instead of lowered (e.g., [u] formants with [i] spectral tilt—see Fig. 2). The bandwidth of the filter used to adjust spectral tilt was determined by a simple recursive algorithm based on the desired attenuation or amplification of F_3 relative to F_1. The frequency response of the filter was constrained by $F=0.375$ BW—where F is the center frequency of the pole or zero, and BW is the bandwidth—in order to achieve critical damping (Klatt and Klatt, 1990). All stimuli were normalized in rms amplitude. The four extreme points of the stimulus matrix are illustrated in Fig. 1 and the procedure for crossing the spectral tilt is illustrated in Fig. 2.

3. Procedure

Subjects listened to stimuli via headphones in a sound attenuated booth. Stimulus presentation and response collec-

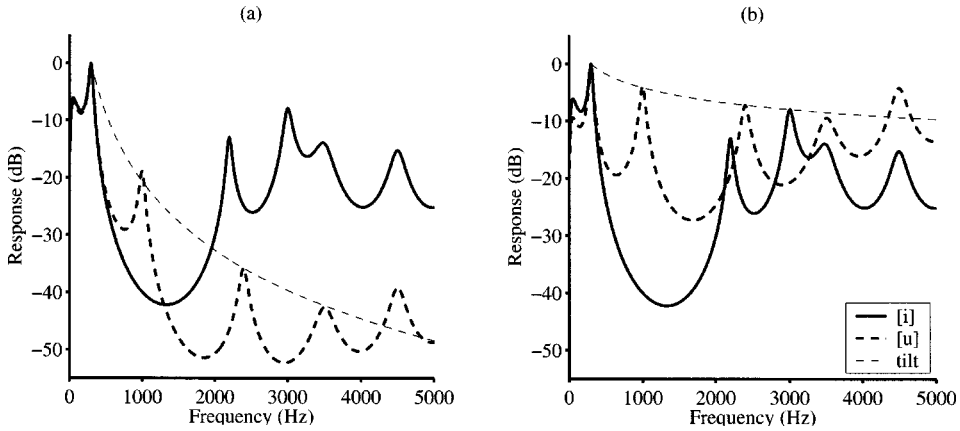


FIG. 2. Stimulus generation. Hybrid stimuli were generated by setting the spectral tilt of one vowel to the natural spectral tilt of another. In this example, the spectral tilt of [u] is raised to match that of [i] to generate a hybrid stimulus. The left panel shows the spectral envelopes of [i] and [u] while the dashed line indicates the spectral tilt of normally synthesized [u]. The right panel shows the spectrum of [u] after the spectral tilt has been raised to match that of [i]. For the purposes of this experiment, spectral tilt was defined as dB/octave between F_1 and F_3 peaks. On the right panel, the spectral tilt curve passes through the F_3 peaks of both the [i] and [u] spectral envelopes.

tion were under the control of an 80486-25 microcomputer. Following D/A conversion (Ariel DSP-16), stimuli were low-pass filtered (4.8-kHz cutoff frequency, frequency Devices, #677), amplified (Stewart HDA4), and presented to subjects via headphones (Beyer DT-100) at a level of 75 dB SPL. Subjects identified each vowel by pressing buttons labeled either ''ee'' or ''oo''. Each of the 49 vowels was presented eight times in each of two 15-min sessions for a total of 784 responses from each subject.

B. Results

Figure 3 illustrates responses pooled across subjects for each of the 49 stimuli as a function of both F_2 frequency and spectral tilt in dB/octave.

Responses were modeled for each subject via logistic regression (McCullagh and Nelder, 1989) with both spectral tilt and F_2 frequency as predictors of the probability of [i] response. Because of the quadratic nature of the category boundary, squares and cross-products of the predictors were included in the model.

The 50% boundaries for the 14 subjects based on the regression fits are shown by dashed lines in Fig. 4 indicating a large amount of between-subject variation, while the fitted boundary based on the aggregated data is given by the solid line. Both spectral tilt and formant frequencies appear to have strong perceptual effects jointly and separately for these stimuli.

To evaluate the importance of either spectral tilt or formant frequency in listeners' responses, Hotelling's T^2 statistic, or multivariate *t*-test for one-sample repeated measures, was calculated for the random subject regression coefficients that include appropriate terms in the model (Davis, 2002; Gumpertz and Pantula, 1989). For example, tests for spectral

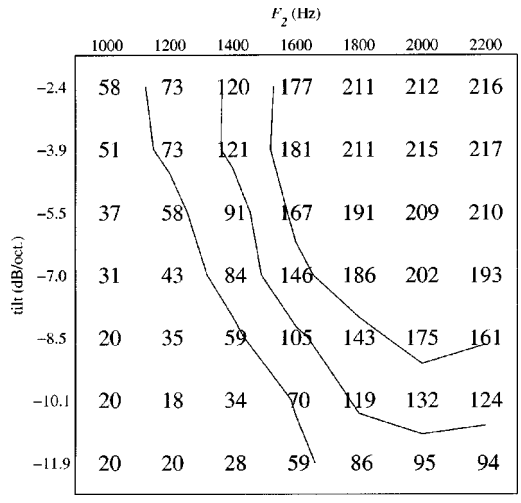


FIG. 3. [i] responses pooled across speakers (out of a maximum of 224) from experiment 1. Contour lines indicate 30%, 50%, and 70% cutoffs. Steps along the ordinate are not evenly spaced as they correspond to stimulus steps (which increase linearly in F_2 and F_3 but nonlinearly in spectral tilt).

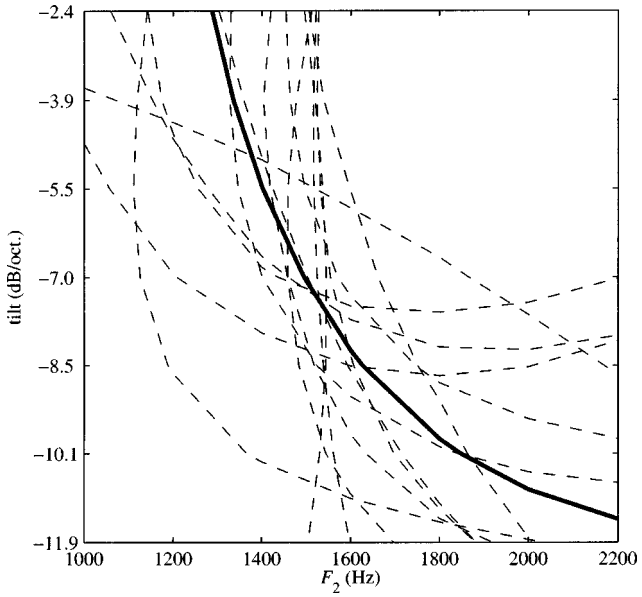


FIG. 4. Fitted regression boundaries. Dashed lines indicate fitted 50% cutoffs for each subject. The solid line indicates fitted boundary for aggregated data.

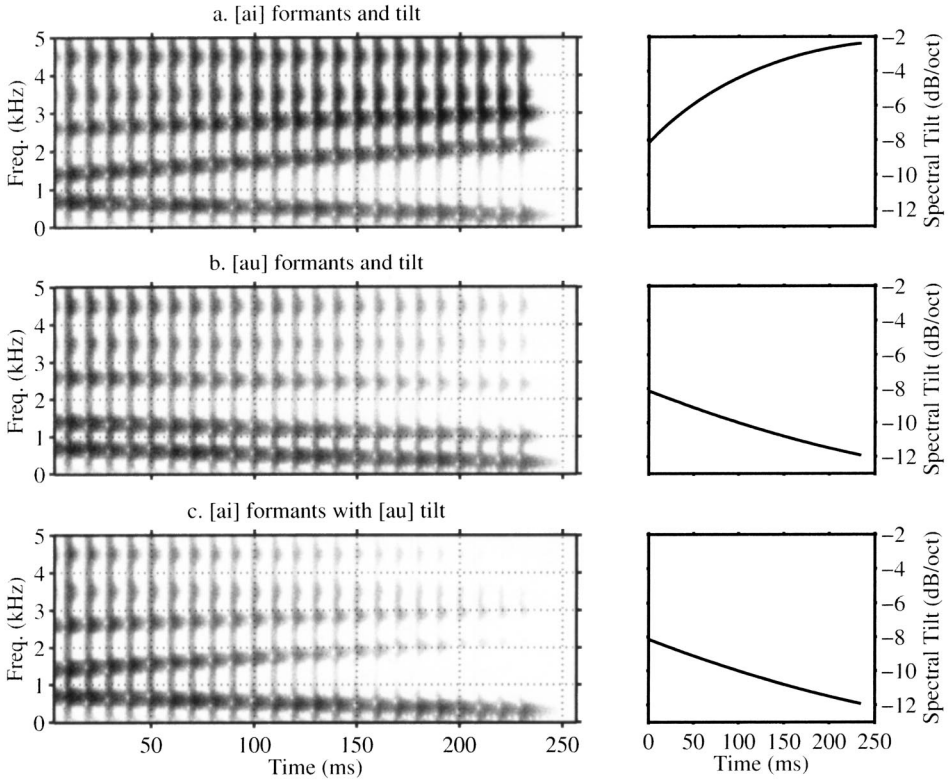


FIG. 5. Two unaltered and one hybrid stimuli from experiment 2. (a) and (b) Spectrograms from the original [ai] and [au] stimuli with unaltered spectral tilt. (c) A spectrogram of a stimulus with formants appropriate for [ai] but with spectral tilt appropriate for [au]. Plots on the right-hand side show spectral tilt as a function of time. The right-hand plot corresponding to (c) shows the that the spectral tilt of [ai] has been matched to that of [au] for each 5 ms frame.

tilt include coefficients from each subject for the square of the tilt parameter, the cross product of tilt and formant frequency, as well as the simple tilt coefficient. The effect for spectral tilt was significant ($F_{3,11}=7.67; p<0.01$) as was the effect for formant frequency ($F_{3,11}=21.41; p<0.001$). The effect size D^2, or Mahalanobis distance (Stevens, 1992), for tilt was 1.94 while effect size for formant frequency was much stronger ($D^2=5.42$).

C. Discussion

Results from the previous experiment showed that both formant frequencies and global spectral tilt contribute to identification of [i] and [u] by listeners. To further evaluate the relative importance of spectral tilt and formant frequencies on vowel perception, a second experiment was designed using the diphthongs [ai] and [au] in a similar conflicting-cue paradigm. If spectral tilt is an important acoustic cue to vowel identity in perception, then this effect should also be observed for more representative kinematic vowel stimuli.

III. EXPERIMENT 2

Experiment 2 is otherwise identical to the first with the exception that formants are kinematically specified. Aside from being representative of English diphthongs, stimuli from experiment 2 are also more representative of phonologically monophthongal English vowels which are produced more generally with continuously changing spectral characteristics (Hillenbrand *et al.*, 1995) that have been shown to be important for perception (Hillenbrand and Nearey, 1999; Nearey and Assmann, 1986).

Figures 5(a) and (b) show spectrograms of the synthetic diphthongs [ai] and [au] for which the offsets are spectrally identical to the extreme [i] and [u] from the first experiment, respectively. Spectral slope as a function of time is indicated on the right-hand side of the figure. Figure 5(c) shows a spectrogram of a diphthong having the same formant transitions as Fig. 5(a) but with a dynamically specified tilt matching that of the stimulus illustrated by Fig. 5(b).

The second experiment was designed to evaluate the relative importance of spectral tilt and formant frequencies in diphthongs, testing the hypothesis that spectral tilt is an important spectral property for vowel stimuli that are characterized by spectral change over time.

A. Method

1. Subjects

Twelve subjects were recruited in the same manner as described in Sec. I and also met the same criteria. No subject participated in both experiments.

2. Stimuli

A seven-step series of 245-ms diphthongs ranging from [ai] to [au] were generated with offset formant frequencies identical to those of steady-state stimuli described in Sec. II. Parameters at the 5-ms frame starting at 235 ms were set to these values. Onset formant frequencies were 700, 1400, and 2600 Hz for F_1, F_2, and F_3 frequency, respectively, and these values changed linearly to the offset values over 235 ms. All remaining parameters were held constant throughout the duration of the stimuli.

Each of these seven stimuli was crossed with every other stimulus to form a 7×7 matrix—i.e., each new stimulus was produced by taking the spectral tilt of one of the original seven diphthongs and matching it, frame by frame, with an-

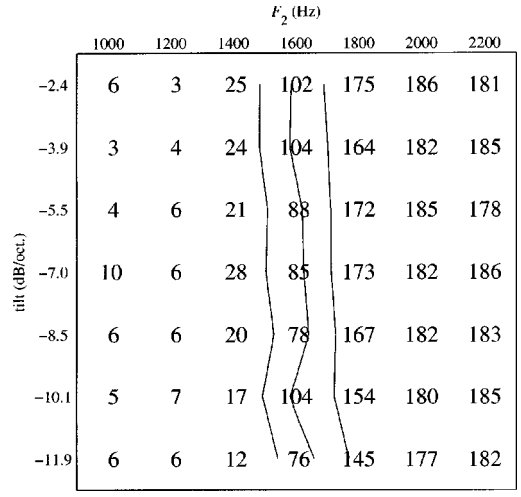


FIG. 6. [ai] responses pooled across speakers (out of a maximum of 192). Contour lines are same as those in Fig. 3. Labels along the ordinate indicate the spectral tilt at the *offset* of the diphthongs.

other of the original seven for a total of 49 stimuli. In this manner, the spectral tilt was matched for each 5-ms synthesis frame: at each frame, the tilt of stimulus *A* was either raised or lowered to match that for stimulus *B* at the same time frame. Figure 5 shows spectrograms of three stimuli from this matrix along with plots of spectral tilt as a function of time.

3. Procedure

The procedure is the same as described in Sec. III with the exception that subjects were asked to choose between buttons labeled ''high'' (/ai/) and ''how'' (/au/) following each stimulus presentation.

B. Results

Figure 6 displays responses for experiment 2 pooled across speakers as a function of both formant frequency and spectral tilt at stimulus offset. Hotelling's T^2 statistics for one-sample repeated measures were calculated for the regression coefficients for each subject estimated via quadratic logistic regression. Using the same statistical procedure described in experiment 1, it was found that formant frequency was significant ($F_{3,9}=16.06$; $p<0.001$; $D^2=4.91$). However, spectral tilt did not reach statistical significance ($F_{3,9}=2.25$; $p\approx0.15$). Figure 7 gives individual fitted category boundaries as well as the estimated boundary from the aggregate data.

Additional Hotelling's T^2 statistics were computed to compare the coefficients obtained in this experiment with those in experiment 1. Coefficients for both formant frequency ($F_{3,22}=8.6$; $p<0.001$; $D^2=4.35$) and spectral tilt ($F_{3,22}=3.93$; $p=0.02$; $D^2=1.99$) were found to be significantly different between those obtained in experiments 1 and 2.

C. Discussion

Unlike results from the first experiment, effects of spectral tilt were no longer statistically significant despite the fact that spectra of stimulus offsets in this experiment were identical to those of steady-state stimuli used in experiment 1. Effects of spectral tilt were substantially reduced when local spectral prominences (formants) changed continuously throughout the stimuli relative to the effects observed in experiment 1. Conversely, it is also possible that the effects of spectral tilt may be substantially enhanced in experiment 1 relative to the present experiment.

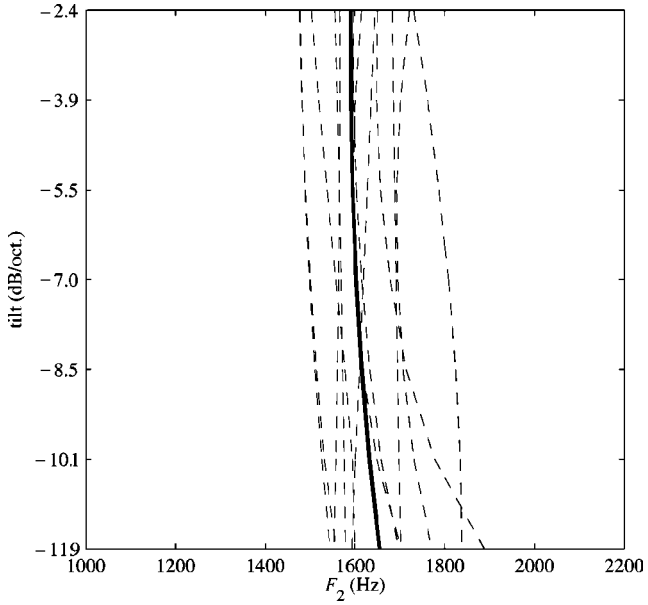


FIG. 7. Fitted regression boundaries. Dashed lines indicate fitted 50% cutoffs for each subject. The solid line indicates fitted boundary for aggregated data.

There are several possible reasons why the perceptual dependence on spectral tilt was attenuated in experiment 2. First, the experiments used stimuli representing completely different phonemes, and it is conceivable that relative weighting of spectral tilt and formant peaks varies between stimulus sets.

Second, in the case of the [ai]-[au] continuum, it is also possible that listeners base their decisions on a ''snapshot'' of the stimulus close to the onset thereby reducing the dependence on slowly changing spectral tilt. Experiment 3 addresses these concerns by using a two-dimensional matrix of stimuli with kinematic spectral properties as in experiment 2, but stimuli are designed by interpolating spectral characteristics of those in experiment 1.

IV. EXPERIMENT 3

The final experiment employs a hybrid monophthong/diphthong series sharing characteristics of stimuli from both experiments 1 and 2 in order to more clearly define joint and separate effects of spectral tilt and kinematic spectral change. Offsets of stimuli in this experiment are identical to those of stimuli in experiment 2 as well as to the steady states of the monophthong stimuli in experiment 1. However, onset synthesis parameters were set at those appropriate for [u]. Stimuli ranged between monophthong [u] and diphthong [ui] in both formant frequencies and spectral tilt.

The spectrum of each frame of each stimulus is effectively an interpolation of the steady-state spectra of stimulus

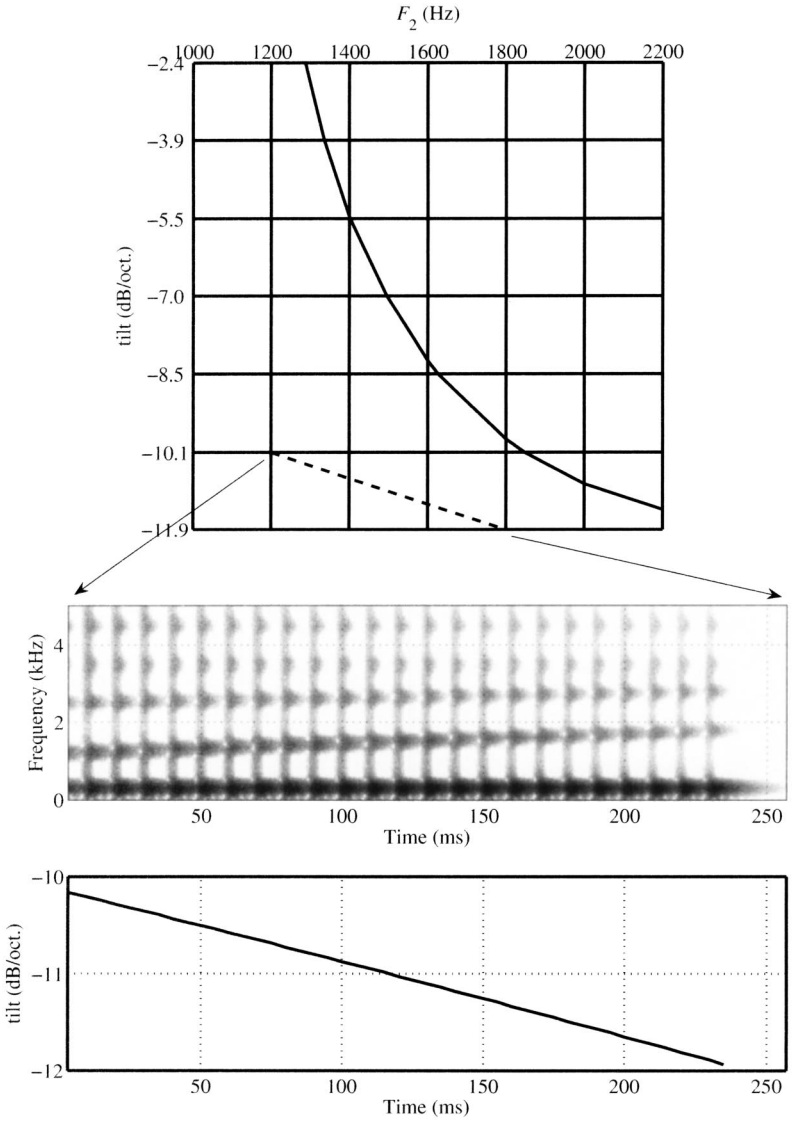


FIG. 8. Example stimulus from experiment 3. The solid line in the top panel gives the 50% cutoff in [i] responses from experiment 1 while the dashed line gives the spectral-tilt/second-formant trajectory for one stimulus with onset F_2 at 1200 and offset at 1800 Hz. The spectrogram for this stimulus is given in the middle panel while the spectral tilt as a function of time is given in the bottom panel. For this stimulus, spectral tilt is made to decrease from −10.1 to −11.9 dB/octave. At all points along the trajectory indicated by the dashed line in the topmost panel, acoustically similar stimuli received a majority of [u] responses from listeners in experiment 1.

matrix used in experiment 1 which ranged from [u] to [i] in both formant frequencies and spectral tilt. The grid in Fig. 8 indicates the locations of each of the 49 steady-state stimuli from experiment 1 in spectral tilt and formant frequency, while the solid curve shows the 50% crossover for /i/ responses to those stimuli. Stimuli in the present experiment with continuously changing F_1 and F_2 frequencies can be visualized as vectors on this grid, each originating at $F_2 = 1200$ Hz and −10.1 dB/oct spectral tilt (stimulus onset) and ending at one of the 49 locations on the grid (stimulus offset). The change in spectral tilt and formant frequency for one stimulus in experiment 3 is shown by the dashed line. If it is the case that spectral tilt plays an important role in perception of English vowels, then the stimulus illustrated by the dashed line should be categorized as /u/ because it does not cross the 50% contour line from experiment 1, i.e., each intermediate time-frame of the diphthong, if presented as a steady-state vowel in the first experiment, would be identified mostly as /u/ by listeners.

However, if spectral tilt is not an important spectral property in the perception of vowels with dynamic formant trajectories, then we expect that listeners will ignore variation in spectral tilt.

A. Subjects

Fourteen subjects were recruited in the same manner and met the same criteria as described in Sec. I. No subject in this experiment participated in either of the first two experiments.

B. Stimuli

Stimuli were designed similarly to those in experiment 2 with the exception that onset parameters for F_1, F_2, and F_3 were set at 300, 1200, and 2500 Hz, respectively. Thus, onset F_2 and F_3 were set at 200 and 100 Hz above the lowest values at the [u] end of the original steady-state series. This gave stimuli at the lowest-F_2 frequency end of the series a modestly diphthongal quality. The onset is illustrated by the leftmost point of the dashed line in Fig. 8.

Each of these stimuli was crossed with every other stimulus in the manner described in experiment 2; spectral tilt was matched between stimuli for every 5-ms frame.

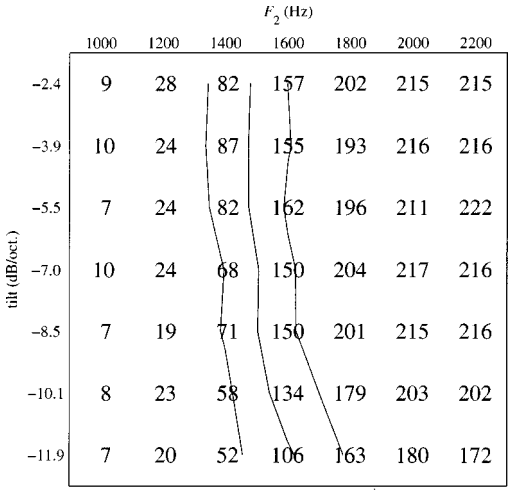


FIG. 9. [ui] responses pooled across speakers (out of a possible 224). Contour lines are the same as those in Figs. 3 and 6. Values indicated along the ordinate and abscissa indicate *offset* parameters.

C. Procedure

The procedure is the same as described in experiments 1 and 2 with the exception that subjects were asked to choose between buttons labeled ''oo'' (/u/) and ''ooey'' (/ui/) following each stimulus presentation.

D. Results

Figure 9 provides aggregate responses for experiment 3 as a function of offset F_2 frequency and global spectral tilt. Hotelling's T^2 statistics were calculated for the quadratic logistic regression coefficients estimated for each subject. Formant frequency was found to be highly significant ($F_{3,11}=58.78$; $p<0.001$; $D^2=14.9$). Spectral tilt was also found to be significant but the effect size was much smaller ($F_{3,11}=4.19$; $p=0.03$; $D^2=1.06$). Based on the individual fitted category boundaries illustrated in Fig. 10, it appears that much of this effect is due to the responses of only two subjects who show a pattern of responses similar to those found in experiment 1. Nevertheless, the effect size for spectral tilt in experiment 3 was much smaller compared with the effect size found in experiment 1 ($D^2=1.94$).

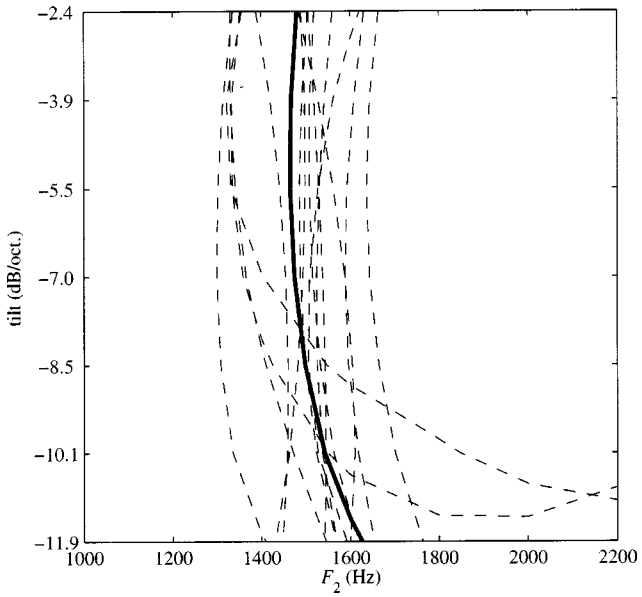


FIG. 10. Fitted regression boundaries. Dashed lines indicate 50% cutoffs for each subject while the solid line indicates fitted boundary for the aggregate data.

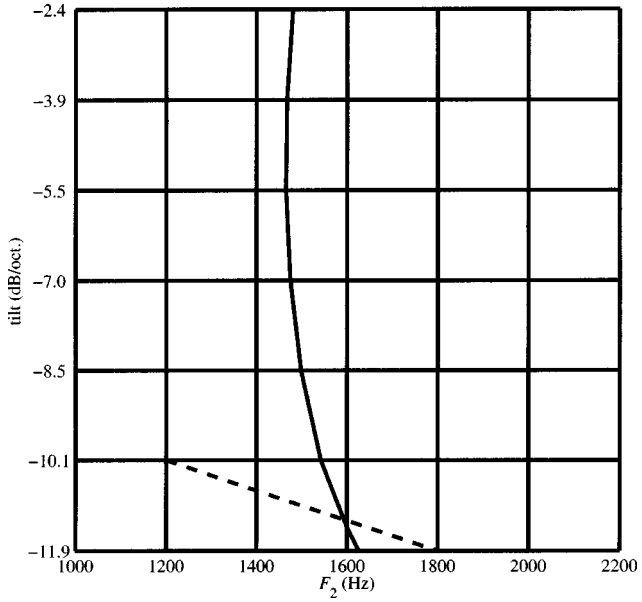


FIG. 11. Spectral-tilt/second-formant trajectory for one stimulus in experiment 3. The trajectory is the same as that illustrated in Fig. 8. Solid curved line indicates 50% boundary from aggregate data in experiment 3.

Coefficients for spectral tilt were found to be significantly different from those obtained in experiment 1 ($F_{3,24}=17.6$; $p<0.001$; $D^2=2.05$), but not for formant frequency ($F_{3,24}=2.97$; $p=0.052$; $D^2=1.38$). Coefficients for formant frequencies were found to be significantly different from those found in experiment 2 ($F_{3,22}=8.6$; $p<0.001$; $D^2=4.35$). This is not surprising because the stimuli in each matrix represent very different speech sounds. Coefficients for spectral tilt were also found to be significantly different between experiments 2 and 3 ($F_{3,22}=2.52$; $p=0.084$; $D^2=1.99$).

E. Discussion

Figure 11 again shows the same spectral-tilt/formant-frequency trajectory as illustrated in Fig. 8 but with the 50% /ui/ response contour from experiment 3. The majority of responses (163 of 224) to this stimulus were /ui/ despite the fact that steady-state stimuli with spectral properties identical to each of the intermediate frames of this kinematic stimulus were identified as /u/ in experiment 1.

Based on results from experiment 1, we would have expected to observe a majority of /u/ responses for sharply falling spectral tilts (−11.9 dB/octave). However, Fig. 11 shows that this was not the case. Likewise, if listeners responded on the basis of a ''snapshot'' close to the onset of the stimuli, we would have observed many more /u/ responses overall given that onset spectral properties were appropriate for /u/.

V. GENERAL DISCUSSION

Experiment 1 demonstrated that *both* spectral tilt and formant peak frequency contribute to discrimination of synthetic steady-state approximations to [i] and [u]. This observation extends the results of Ito *et al.* (2001) who found that listeners could identify vowels on the basis of gross spectral properties alone. Although the present data lend evidence for the importance of both spectral tilt and F_2-peak frequency for vowel perception, results in the second and third experiments suggest that this effect may be limited to stimuli with unchanging spectral characteristics.

Experiment 2 used kinematically specified vowels more typical of naturally produced English vowels with the assumption that listeners could better identify vowels on the basis of *change* in formant frequency relative to static peak frequencies in steady-state monophthongs. Experiment 3 further explored this possibility. Each time frame of each stimulus in the third experiment was interpolated from the stimulus matrix used in the first experiment. However, data from experiment 1, which used stationary vowels, did not predict the pattern of responses to these kinematically specified vowels. The primary difference between the two experiments is that stimuli in experiment 3 were identified on the basis of formant transitions instead of stationary stimulus characteristics.

Results from these experiments show that spectral tilt appears to be much more effective perceptually for steady-state vowels than for vowels with changing spectral properties. Steady-state approximations to English monophthongs, however, are not perceptually or acoustically equivalent to naturally produced English monophthongs. For example, Hillenbrand *et al.* (1995) have shown that almost all vowels in American English, including those that are traditionally viewed as monophthongs such as /ɪ/ and /ɛ/, include substantial changes in formant frequency. Based on the present findings, it is unlikely that spectral tilt plays a dominant role in perception of most naturally produced vowels.

One explanation for the present results may be that formant frequency changes in experiments 2 and 3 are simply more perceptually salient than changes in spectral tilt. From an ecological perspective, this is a reasonable conclusion because changes in spectral tilt often occur independently of linguistic information. For example, spectral tilt varies with changes in transmission channel characteristics such as the change in room acoustics that occurs as one moves from a narrow hallway into a large room. From the point of view of speech perception, these types of changes in spectral information should generally be ignored by the listener because they do not help to convey linguistic meaning. In contrast, there are few natural situations in which frequencies of formant peaks change due to external factors.

Nevertheless, listeners still identified synthetic steady-state vowels on the basis of both spectral tilt *and* formant frequency in experiment 1. Here, we propose that this may be a consequence of unchanging stimuli being perceptually impoverished. If formant peak information is perceptually attenuated for steady state stimuli, then listeners may become more likely to attend to other less reliable properties such as spectral tilt.

Previously, the present authors (Kluender *et al.*, 2003) have noted that, like all sensorineural systems, the auditory system responds most vigorously to change. The fact that perceptual effectiveness of sustained unchanging acoustic energy quickly attenuates has long been known (Hood, 1950). With respect to speech, Summerfield and colleagues' studies of ''auditory afterimages'' demonstrate how local changes in amplitude within the spectrum are auditorily enhanced (Summerfield *et al.*, 1984, 1987). Studies concerning spectral contrast between adjacent coarticulated consonants and vowels (e.g., Holt *et al.*, 2000; Kluender *et al.*, 2003; Lotto *et al.*, 1997; Lotto and Kluender, 1998) provide further evidence that relative change in spectral composition exerts powerful effects on phonetic perception. For these well-established examples of constrastive processes across adjacent spectral regions, effects appear to be spectrally local—altering perceptual effectiveness of energy in spectral regions on the scale of individual formants. For example, Coady *et al.* (2003) showed that contrastive effects of preceding vowels on perception of following stop consonants are maintained when only a few harmonics in the region of F_2 are presented in place of full spectrum vowel sounds.

With respect to neurophysiological processes underlying this sensitivity to change, Delgutte *et al.* (1996) argue that data from recordings in auditory nerve reveal how adaptation serves to enhance spectral contrast between successive speech signals. In contrast, some researchers (e.g., Houtgast, 1974; Moore and Glasberg, 1983) have suggested that adaptation serves mostly to enhance onsets selectively, with suppression being hypothesized to be a process through which differences in level of adjacent spectral regions in complex spectra (e.g., formant peaks and spectral valleys in speech signals) are preserved and/or enhanced. For example, Viemeister and Bacon (1982) note that different frequency components of a signal serve to suppress one another, and suggest that what is relevant is not so much that channels for the signal are unadapted relative to other channels. Instead, they hypothesize that nonsignal channels are adapted such that their ability to suppress the signal is attenuated. This explanation is consistent with studies of two-tone suppression which has been cast as an instance of lateral inhibition in hearing (Houtgast, 1972). Investigators have argued that suppression helps to provide sharp tuning (e.g., Wightman *et al.*, 1977; Festen and Plomp, 1981), and with respect to speech perception, Houtgast (1972) has argued that this process serves to sharpen the neural projection of a vowel spectrum in a fashion that effectively provides formant extraction. Whether one accepts explanations based upon adaptation or adaptation of suppression, the conclusion is that when vowel signals do not change across time, formant peaks lose perceptual effectiveness.

Whatever the underlying neurophysiological or psychoacoustic explanation, the potency of change for perception of speech and other acoustic signals was repeatedly revealed in listeners' almost exclusive reliance upon changing spectral peaks for all non-steady-state stimuli used in the present studies. To the extent that perceptual effectiveness of stationary formants in synthetic steady-state vowels is attenuated following onset, other acoustic information such as spectral

tilt may be expected to play a more significant role. This hypothesis is consistent with data from listeners with hearing impairment, who suffer from signal degradation including spectral smearing, showing that, relative to listeners with normal hearing, spectral tilt plays a greater role in perception of consonants (Lindholm *et al.*, 1988). How listeners come to exploit global spectral attributes such as tilt in the absence of more reliable detailed spectral information is unclear. At this juncture, it is unknown whether tilt provides independent information for phonetic perception, or whether tilt comes to serve perception as a consequence of experienced covariance with putatively more primary acoustic attributes such as formant peaks.

With the exception of stimulus onset and offset, stimuli in experiment 1 include no acoustic change across time. Consequently, listeners were left to perform a task that bears little resemblance to perception of naturally produced speech. If gross spectral properties such as tilt tend to be perceptually effective only when time-varying spectral information is unavailable, then one might suggest that tilt is relatively uninfluential for perception of much natural speech, including English monophthongs that naturally include substantial spectral change across time. One should be cautious, however, before concluding that gross spectral properties never contribute to perception of speech. Some English vowels are more likely to approximate steady-state characteristics, and vowels in other languages (e.g., Spanish) are not produced with as much spectral change as are vowels in English. These facts leave open the possibility that the steady-state vowels used by Ito *et al.* (2001) are more appropriate approximations of natural Japanese vowels than English vowels. To the extent that this is true, data from experiment 1 may well represent relative contributions of spectral tilt and formant peak frequencies for such cases.

Aside from these possible limited exceptions, natural speech varies spectrally nearly continuously across time for all languages. The present findings serve to further emphasize that stimulus change dominates auditory perception of speech and perception most generally.

ACKNOWLEDGMENTS

This work was supported by Grant No. R01 DC04072 to K.R.K. from the National Institutes of Health. We would also like to thank Anthony Watkins and Peter Assmann for their helpful comments on this paper.

Assmann, P. F. (**1991**). ''The perception of back vowels: Centre of gravity hypothesis,'' Q. J. Exp. Psychol. **43A**, 423–428.

Beddor, P. S., and Hawkins, S. (**1990**). ''The influence of spectral prominence on perceived vowel quality,'' J. Acoust. Soc. Am. **87**, 2684–2704.

Bladon, R. A. W. (**1982**). ''Arguments against formants in the auditory representation of speech,'' in *The Representation of Speech in the Peripheral Auditory System*, edited by R. Carlson and B. Granström (Elsevier Biomedical, Amsterdam), pp. 95–102.

Bladon, R. A. W., and Lindblom, B. (**1981**). ''Modeling the judgement of vowel quality differences,'' J. Acoust. Soc. Am. **69**, 1414–1422.

Chistovich, L. A. (**1985**). ''Central auditory processing of peripheral vowel spectra,'' J. Acoust. Soc. Am. **77**, 789–805.

Coady, J. A., Kluender, K. R., and Rhode, W. S. (**2003**). ''Effects of contrast between onsets of speech and other complex spectra,'' J. Acoust. Soc. Am. **114**, 2225–2235.

Davis, C. S. (**2002**). *Statistical Methods for the Analysis of Repeated Measurements* (Springer, New York).

Delattre, P., Liberman, A. M., Cooper, F. S., and Gerstman, L. J. (**1952**). ''An experimental study of the acoustic determinants of vowel color; Observations of one- and two-formant vowels synthesized from spectrographic patterns,'' Word **8**, 195–210.

Delgutte, B., Hammond, B. M., Kalluri, S., Litvak, L. M., and Cariani, P. A. (**1996**). ''Neural encoding of temporal envelope and temporal interactions in speech,'' in *Auditory Basis of Speech Perception*, edited by W. Ainsworth and S. Greenberg, Keele, UK (European Speech Communication Association), pp. 1–9.

Festen, J. M., and Plomp, R. (**1981**). ''Relations between auditory functions in normal hearing,'' J. Acoust. Soc. Am. **70**, 356–369.

Gumpertz, M., and Pantula, S. G. (**1989**). ''A simple approach to inference in random coefficient models,'' Am. Stat. **43**, 203–210.

Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). ''Acoustic characteristics of American English vowels,'' J. Acoust. Soc. Am. **97**, 3099–3111.

Hillenbrand, J. M., and Nearey, T. M. (**1999**). ''Identification of resynthesized /hVd/ utterances: Effects of formant contour,'' J. Acoust. Soc. Am. **105**, 3509–3523.

Holt, L. L., Lotto, A. J., and Kluender, K. R. (**2000**). ''Neighboring spectral content influences vowel identification,'' J. Acoust. Soc. Am. **108**, 710–722.

Hood, J. D. (**1950**). ''Studies in auditory fatigue and adaptation,'' Acta Oto-Laryngol., Suppl. **92**, 1–57.

Hose, B., Langner, G., and Scheich, H. (**1983**). ''Linear phoneme boundaries for German synthetic two-formant vowels,'' Hear. Res. **9**, 13–25.

Houtgast, T. (**1972**). ''Psychophysical evidence for lateral inhibition in hearing,'' J. Acoust. Soc. Am. **51**, 1885–1894.

Houtgast, T. (**1974**). ''Auditory analysis of vowel-like sounds,'' Acustica **31**, 320–324.

Ito, M., Tsuchida, J., and Yano, M. (**2001**). ''On the effectiveness of whole spectral shape for vowel perception,'' J. Acoust. Soc. Am. **110**, 1141–1149.

Kiefte, M., Kluender, K. R., and Rhode, W. S. (**2002**). ''Synthetic speech stimuli spectrally normalized for nonhuman cochlear dimensions,'' ARLO **3**, 41–46.

Klatt, D. H. (**1980**). ''Software for a cascade/parallel formant synthesizer,'' J. Acoust. Soc. Am. **67**, 971–995.

Klatt, D. H. (**1982**). ''Speech processing strategies based on auditory models,'' in *The Representation of Speech in the Peripheral Auditory System*, edited by R. Carlson and B. Granström (Elsevier Biomedical, Amsterdam), pp. 181–196.

Klatt, D. H., and Klatt, L. C. (**1990**). ''Analysis, synthesis, and perception of voice quality variations among female and male talkers,'' J. Acoust. Soc. Am. **87**, 820–857.

Kluender, K. R., Coady, J. A., and Kiefte, M. (**2003**). ''Sensitivity to change in perception of speech,'' Speech Commun. **41**, 59–69.

Lindholm, J. M., Dorman, M., Taylor, B. E., and Hannley, M. T. (**1988**). ''Stimulus factors influencing the identification of voiced stop consonants by normal-hearing and hearing-impaired adults,'' J. Acoust. Soc. Am. **83**, 1608–1614.

Lotto, A. J., and Kluender, K. R. (**1998**). ''General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification,'' Percept. Psychophys. **60**, 602–619.

Lotto, A. J., Kluender, K. R., and Holt, L. L. (**1997**). ''Perceptual compensation for coarticulation by Japanese quail (*Coturnix coturnix japonica*),'' J. Acoust. Soc. Am. **102**, 1134–1140.

McCullagh, P., and Nelder, J. A. (**1989**). *Generalized Linear Models*, 2nd ed. (Chapman & Hall, London).

Moore, B. C. J., and Glasberg, B. R. (**1983**). ''Suggested formulas for calculating auditory-filter bandwidths and excitation patterns,'' J. Acoust. Soc. Am. **74**, 750–753.

Nearey, T. M., and Assmann, P. F. (**1986**). ''Modeling the role of inherent spectral change in vowel identification,'' J. Acoust. Soc. Am. **80**, 1297–1308.

Rosner, B. S., and Pickering, J. B. (**1994**). *Vowel Perception and Production* (Oxford University Press, New York).

Stevens, J. (**1992**). *Applied Multivariate Statistics for the Social Sciences* (Erlbaum, Hilsdale, NJ), 2nd ed.

Summerfield, Q., Haggard, M. P., Foster, J., and Gray, S. (**1984**). ''Perceiving vowels from uniform spectra: Phonetic exploration of an auditory aftereffect,'' Percept. Psychophys. **35**, 203–213.

Summerfield, Q., Sidwell, A., and Nelson, T. (**1987**). ''Auditory enhancement of changes in spectral amplitude,'' J. Acoust. Soc. Am. **81**, 700–707.
Viemeister, N. F., and Bacon, S. P. (**1982**). ''Forward masking by enhanced components in harmonic complexes,'' J. Acoust. Soc. Am. **71**, 1502–1507.
Wightman, F., McKee, T., and Kramer, M. (**1977**). ''Factors influencing frequency selectivity in normal and hearing-impaired listeners,'' in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London), pp. 295–310.

Interactions between cochlear implant electrode insertion depth and frequency-place mapping

Deniz Başkent[a)]
Department of Biomedical Engineering, University of Southern California, Los Angeles, California 90089 and House Ear Institute, 2100 West Third Street, Los Angeles, California 90057

Robert V. Shannon
Department of Biomedical Engineering, University of Southern California, Los Angeles, California 90089 and House Ear Institute, 2100 West Third Street, Los Angeles, California 90057

(Received 31 December 2003; revised 13 December 2004; accepted 15 December 2004)

While new electrode designs allow deeper insertion and wider coverage in the cochlea, there is still considerable variation in the insertion depth of the electrode array among cochlear implant users. The present study measures speech recognition as a function of insertion depth, varying from a deep insertion of 10 electrodes at 28.8 mm to a shallow insertion of a single electrode at 7.2 mm, in four Med-El Combi 40+ users. Short insertion depths were simulated by inactivating apical electrodes. Speech recognition increased with deeper insertion, reaching an asymptotic level at 21.6 or 26.4 mm depending on the frequency-place map used. Başkent and Shannon [J. Acoust. Soc. Am. **116**, 3130–3140 (2004)] showed that speech recognition by implant users was best when the acoustic input frequency was matched onto the cochlear location that normally processes that frequency range, minimizing the spectral distortions in the map. However, if an electrode array is not fully inserted into the cochlea, a matched map will result in the loss of considerable low-frequency information. The results show a strong interaction between the optimal frequency-place mapping and electrode insertion depth. Consistent with previous studies, frequency-place matching produced better speech recognition than compressing the full speech range onto the electrode array for full insertion ranges (20 to 25 mm from the round window). For shallower insertions (16.8 and 19.2 mm) a mild amount of frequency-place compression was better than truncating the frequency range to match the basal cochlear location. These results show that patients with shallow electrode insertions might benefit from a map that assigns a narrower frequency range than patients with full insertions. © *2005 American Institute of Physics.* [DOI: 10.1121/1.1856273]

PACS numbers: 43.71.Es, 43.71.Pc, 43.66.Ts [KWG] Pages: 1405–1416

I. INTRODUCTION

Recent studies on cochlear implants and simulations of cochlear implants have suggested that speech recognition is optimized when the frequency information is presented to the normal acoustic tonotopic cochlear location (Dorman *et al.*, 1997; Fu and Shannon, 1999a, b; Başkent and Shannon, 2003, 2004). Frequency-place maps that are shifted or distorted relative to the normal tonotopic map reduce speech recognition. However, if an electrode array is not fully inserted, the lowest frequency represented by the most apical electrode of that array may be 2000 Hz or higher. Matching frequency information to the acoustic tonotopic place in this case will result in the loss of a range of frequencies that is critical for speech recognition. The present study investigates the trade-off between such loss of low-frequency information when the acoustic input range is matched to the stimulation range, and distortion in the frequency-place mapping when a wider range of input acoustic range is assigned compressively.

More speech information could theoretically be delivered with longer and more deeply inserted electrode arrays. Blamey *et al.* (1992) and Skinner *et al.* (2002) found a significant correlation between open-set speech recognition scores of Nucleus 22 users and the insertion depths of the electrode arrays. Yet, Hodges *et al.* (1999) found no correlation between speech recognition and electrode insertion depth for insertion depths ranging from 17 to 25 mm. One complicating factor in these studies was that the speech performance was compared across different subjects with different array insertions rather than within subjects. If different pathologies contributed to the differences in electrode insertion depth, those pathologies might have also affected the residual nerve survival, and so speech recognition. In such a case poorer speech recognition with short electrode insertion may not be due to the short insertion *per se*, but rather due to the covarying pathology. Studies that simulated different insertion depths within subjects by selective activation of more apical electrodes found that deeper insertion generally resulted in improved speech recognition (Kileny *et al.*, 1998; Hochmair *et al.*, 2003).

Even though current electrode designs are intended to achieve array insertions as deep as 30 mm (Gstoettner *et al.*, 1999; Hochmair *et al.*, 2003), obstructions in the cochlea as a result of new bone formation, cochlear otosclerosis, or anatomical abnormalities might still prevent the full insertion of the array (Cohen and Waltzman, 1993). For example, three studies that used imaging methods to assess actual electrode

a)Electronic mail: dbaskent@hei.org

insertion depth (Cohen *et al.*, 1996; Ketten *et al.*, 1998; Skinner *et al.*, 2002) observed a range of insertions of Nucleus patients from 2 to 26 mm from the round window. There are also cases where a short array is used on purpose to preserve the residual acoustic hearing for a combined electric-acoustic stimulation (Kiefer *et al.*, 2004).

With such large variation in electrode insertion depths it becomes crucial to customize the frequency-place map for the individual implant user to achieve the best speech perception. Başkent and Shannon (2003) measured speech recognition by normal-hearing (NH) listeners in conditions that simulated cochlear implant processing. Their results suggested that implant users with full electrode array insertions (from 20 to 25 mm) would have the best speech recognition when the acoustic information was mapped onto the matching tonotopic range in the cochlea, even if that mapping resulted in the elimination of a considerable amount of acoustic information. A second study confirmed this finding with Med-El Combi 40+ users (Başkent and Shannon, 2004). In an earlier study, Whitford *et al.* (1993) modified the frequency-place map to match the characteristic frequencies of the electrode locations in the cochlea. They assigned the acoustic input at 3 kHz to the electrode at the cochlear location with the characteristic frequency of 3 kHz. The remaining acoustic input range was redistributed over the remaining electrodes around the matched electrode. Nucleus-22 users with array insertions from 20.5 to 23.5 mm had improved open-set sentence recognition scores in low levels of noise with the modified map. However, Eyles *et al.* (1995) did not observe significant improvement with Nucleus-22 patients with shallower insertion depths (from 14 to 21.5 mm) when tested with similar mapping strategy.

These studies imply that there is a strong interaction between the optimum frequency-place map and the electrode array insertion depth. It is important to quantify this interaction, considering that the electrode array insertion depth varies widely across individual patients. Speech recognition performance is generally very low for patients with shallow insertions. Many implant speech processors map the full range of frequencies onto the electrodes compressively even in cases of short electrode insertions, which might result in substantial frequency-place compression and shifting. One reason for a compressed map is to provide the patient with the full range of acoustic information with the assumption that she would eventually learn to make use of the information even if it is distorted in terms of cochlear tonotopic place. Several studies have shown evidence of partial adaptation (Rosen *et al.*, 1999; Fu *et al.*, 2002), but it is not clear if patients can fully adapt to a large distortion in the mapping and, if so, how much time such an adaptation would require. Studies of frequency-place distortions (Başkent, 2003; Başkent and Shannon, 2003, 2004) or frequency place shifts (Fu and Shannon, 1999b) showed a tolerance of only a few mm for such tonotopic distortions. Trinh *et al.* (2000) observed no significant improvement in speech recognition with five of eight Nucleus 22 users with partial insertions, over periods of time ranging from 12 to 60 months. This finding suggests that patients with short electrode insertions may only have a limited ability to adapt to the frequency-place compression normally assigned in such cases.

On the other hand, a strict matching of frequency information to cochlear place with short insertions would result in the loss of substantial low-frequency information that is important for speech. The Speech Intelligibility Index (SII), which predicts the speech recognition performance from the amount of audible acoustic information available (ANSI, 1997), weights frequency information between 1 and 3 kHz most heavily. When the speech frequency range is converted to cochlear location (Greenwood, 1990), the cochlear region around 14 to 28 mm from the round window receives speech information, with the most critical information concentrated between 18 and 25 mm. The 900-Hz region (21.5 mm) is important for the distinction of the first two formants of vowels, while the 1.5-kHz region (18.5 mm) is important for the distinction of high and low second formants. If the insertion depth is shallower than 19 mm, and frequency information is matched to the normal acoustic tonotopic place, then all frequencies lower than 1.3 kHz will be lost. Such a truncated map might be harmful for speech recognition, even though the acoustic information is delivered to the correct tonotopic location (Faulkner *et al.*, 2003).

The present study measures phoneme and sentence recognition in Med-El Combi 40+ implant users as a function of electrode array insertion depth and stimulation range. Experiment 1 extends the results from Başkent and Shannon (2003, 2004) to a wider range of insertion depths (varying from a shallow insertion depth of 7.2 mm with a single electrode to a deep insertion of 28.8 mm with an array of 10 electrodes), where the performances with compressed and matched maps are compared. Experiment 2 explores an optimum map for the shallow insertions of 19.2 and 16.8 mm by systematically changing the map from matched to fully compressed.

II. EXPERIMENTAL METHOD

A. Subjects

Four Med-El Combi 40+ users, aged 25–62, participated in the experiments. All were reported to have full electrode insertions at surgery. An insertion depth of 31 mm, which complies with company specifications, was assumed for all subjects. Information about subjects, such as duration and type of deafness, duration of implant use, baseline speech recognition scores, and the frequency ranges used in the clinical processor, is summarized in Table I. Table II additionally shows the center frequencies of the clinical maps the subjects use most.

M1, M3, and M4 were postlingually, and M2 was prelingually deafened. All subjects were born into hearing families, used oral communication as their main communication mode, and had been provided with speech correction therapies for long periods of time. Only M2 has used sign language frequently as an additional communication mode. All patients could converse over the telephone with their implants.

TABLE I. Information about Med-El Combi 40+ users.

Subject	Age	Duration of profound deafness (years)—reason of deafness	Experience with CI (years)	Baseline vowel score (corrected for chance)	Baseline consonant score (corrected for chance)	Baseline IEEE sentence score	Overall acoustic input frequency range of the original map and number of the electrodes activated
M1	39	30—High fever	2.5	60.0	55.3	38.2	300–5500 Hz, 6 or 12 electrodes later: 200–8500 Hz, 10, 11, or 12 electrodes
M2	25	From birth—Unknown	5	70.0	85.9	84.5	300–7000 Hz, 9 or 12 electrodes
M3	62	12—Noise exposure	1	68.2	70.2	92.8	300–5500 Hz, all 12 electrodes
M4	46	26—Unknown	2	82.5	86.7	93.9	Map 1 and 2: 300–5500 Hz Map 3: 300–7000 Hz 9 electrodes

B. Speech stimuli

The speech recognition tasks consisted of medial vowel and consonant identification, and sentence recognition.

Vowel stimuli (Hillenbrand *et al.*, 1995) consisted of 12 medial vowels, including 10 monophthongs and 2 diphthongs, presented in /h/-vowel–/d/ context (heed, hid, head, had, hod, hawed, hood, who'd, hud, heard, hayed, hoed) and spoken by five female and five male talkers. Chance level on this test was 8.33% correct, and the single-tailed 95% confidence level was 12.48% correct based on a binomial distribution.

Consonant stimuli (Shannon *et al.*, 1999) consisted of 20 medial consonants (b tʃ d ð f g ʒ k l m n p r s ʃ t v w j z) presented in /a/-consonant–/a/ context and spoken by three male and three female talkers. Chance performance level for this test was 5% correct, and the single-tailed 95% confidence level was 8.27% correct based on a binomial distribution.

IEEE sentences (IEEE, 1969), spoken by a single male speaker, are phonetically balanced across lists and the predictability of the words is relatively low. For each condition, the percent-correct score was acquired from 20 sentences of varying length from each listener. The length of each sentence varied from 5 words to 12 words. Lists of 20 sentences for each condition were prepared such that the average word length per sentence was 7 to 9 words. Sentences were presented without any context information, and no sentences were repeated to an individual listener.

C. Med-El Combi 40+ implant system

Combi 40+ electrode array consists of 12 electrode pairs equally distributed over 26.4-mm total length. The electrodes are numbered 1 to 12 from apex to base. The array is inserted through a cochleostomy around 4 mm from round window. In the study we used a TEMPO+ processor worn behind the ear, which can process frequencies from 200 Hz to 8.5 kHz.

D. Procedures

The study explores the acute effects of frequency-place maps on speech recognition. In both experiments the subjects were tested right after receiving the experimental processor.

The experimental processor was programed with a different frequency-place map before each test condition. All stimuli were presented via a loudspeaker in a sound field at

TABLE II. Center frequencies of clinical maps the subjects used most.

	Bandpass filter center frequencies for 12 electrodes (Hz)											
Subject	1	2	3	4	5	6	7	8	9	10	11	12
M1,M3	338	430	549	701	894	1137	1444	1845	2349	2987	3889	4918
M2	358	507	Off	722	1017	Off	1445	2057	Off	2890	4225	6013
M4	352	487	Off	672	930	Off	1273	1771	Off	2420	3456	4544

70 dB on an A-weighted scale. Phonemes were presented to the listeners in random order via custom software (Robert, 1998). Subjects were asked to select the phoneme they heard from the set of all possible phonemes displayed on the screen. Recognition of words in sentences was evaluated using custom software (TIGER Speech Recognition System developed by Qian-Jie Fu) which presented the sentences in random order. Subjects were asked to type the sentence they heard, and the number of words heard correctly was counted.

The implant processor, fitting software STUDIO+, and the fitting box were provided by Med-El for use in the experiments. The threshold and maximum comfort levels were measured with the research processor for each patient before testing and these levels were used throughout the experiments. Maximum stimulation rate was limited to 2 kpps per electrode. The electrodes were stimulated in monopolar mode and the processing strategy was continuous interleaved sampling (CIS).

The STUDIO+ software allows the programmer considerable flexibility in manually entering the bandpass filter cutoff frequencies that are assigned to each electrode. In every condition the acoustic input range was converted to cochlear distance in mm by the Greenwood mapping function. The range was divided into bandpass filters with equal bandwidths in mm, which were then converted to frequency in Hz with the Greenwood mapping function.

Greenwood's function describes the characteristic frequency along the organ of Corti as a function of cochlear place. In healthy cochlea, the maximum displacement on the basilar membrane shifts as a function of sound level. However, in implants, the place information does not come from the basilar-membrane motion. Instead, it is conveyed by selective stimulation of the spiral ganglia electrically. In our calculations, we assumed that Greenwood's function similarly holds at the level of spiral ganglia, and ignored effects of stimulation level.

We also assumed an average length of 35 mm for the cochlea in the calculations. Ulehlova *et al.* (1987) measured a range of 28 to 40 mm, with an average length of 34.2 mm, for human cochleae. However, Başkent and Shannon (2004) showed that a change of a few mm in the assumed length for the cochlea did not have a significant effect on speech recognition.

A third assumption in the study is uniform distribution of the nerve survival pattern. Any possible dead regions in the auditory nerve would be difficult to locate with implant users. Since we have not observed any abnormal threshold patterns during the fitting, we assumed the simplest case of uniform distribution for nerve survival pattern.

III. EXPERIMENT 1: FREQUENCY-PLACE MAPPING WITH VARYING INSERTION DEPTH

A. Experimental setup

In experiment 1, we changed the effective insertion depth by turning off the most apical electrode for each successive condition.

In the baseline condition, the 10 middle electrodes (numbered 2 to 11) were activated. We assumed an insertion depth of 31 mm for full insertion of the entire array. With this assumption, the middle 10 electrodes cover a cochlear range between 7.2 and 28.8 mm from the round window, which, according to the Greenwood frequency-place function, corresponds to a range of acoustic frequencies between 244 Hz and 7.5 kHz. Thus, the baseline condition simulates a deeply inserted (at 28.8 mm) array of 10 electrodes with a stimulation range of 21.6 mm. From this deep insertion condition the most apical electrode was turned off for each successive experimental condition. Because the electrode separation is 2.4 mm, each condition created an insertion depth that was 2.4 mm shallower than the previous condition. Note the number of the electrodes decreases by one with each successive condition as well.

In the first part of the experiment the acoustic input range was kept the same for all insertion depths (244 Hz–7.5 kHz). Starting from the baseline condition of 10 electrodes, where the acoustic input range matched the stimulation range, the map becomes increasingly compressive with each successive condition, as shown for the partial insertion condition of 6 electrodes in the upper part of Fig. 1. The acoustic input range assigned onto the electrodes is shown by the frequency bands on top, while the stimulation region covered by the actual electrode array is shown in the bottom of the map. These conditions simulate the clinical approach where the patient gets the same standard input frequency range in the map regardless of the actual location of the electrode array or the number of electrodes activated in her cochlea.

In the second part of the experiment, the acoustic input frequency range was truncated to match the stimulation range for all insertion depth conditions. In such a matched map, as the stimulation range gets smaller with each condition, the patient receives less acoustic information, as shown in lower portion of Fig. 1.

All compressed and matched conditions for all number of electrodes are summarized in Table III. The frequencies in the table show the actual frequencies STUDIO+ used rather than the theoretical values calculated with Greenwood mapping function.

B. Results and discussion

The individual percent-correct scores are shown in Fig. 2 as a function of insertion depth, expressed both in cochlear distance from the round window and in number of electrodes activated. The top, middle, and bottom rows present vowel, consonant, and sentence recognition scores, respectively. Vowel and consonant scores were corrected for chance level. M1 was excluded from the sentence recognition test due to her inability to recognize IEEE sentences at a reasonable level with the experimental maps. The open symbols show the percent-correct scores with the compressed map (upper part of Fig. 1) and the filled symbols show scores with the matched map (bottom part of Fig. 1). The dashed lines show the performances of subjects with their own processor maps, for which the center frequencies are shown in Table II. The baseline condition of 10 electrodes is slightly different from this map. Some subjects performed as well with the baseline condition of 10 electrodes as they did with their own map,

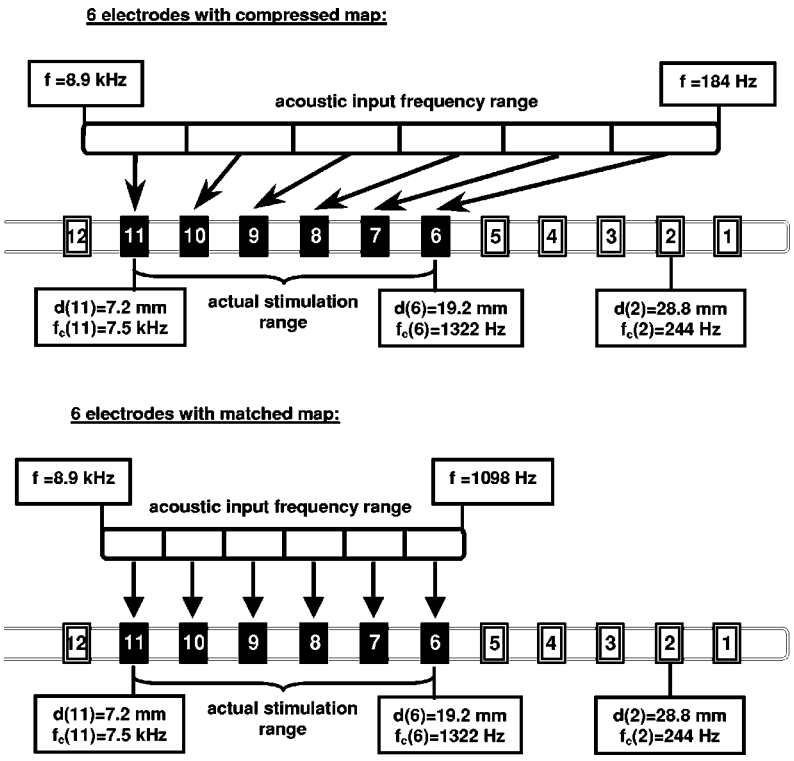


FIG. 1. Compressed and matched maps for the partial insertion condition of 6 electrodes, simulated by activating electrodes 6–11. In each map, the frequency bands on top show the acoustic input range. The lowest and highest frequencies are given at both ends of the input range. The bottom row of each map shows the electrode array with all 12 electrodes, with active electrodes shown in black. The stimulation range in the cochlea is determined by the position of the active electrodes. Assumed distances of the electrodes 2, 6, and 11 from the round window and the center frequencies calculated by Greenwood function based on these distances are shown under the electrodes. The upper part of the figure shows the compressed map, where the wide acoustic range is assigned onto the stimulation range compressively. The lower part of the figure shows the matched map, where the acoustic input range is truncated to match the stimulation range.

while some subjects performed worse. The subjects who performed worse might be more sensitive to spectral changes in the frequency-place mapping.

Figure 2 shows that performance dropped with both matched and compressed maps as the insertion depth became shallower. Findings from previous studies imply that there are several factors contributing to the decrease in performance: decrease in the number of electrodes (Friesen *et al.*, 2001), transmitting a smaller portion of the speech spectrum with the matched map (Faulkner *et al.*, 2003; Hochmair *et al.*, 2003), spectral distortions with the compressed map (Başkent and Shannon, 2003, 2004).

The matched and compressed maps produced similar scores for the deep insertion condition of 26.4 mm, probably due to the minimal difference in the acoustic input ranges of these two maps (Table III). At insertion ranges from 19.2 to

TABLE III. Frequency-place mapping conditions with compressed and matched maps for varying insertion depths. The conditions also used in experiment 2 are denoted by symbols Δ, *, +, and ◇.

Condition	Number of active electrodes, electrodes employed	Length of the active electrode array	Insertion depth of the active electrode array (mm)	Input frequency range: center frequency range (Hz), total analysis range (Hz)	
				Compressed map	Matched map
10	10 (2–11)	21.6 mm	28.8	273–7.3 k 214–8.4 k	273–7.3 k 214–8.4 k
9	9 (3–11)	19.2 mm	26.4	272–7.3 k 212–8.5 k	393–7.3 k 312–8.4 k
8	8 (4–11)	16.8 mm	24	269–7.3 k 206–8.5 k	609–7.3 k 496–8.4 k
7	7 (5–11)	14.4 mm	21.6	265–7.2 k 199–8.6 k	906–7.3 k 745–8.4 k
6	6 (6–11)	12 mm	19.2	275–7.1 k 201–8.8 k$^{\Delta}$	1332–7.3 k 1108–8.4 k*
5	5 (7–11)	9.6 mm	16.8	293–6.4 k 207–8.1 k^{+}	1896–7.3 k 1583–8.4 k$^{\diamond}$
4	4 (8–11)	7.2 mm	14.4	342–6.4 k 238–8.1 k	2676–7.3 k 2250–8.4 k
3	3 (9–11)	4.8 mm	12	461–5.8 k 315–7.5 k	3803–7.3 k 3199–8.4 k
2	2 (10–11)	2.4 mm	9.6	1.0 k–4.8 k 682–6.5 k	5557–7.3 k 4673–8.4 k
1	1 (11)	single electrode	7.2	4.4 k 3.1 k–6.1 k	—

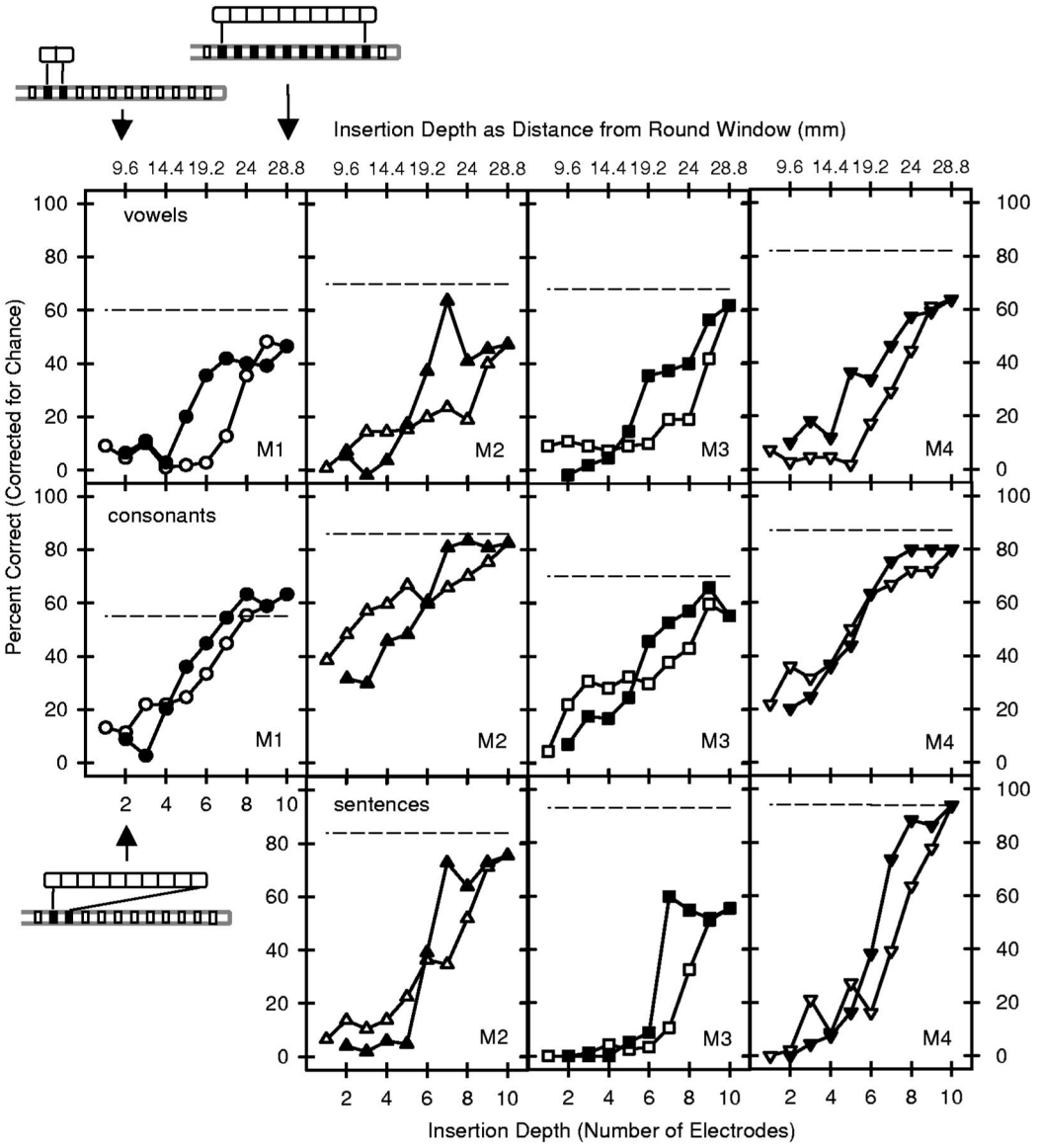


FIG. 2. Individual percent-correct scores from Med-El Combi 40+ users. Filled symbols show the scores when the acoustic frequency range was matched to the cochlear stimulation region (matched map). The figures on top of the first column show the matched maps with 2 and 10 electrodes. Open symbols show the scores when the entire acoustic bandwidth was compressed into the shorter stimulation region (compressed map). The compressed map with the full acoustic input frequency and two apical electrodes activated is shown in the figure under the first column. Dashed lines show the scores the subjects obtained with the map they used most with their own processors. Vowel and consonant recognition scores are corrected for chance level.

24 mm, all subjects performed better with the matched map, consistent with previous studies with fully inserted electrode arrays. Even though the compressed map assigned a larger frequency range, it produced poorer speech recognition due to the distortion in the frequency-place mapping. However, the advantage of the matched map vanished at shallow insertions and compression usually produced better performance. The crossover point was at the insertion depth of 14.4 mm for M1, in the range of 16.8–19.2 mm for M2, and 14.4–19.2 mm for M3 and M4. This transition demonstrates the trade-off between the frequency range delivered to the electrodes and the accuracy of the frequency-place mapping.

A minor but interesting point is that subjects M2 and M4 had higher consonant recognition scores than M1 and M3, which probably also contributed to their higher sentence recognition scores. Note that M2 still identified 40% and M4 identified 20% of consonants correctly even when they had only one electrode active (in Fig. 2). In this condition there is no spectral resolution; the entire spectral range is provided to a single electrode. The carrier pulse rate on the single electrode is simply modulated by the broadband envelope of the speech signal. This observation suggests that M2 and M4 make better use of temporal cues.

To explore a common pattern, the average scores of the subjects are plotted in Fig. 3. Similar to Fig. 2, filled circles show the average percent-correct scores of all patients for phonemes, and average scores of M2, M3, and M4 for sentences, with the matched map. Open circles show the average scores when the subjects were tested with the compressed map. Performance levels with both maps dropped significantly ($p<0.001$) with decreasing insertion depth, for vowels, consonants, and sentences, as shown by one-way repeated measures ANOVAs.

A paired t-test was applied to the scores to compare the

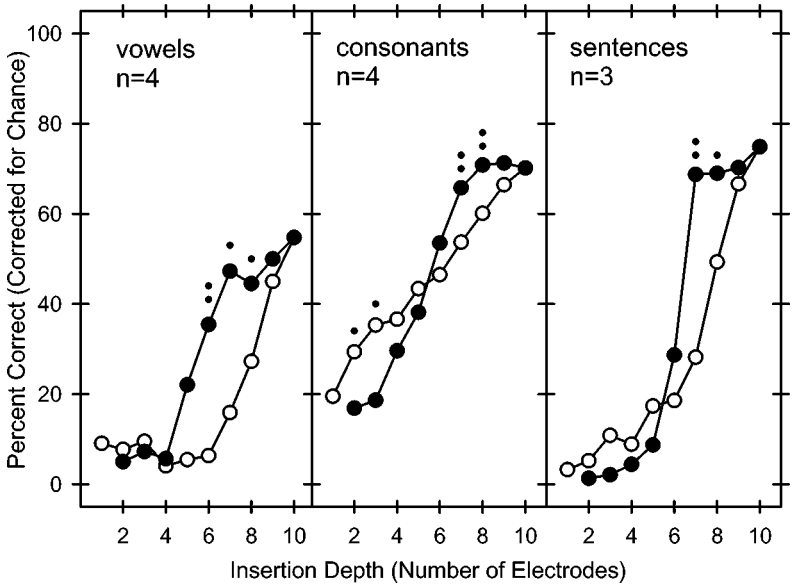


FIG. 3. Average percent-correct scores of M1, M2, M3, and M4 for phonemes, and M2, M3, and M4 for sentences. The open circles show the scores when a wide acoustic range was mapped to electrodes (compressed map), whereas the filled circles show the scores when the acoustic range was truncated to match the stimulation range (matched map). Small dots on top of the scores show the significance level of the difference between the scores from two maps with paired *t*-test: one dot for $p<0.05$, two dots for $p<0.01$.

two mapping conditions. The level of significance is shown by small dots in Fig. 3. The *t*-test shows a significant advantage of the matched map for moderate insertion depths (6–8 electrodes for vowels, and 7–8 electrodes with consonants and sentences) and a significant advantage in consonant recognition for the compressed map at very shallow insertions (2–3 electrodes). At insertion depths shallower than 4 electrodes vowel and sentence recognition performances are at floor level and there is no significant difference between compressed and matched maps.

An interesting observation is that scores for both maps reach an asymptotic level with increasing insertion depth (and hence increasing number of electrodes) before the baseline condition of 10 electrodes. The improvement in the performance stops at the insertion depth condition of 9 electrodes for the compressed map, and 7 electrodes for the matched map, as shown by a *posthoc* Tukey test. The frequencies corresponding to these ranges contain useful spectral information for vowel recognition. Therefore, their inclusion in the overall acoustic input range would be expected to increase the performance. On the other hand, this result is consistent with previous implant studies that showed little improvement in speech recognition as the number of electrodes was increased above 7 (Fishman *et al.*, 1997; Friesen *et al.*, 2001). The lack of a significant improvement with further insertion than 7 electrodes (21.6 mm in cochlear distance) might also reflect decreased frequency selectivity in the middle and apical turns of the cochlea when stimulated electrically. Spiral ganglia are located near the habenula perforata, near the medial wall of the scala tympani in the basal turn, but are located more centrally in the modiolus apically. The modiolus gets narrower at the apex so the spiral ganglia are packed in a tight bundle, making selective tonotopic activation difficult. As a result, the anatomy of the cochlea might also be limiting the potential improvement with deeper insertions (Cohen *et al.*, 1996).

Figures 2 and 3 show the trade-off between the amount of acoustic information available versus the accuracy of the location where the information is mapped. The scores in these figures show two extremes, where either the widest acoustic frequency range available was assigned, or it was limited to the matching stimulation range at the expense of losing important acoustic information. It is possible that there might be an optimum range in between where some information is included by truncating the acoustic range less severely and producing a milder distortion in the location where the information is mapped.

The following experiment is designed to explore the possibility of such an optimum map at two insertion depths simulating shallow insertions: 19.2 and 16.8 mm.

IV. EXPERIMENT 2: FREQUENCY-PLACE MAPPING AT 19.2- AND 16.8-MM INSERTION DEPTHS

A. Experimental setup

In the previous experiment, the matched map resulted in significantly better vowel recognition compared to the compressed map at 19.2-mm insertion [with 6 electrodes (6–11) active]. There was little difference between matched and compressed maps for consonant and sentence recognition. When an insertion depth of 16.8 mm was simulated [with 5 active electrodes (7–11)] the performance levels with the matched and compressed maps were not significantly different for all speech materials. Possibly the matched map leaves out too much low-frequency information at these relatively short insertions, while the compressed map introduces too much distortion in the speech patterns. There might be an optimum trade-off between these two extreme maps where a relatively wider acoustic range is assigned onto a relatively accurate cochlear location. To explore this possibility, the frequency-place map was changed from the matched map to the compressed map gradually, while the same set of electrodes was used for each condition.

An insertion depth of 19.2 mm was simulated by activating electrodes 6–11. These electrodes cover 12 mm, from 7.2 to 19.2 mm from the round window, with the assumption of 31 mm for full insertion. The frequency range assigned onto the electrodes was first matched to this stimulation

TABLE IV. Compression conditions for the array of 6 electrodes (electrodes 6–11) inserted 19.2-mm deep and covering 12 mm in the cochlea. The acoustic information assigned onto the array increases as the map changes from 0-mm matching condition to +8-mm compression while the stimulation region remains the same. The maps with * and Δ are same conditions from experiment 1, as shown in Table III.

Frequency-place mismatch condition	Range of acoustic input (mm)	Bandpass filter center frequencies (Hz)						Frequency range of analysis bands (Hz)
		6	7	8	9	10	11	
0 mm (matching)	7.2–19.2	1332	1896	2676	3803	5558	7310	1108–8.4 k*
+1 mm (compression)	7.2–20.2	1132	1664	2472	3597	5146	7290	924–8.5 k
+2 mm (compression)	7.2–21.2	967	1479	2294	3397	5102	7269	773–8.5 k
+3 mm (compression)	7.2–22.2	821	1327	2075	3221	5059	7248	643–8.6 k
+4 mm (compression)	7.2–23.2	695	1160	1894	3051	4695	7226	531–8.6 k
+5 mm (compression)	7.2–24.2	584	1027	1733	2902	4650	7203	438–8.6 k
+6 mm (compression) (clinical setting)	7.2–25.2	489	904	1595	2661	4609	7181	354–8.6 k
+7 mm (compression) (clinical setting)	7.2–26.2	408	787	1446	2535	4300	7158	284–8.7 k
+8 mm (compression) (clinical setting)	7.2–27.2	275	630	1225	2285	4214	7109	201–8.8 k$^{\Delta}$

range, then made wider by adding lower frequencies in steps of 1 mm in cochlear distance. In the second part of the experiment, electrodes 7–11 covering 9.6 mm, from 7.2 to 16.8 mm, were activated to simulate a 16.8-mm insertion depth. The frequency-place map was changed from the matched to the compressed map in 1.5-mm steps. The conditions are summarized in Tables IV and V for insertion depths of 19.2 and 16.8 mm, respectively. The compression conditions from +6 to +8 mm for 19.2-mm insertion depth, and from +7.5 to +10.5 mm for 16.8-mm insertion depth are similar to the clinically available maps.

B. Results and discussion

Figure 4 shows the vowel and consonant recognition percent scores (corrected for chance level) as a function of increasing frequency-place compression. The top panels show the scores with 6 electrodes inserted to 19.2 mm. The bottom panels show the scores with 5 electrodes inserted to 16.8 mm. The thin lines with open symbols show the individual scores, while the thick lines show the average performance of all subjects. The same symbols from Fig. 2 were used to represent scores from individual subjects. In each panel, the area between the vertical dashed lines shows the maps that can be programed in the clinic with the standard fitting procedure. The 0-mm matching condition in Fig. 4 is the same as the matched map shown with filled symbols in Fig. 2 (shown by the maps with * and ◇ in Tables III–V). The +8-mm compression in the upper panels and +12-mm compression in the lower panels are the compressed maps with widest frequency range of acoustic input shown with open symbols in Fig. 2 for 19.2- and 16.8-mm insertion depth conditions (shown by the maps with Δ and + in Tables III–V), respectively. Because the electrode array at 16.8 mm is shorter, the assignment of the full acoustic frequency range effectively results in more compression than the 19.2-mm insertion depth.

A repeated-measures one-way ANOVA shows that compression had a significant effect on the vowel recognition performance at 19.2 mm [$F(8,24)=14.06$, $p<0.001$] and 16.8-mm insertions [$F(8,24)=11.27$, $p<0.001$]. At both insertion depths a clear peak was observed in the vowel recognition performance with an optimal map of a few mm compression.

At 19.2-mm insertion, the peak performance was obtained with a compression of +2 to +3 mm. These optimal maps resulted in a performance level 10% higher than the 0-mm matched map (not significant by paired t-test), 20% higher than the compression map most similar to that offered by the clinical fitting program (+6-mm compression, $p<0.01$), and 35% higher than the compression condition where the full acoustic frequency range was assigned to the electrodes (+8-mm compression, $p<0.01$).

This finding implies that including acoustic information as low as 700 Hz (from Table IV) in the acoustic input range improved the performance, but adding lower frequencies started decreasing the performance. If a patient had a shallow insertion of 19.2 mm, the closest value offered by the clinical program for the low end of the frequency range would be 350 Hz. This mapping is shown by the +6-mm compression

TABLE V. Compression conditions for the array of 5 electrodes (electrodes 7–11) inserted 16.8-mm deep and covering 9.6 mm in the cochlea. The acoustic information assigned onto the array increases as the map changes from 0-mm matching condition to +12-mm compression while the stimulation region remains the same. The maps with ◇ and + are same conditions from experiment 1, as shown in Table III.

Frequency-place mismatch condition	Range of acoustic input (mm)	Bandpass filter center frequencies (Hz)					Frequency range of analysis bands (Hz)
		7	8	9	10	11	
0 mm (matching)	7.2–16.8	1899	2676	3802	5558	7310	1583–8.4 k◇
+1.5 mm (compression)	7.2–18.3	1510	2297	3403	5102	7279	1221–8.5 k
+3 mm (compression)	7.2–19.8	1203	1949	3062	4711	7238	941–8.6 k
+4.5 mm (compression)	7.2–21.3	948	1641	2772	4632	7191	715–8.7 k
+6 mm (compression)	7.2–22.8	737	1378	2441	4278	7145	533–8.7 k
+7.5 mm (compression) (clinical setting)	7.2–24.3	582	1155	2215	4191	7096	412–8.8 k
+9 mm (compression) (clinical setting)	7.2–25.8	460	975	1981	3877	7084	326–8.8 k
+10.5 mm (compression) (clinical setting)	7.2–27.3	361	838	1816	3631	7084	256–8.8 k
+12 mm (compression)	7.2–28.8	293	718	1635	3631	6354	207–8.1 k+

(from Table IV). At this condition two subjects performed worse than the matched map even though the matched map discards all information below 1 kHz. The +6-, +7-, and +8-mm compressed maps are the only choices offered by the standard clinical fitting program, and they are clearly not the optimal maps for such shallow insertion.

With a 16.8-mm insertion, a peak performance was observed with +3- and +4.5-mm compression conditions. The performance was 20% higher compared to the 0-mm matched condition ($p<0.05$, by paired t-test), 20% higher than the closest compression map offered by the clinical fitting program (+7.5-mm compression, $p<0.05$), and 30% higher than the compression condition where the full acoustic frequency range was assigned to the electrodes (+12 compression, $p<0.05$). A *posthoc* Tukey test showed that there was no significant difference for conditions between +3- and +6-mm compression on vowel recognition, which corresponds to a range of low-end frequencies for stimulation range from 536 to 941 Hz (Table III). In the +4.5-mm compression condition all frequencies higher than 715 Hz were assigned onto the electrode array, but adding further lower frequencies increased the amount of frequency-place compression and reduced performance.

Consonant scores did not change significantly with compression, but a small peak around +4- and +5-mm compression and a slight drop of 10% for extreme compression of +8 mm were observed at 19.2-mm insertion. At 16.8-mm insertion the effect of compression on consonant recognition was significant [$F(8,24)=7.54$, $p<0.001$, by repeated-measures one-way ANOVA]. At this depth, the optimal range was much wider compared to vowels. Maps from +1.5- to +7.5-mm compression resulted in better consonant recognition compared to 0-mm matched or highly compressed maps ($p<0.05$).

The results support the hypothesis that, for shallow electrode insertions, a compromise between the amount of low-frequency information provided and the accuracy of the mapping of that information to cochlear place might be beneficial. For example, Fig. 4 shows that by choosing a mild frequency-place compression, vowel recognition performance can instantly be increased by 20%–30% compared to the clinical maps, which produce more severe compression. At a low performance level (due to the shallow insertion) such an increase (from 20% to 45% for vowels, and from 40% to 55% for consonants, for example) will have a significant effect on the patient's speech understanding.

V. GENERAL DISCUSSION

The results of the present study mainly show that there is a strong interaction between the frequency range assigned to the electrodes and the distortion in the frequency-electrode mapping for a wide range of electrode insertions.

Speech recognition generally increases as the insertion depth and number of electrodes activated increase, reaching an asymptotic level at an insertion smaller than the baseline condition (10 electrodes, at 28.8-mm insertion). For full insertion ranges of 20 to 25 mm a matched map with less spectral distortions results in better speech recognition. For shorter insertions, however, a map that reduces the input frequency range to preserve the normal acoustic mapping eliminates too much low-frequency information that is important

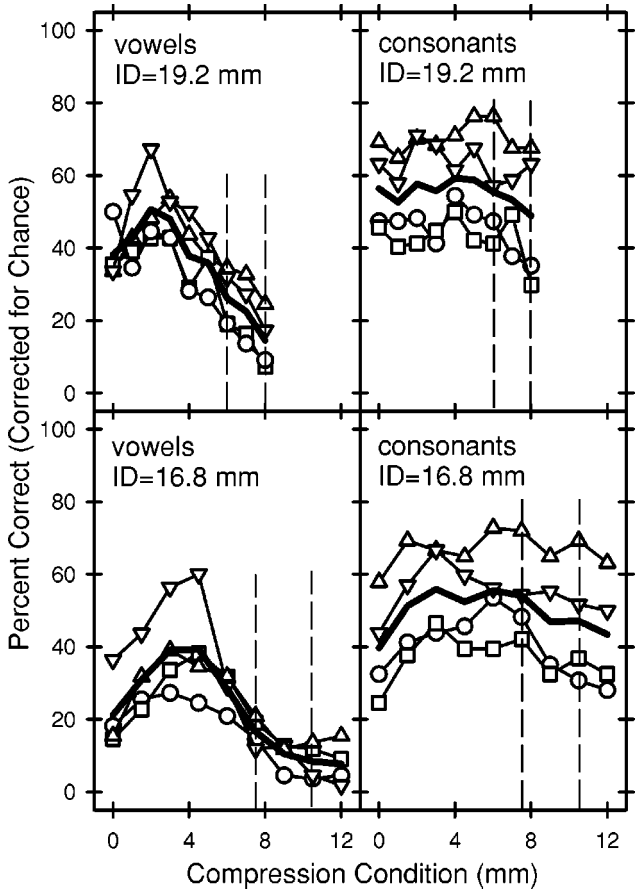


FIG. 4. Individual percent-correct scores (shown by open symbols and thin lines) superimposed with average performances of all subjects (shown by thick lines). In the top row, the frequency-place map is changed from perfect match (0-mm condition) to the compressed map (+8-mm compression) in steps of +1-mm cochlear distance, when 6 electrodes at 19.2-mm insertion depth were activated. In the bottom row the map is changed from perfect match to the compressed map (+12-mm compression) in steps of +1.5 mm when 5 electrodes at 16.8-mm insertion depth were activated. The same symbols from Fig. 2 were used to represent scores of individual subjects. The maps between the vertical dashed lines show the clinically available maps.

for speech. Compressing the full acoustic range onto a short electrode insertion also results in poor speech recognition because of the distortion (compression) in the frequency-place mapping (experiment 1). Optimal recognition of spectrally sensitive stimuli like vowels occurs with a compromise between these two extreme maps (experiment 2).

Several studies have previously shown that speech recognition increases with deeper electrode insertion. For example, Hochmair *et al.* (2003) observed an improvement in speech perception from a shallow insertion of 20 mm to a deep insertion of 30 mm. In the present study we observed an increase in scores up to 26-mm insertion with the compressed map, and 22 mm with the matched map, but no further improvement for deeper insertions. Ideally, inclusion of the lower frequencies in the input acoustic range at deeper insertions would be expected to increase speech recognition. However, the difficulty of selective stimulation of the auditory nerves at such deep apical regions of the cochlea might have a limiting effect. The cochlea is coiled more tightly towards the apical end, physically restricting a deep insertion. Even if the deep insertion is achieved, it is difficult to stimulate spiral ganglia of different characteristic frequencies selectively because they are more densely clustered in the apical turn.

A second factor contributing to the asymptotic performance might be the reduced spectral channels of implant users. It has been shown that implant patients only utilize a limited number of stimulation channels, regardless of the number of stimulating electrodes (Fishman *et al.*, 1997; Friesen *et al.*, 2001). Adding more electrodes (to the 9 electrodes with the compressed map and 7 electrodes with the matched map) similarly did not result in a significant improvement in performance in the present study.

Many studies have shown that speech recognition is adversely affected by spectral distortions such as a spectral shift between the acoustic input range and the stimulation range (Dorman *et al.*, 1997; Fu and Shannon, 1999b), or nonlinear distortions (Shannon *et al.*, 1998), even when the same speech information was used. In more recent studies by Başkent and Shannon (2003, 2004) the acoustic frequency range was systematically made wider or narrower than the tonotopic stimulation range. The results showed that both normal hearing and implant subjects were sensitive to abrupt frequency-place distortions. The subjects had only a limited tolerance of a few mm and performance dropped significantly with further distortion, especially in vowel and sentence recognition tests. These studies generally used an insertion depth ranging from 20 to 25 mm, which represent the ''full-insertion'' range for cochlear implants. Consistent with the previous studies, the present study showed a matched map is advantageous over a compressed map with spectral distortions for full insertion ranges.

Note that there are several assumptions used in the present study. There are many unknown factors with implant users such as individual cochlear length, electrode array insertion depth and its lateral distance from modiolus, nerve survival patterns, and the best frequencies of the nerves actually stimulated by each electrode. We did not have radiographic images of the implants for precise calculations of cochlear dimensions or electrode array positions in the scala tympani. We simply assumed a typical value of 35 mm for the cochlear length, an insertion depth of 31 mm for the full array, a medial location from the modiolus, and functioning nerves uniformly distributed along the organ of Corti. We also assumed that the Greenwood mapping function holds at the spiral ganglia level. Similar assumptions were made by Başkent and Shannon (2004) to match the acoustic input range to the stimulation range in Med-El users. Such studies show that an initial map can be estimated with similar assumptions to the present study, and it can further be fit for the individual patient functionally by using tests with a small set of spectrally sensitive stimuli such as vowels.

The latest generation implants are designed to be inserted much deeper (up to 31 mm) than the conventional full insertion of 25 mm. Despite the improvements in implant designs and surgical techniques, there are still implant patients who receive partially inserted electrode arrays, mostly due to bone and fiber occlusions in the cochlea. In such short insertion cases it is not clear whether it is better to match the

acoustic frequencies to the actual electrode tonotopic range, thus losing low-frequency information, or to present a wider acoustic range to the electrodes, resulting in a distortion in the frequency-place mapping. Faulkner *et al.* (2003) showed that matching the frequency to the tonotopic place for insertions shorter than 19 mm was detrimental to speech recognition. The results of the present study also showed that matching was detrimental to speech recognition for short insertions; the best speech recognition was achieved with a compromise between compressing the entire frequency range onto a short insertion electrode and truncating the frequency range to match the short electrode tonotopic place.

The results in the present study show the acute effects of frequency-place maps on speech recognition. An important consideration in actual implant users would be the role of long-term learning. If the frequency range is matched to the electrode tonotopic location the resulting mapping will eliminate low-frequency information. Adding low-frequency information will result in frequency-place distortion. The low-frequency information is presented to the listener in this case, but in a distorted form that might be learned over time. However, it is not clear how flexible the speech pattern recognition in the central nervous system is. Over a lifetime of a normal-hearing listener the brain learns to recognize speech patterns based on the normal tonotopic distribution of frequency information in the cochlea. When hearing is lost and later restored by a cochlear implant, the implant may not provide the brain with the same distribution of tonotopic information as a normal cochlea, depending on the electrode insertion depth and the frequency mapping. Speech pattern recognition in normal hearing is based on a physiological ''hard-wired'' tonotopic representation from the cochlea to the brain. In a cochlear implant any range of frequency information can theoretically be presented to any electrode, so the frequency-place mapping is a manipulable factor in implant fitting. How much distortion in the frequency-place mapping is learnable? What are the trade-offs between frequency range and distortion in frequency-place mapping? The present results suggest several factors that should be considered in selecting the frequency-electrode mapping.

In present clinical practice the general approach is to provide the patient with as much acoustic input as available, regardless of her specific implant configuration such as the insertion depth, in the belief that eventually the patient would learn to make use of this abundant information regardless of any frequency-place distortions. The results of Fig. 4 suggest that this approach may not be optimal. Vowel recognition with the widest frequency range was significantly poorer than the conditions that used a modest amount of compression. The widest frequency range did not add much information that is important for speech and may have degraded the recognition of vowels because it reduces the tonotopic resolution within the most important speech range. So, even if the compressed frequency-place mapping can be learned, the loss of resolution may limit performance with a wide frequency range.

From the present results we infer that there are three factors in frequency-place mapping that determine vowel recognition: the match between frequency and place (especially on the apical end; see Başkent, 2003), the amount of low-frequency information deleted, and the frequency resolution. While cochlear implant listeners might be able to learn a distorted pattern of tonotopic activity, they probably cannot overcome the loss of information caused by truncating the frequency range and the loss of frequency resolution within the speech range. Processor settings to optimize the transmission of spectral cues should include the most important frequency range for speech, should maximize the spectral resolution within that range, and minimize the distortion between the presented frequency place mapping and the original acoustic tonotopic map.

ACKNOWLEDGMENTS

The authors would like to thank Med-El Corp., especially Amy Barco and Peter Nopp, for providing the equipment and software, help recruiting subjects and with their travel expenses, Dawna Mills, Pam Fiebig, and Michelle Colburn, for help recruiting subjects, and the subjects for their valuable efforts. Also, the comments by Stefan Brill, David Nelson, and an anonymous reviewer are greatly appreciated. Funding for this research was provided in part by NIDCD Grant R01-DC-01526 and Contract N01-DC-92100.

ANSI (**1997**). ''S35-1997, Methods for calculation of the speech intelligibility index,'' (American National Standards Institute, New York).

Başkent, D. (**2003**). ''Speech recognition under conditions of frequency-place compression and expansion,'' Ph.D. thesis, University of Southern California, CA.

Başkent, D., and Shannon, R. V. (**2003**). ''Speech recognition under conditions of frequency-place compression and expansion,'' J. Acoust. Soc. Am. **113**, 2064–2076.

Başkent, D., and Shannon, R. V. (**2004**). ''Frequency-place compression and expansion in cochlear implant listeners,'' J. Acoust. Soc. Am. **116**, 3130–3140.

Blamey, P. J., Pyman, B. C., Clark, G. M., Dowell, R. C., Gordon, M., Brown, A. M., and Hollow, R. D. (**1992**). ''Factors predicting postoperative sentence scores in postlinguistically deaf adult cochlear implant patients,'' Ann. Otol. Rhinol. Laryngol. **101**, 342–348.

Cohen, L. T., Busby, P. A., Whitford, L. A., and Clark, G. M. (**1996**). ''Cochlear implant place psychophysics. I. Pitch estimation with deeply inserted electrodes,'' Audiol. Neuro-Otol. **1**, 265–277.

Cohen, N. L., and Waltzman, S. B. (**1993**). ''Partial insertion of the Nucleus multichannel cochlear implant: Technique and results,'' Am. J. Otol. **14**, 357–361.

Dorman, M. F., Loizou, P. C., and Rainey, D. (**1997**). ''Simulating the effect of cochlear-implant electrode insertion depth on speech understanding,'' J. Acoust. Soc. Am. **102**, 2993–2996.

Eyles, J. A., Boyle, P. J., and Burton, M. J. (**1995**). ''Characteristic frequency mapping in subjects using the Nucleus 22-channel cochlear implant system with partial and full insertions,'' Ann. Otol. Rhinol. Laryngol. Suppl. **116**, 356–358.

Faulkner, A., Rosen, S., and Stanton, D. (**2003**). ''Simulations of tonotopically mapped speech processors for cochlear implant electrodes varying in insertion depth,'' J. Acoust. Soc. Am. **113**, 1073–1080.

Fishman, K. E., Shannon, R. V., and Slattery, W. H. (**1997**). ''Speech Lang. recognition as a function of the number of electrodes used in the SPEAK cochlear implant speech processor,'' J. Speech Lang. Hear. **40**, 1201–1215.

Friesen, L. M., Shannon, R. V., Başkent, D., and Wang X. (**2001**). ''Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants,'' J. Acoust. Soc. Am. **110**, 1150–1163.

Fu, Q.-J., and Shannon, R. V. (**1999a**). ''Effects of electrode location and spacing on phoneme recognition with the Nucleus-22 cochlear implant,'' Ear Hear. **20**, 321–331.

Fu, Q.-J., and Shannon, R. V. (**1999b**). ''Recognition of spectrally degraded and frequency-shifted vowels in acoustic and electric hearing,'' J. Acoust. Soc. Am. **105**, 1889–1900.

Fu, Q.-J., Shannon, R. V., and Galvin, J. J. (**2002**). ''Perceptual learning following changes in the frequency-to-electrode assignment with the Nucleus-22 cochlear implant,'' J. Acoust. Soc. Am. **112**, 1664–1674.

Greenwood, D. D. (**1990**). ''A cochlear frequency-position function for several species—29 years later,'' J. Acoust. Soc. Am. **87**, 2592–2605.

Gstoettner, W., Franz, P., Hamzavi, J., Plenk, H., Baumgartner, W., and Czerny, C. (**1999**). ''Intracochlear position of cochlear implant electrodes,'' Acta Oto-Laryngol. **119**, 229–233.

Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). ''Acoustic characteristics of American English vowels,'' J. Acoust. Soc. Am. **97**, 3099–3111.

Hochmair, I., Arnold, W., Nopp, P., Jolly, C., Müller, J., and Roland, P. (**2003**). ''Deep electrode insertion in cochlear implants: Apical morphology, electrodes and speech perception results,'' Acta Oto-Laryngol. **123**, 612–617.

Hodges, A. V., Villasuso, E., Balkany, T., Bird, P. A., Butts, S., Lee, D., and Gomez, O. (**1999**). ''Hearing results with deep insertion of cochlear implant electrodes,'' Am. J. Otol. **20**, 53–55.

Institute of Electrical and Electronics Engineers (**1969**). ''IEEE recommended practice for speech quality measurements.''

Ketten, D. R., Skinner, M. W., Wang, G., Vannier, M. W., Gates, G. A., and Neely, J. G. (**1998**). ''*In vivo* measures of cochlear length and insertion depth of nucleus cochlear implant electrode arrays,'' Ann. Otol. Rhinol. Laryngol. Suppl. **175**, 1–16.

Kiefer, J., Gstoettner, W., Baumgartner, W., Pok, S. M., Tillein, J., Ye, Q., and Von Ilberg, C. (**2004**). ''Conservation of low-frequency hearing in cochlear implantation,'' Acta Oto-Laryngol. **124**, 272–280.

Kileny, P. R., Zwolan, T. A., Telian, S. A., and Boerst, A. (**1998**). ''Performance with the 20+2L lateral wall cochlear implant,'' Am. J. Otol. **19**, 313–319.

Robert, M. E. (**1998**). ''CONDOR: Documentation for Identification Test Program,'' Los Angeles; House Ear Institute.

Rosen, S., Faulkner, A., and Wilkinson, L. (**1999**). ''Adaptation by normal listeners to upward spectral shifts of speech: Implications for cochlear implants,'' J. Acoust. Soc. Am. **106**, 3629–3636.

Shannon, R. V., Zeng, F.-G., and Wygonski, J. (**1998**). ''Speech recognition with altered spectral distribution of envelope cues,'' J. Acoust. Soc. Am. **104**, 2467–2476.

Shannon, R. V., Jensvold, A., Padilla, M., Robert, M. E., and Wang, X. (**1999**). ''Consonant recordings for speech testing,'' J. Acoust. Soc. Am. **106**, L71–L74.

Skinner, M. W., Ketten, D. R., Holden, L. K., Harding, G. W., Smith, P. G., Gates, G. A., Neely, J. G., Kletzker, G. R., Brunsden, B., and Blocker, B. (**2002**). ''CT-Derived estimation of cochlear morphology and electrode array position in relation to word recognition in Nucleus 22 recipients,'' J. Assoc. Res. Otolaryngol. **3**, 332–350.

Trinh, B. T., Bergeron, F., and Ferron, P. (**2000**). ''Long-term follow-up of cochlear implant users with ossified cochlea,'' J. Otolaryngol. **29**, 279–284.

Ulehlova, L., Voldrich, L., and Janisch, R. (**1987**). ''Correlative study of sensory cell density and cochlear length in humans,'' Hear. Res. **28**, 149–151.

Whitford, L. A., Seligman, P. M., Blamey, P. J., McDermott, H. J., and Patrick, J. F. (**1993**). ''Comparison of current speech coding strategies,'' Adv. Oto-Rhino-Laryngol. **48**, 85–90.

Glottal open quotient in singing: Measurements and correlation with laryngeal mechanisms, vocal intensity, and fundamental frequency

Nathalie Henrich[a)]
LAM (UPMC, CNRS, Ministère de la culture), 11 rue de Lourmel, 75015 Paris, France and LIMSI-CNRS, BP 133, F91403 Orsay, France

Christophe d'Alessandro and Boris Doval
LIMSI-CNRS, BP 133, F91403 Orsay, France

Michèle Castellengo
LAM (UPMC, CNRS, Ministère de la culture), 11 rue de Lourmel, 75015 Paris, France

(Received 10 May 2004; revised 24 November 2004; accepted 29 November 2004)

This article presents the results of glottal open-quotient measurements in the case of singing voice production. It explores the relationship between open quotient and laryngeal mechanisms, vocal intensity, and fundamental frequency. The audio and electroglottographic signals of 18 classically trained male and female singers were recorded and analyzed with regard to vocal intensity, fundamental frequency, and open quotient. Fundamental frequency and open quotient are derived from the differentiated electroglottographic signal, using the DECOM (DEgg Correlation-based Open quotient Measurement) method. As male and female phonation may differ in respect to vocal-fold vibratory properties, a distinction is made between two different glottal configurations, which are called laryngeal mechanisms: mechanism 1 (related to chest, modal, and male head register) and mechanism 2 (related to falsetto for male and head register for female). The results show that open quotient depends on the laryngeal mechanisms. It ranges from 0.3 to 0.8 in mechanism 1 and from 0.5 to 0.95 in mechanism 2. The open quotient is strongly related to vocal intensity in mechanism 1 and to fundamental frequency in mechanism 2. [DOI: 10.1121/1.1850031]

PACS numbers: 43.75.Rs [SM] Pages: 1417–1430

I. INTRODUCTION

Voice quality is mainly due to the characteristics of vocal-fold vibratory movement. Thus, a better understanding of these properties would help to characterize voice quality. In this context, the open quotient O_q is a glottal source parameter of considerable interest, as it has been reported to be related to voice qualities such as ''breathy'' and ''pressed'' (e.g., Alku and Vilkman, 1996; Klatt and Klatt, 1990). It is defined as the ratio of the glottal open time over the fundamental period. It is a dimensionless parameter, ranging from 0 (no opening) to 1 (no or incomplete closure). This glottal source parameter is the counterpart of the closed quotient C_q, given that $O_q = 1 - C_q$. It can be measured directly by the use of high-speed visualization (Childers *et al.*, 1990; Timcke *et al.*, 1958), photoglottographic signals (Dejonckere, 1981; Hanson *et al.*, 1990; Kitzing, 1982, 1983; Kitzing *et al.*, 1982; Kitzing and Sonesson, 1974), electroglottographic (EGG) signals (Childers *et al.*, 1990; Hanson *et al.*, 1990; Lecluse, 1977; Lecluse and Brocaar, 1977; Miller *et al.*, 2002; Orlikoff, 1991). Indirect methods have also been used, based on inverse filtering of volume velocity or acoustic signals (Holmberg *et al.*, 1988, 1989, 1995; Sundberg *et al.*, 1999a). It has also been related to the amplitude difference $H_1^* - H_2^*$ between the first two harmonics of the acoustic signal spectrum after a formant-based correction (Hanson, 1995, 1997; Sundberg *et al.*, 1999a).

Most of these studies showed a variation of open quotient with vocal intensity in speech and singing. A decrease of open quotient with increase of vocal intensity was found with high-speed visualization of one male speaker (Timcke *et al.*, 1958), photoglottographic analysis of 20 female speakers (Kitzing and Sonesson, 1974), electroglottographic analysis of ten male speakers (Orlikoff, 1991), and glottal flow analysis of 25 male and 20 female speakers (Holmberg *et al.*, 1988), and six country singers (Sundberg *et al.*, 1999b). Only a slight trend was found by Hanson *et al.* (1990) on photoglottographic analysis of 12 male speakers. No relation was found by Lecluse and Brocaar (1977) on electroglottographic analysis of six untrained male singers, which could be explained by the use of a slightly different open-quotient definition, distinguishing opening time and open time.

The variation of open quotient with fundamental frequency has also been explored. In the case of male speakers, previous research has not shown any relationship between open quotient and fundamental frequency whatever the measurement method (Childers *et al.*, 1990; Hanson *et al.*, 1990; Lecluse and Brocaar, 1977; Timcke *et al.*, 1958). In the case of female speakers, an increase of open quotient with an increase of fundamental frequency was found by Kitzing and Sonesson (1974). Holmberg *et al.* (1989) observed that open

a)Electronic mail: henrich@lam.jussieu.fr

quotient tended to increase with fundamental frequency for both male and female speakers, although the correlation was not strong. Studying trained and untrained male and female singers, Howard (1995) reported a variation in open quotient with fundamental frequency in the case of female singers, depending on the singer's experience, but no relation between these parameters was found in the case of male singers (Howard *et al.*, 1990). No relation between open quotient and fundamental frequency was found in the case of six premier male country singers (Sundberg *et al.*, 1999b).

The differences found between male and female subjects may be due to the use of different laryngeal mechanisms, which differ with respect to the length and thickness of the vocal folds, as well as to the muscular laryngeal tensions involved in the process of voice production (Hirano, 1982; Roubeau, 1993). Indeed, voice production can be divided into four main laryngeal mechanisms, the evidence for such division being provided by the noticeable transitions in the electroglottographic signals during the production of a glissando (Henrich *et al.*, 2003; Roubeau, 1993; Roubeau *et al.*, 1991). The laryngeal mechanisms can be related to the well-known voice registers: *vocal fry* is produced in mechanism 0, the so-called *chest* or *modal* register and male *head* register are produced in mechanism 1, the *falsetto* register (male) or *head* register (female) are produced in mechanism 2 and the *flageolet*, or *whistle*, register is produced in mechanism 3. Mechanisms 1 and 2 are commonly used in speech and singing. In mechanism 1, the vocal folds are thick, leading to vertical phase differences in vibration, and longer closing and opening phases as compared to mechanism 2, where the vocal folds are thin and vibrate without any vertical phase difference (Hollien, 1974). Therefore, it is reasonable to expect lower open-quotient values in mechanism 1 than in mechanism 2. All the studies found in the literature seem to confirm this assumption (Kitzing, 1982; Lecluse, 1977; Lecluse and Brocaar, 1977). A study of the frequency jump at the transition between mechanisms in singing brought to light sudden changes in closed quotient that accompany or precede the transition (Miller *et al.*, 2002). In a study of male singers, Sundberg and Högset (1999) showed that the open-quotient differences between mechanisms were larger for baritones, as compared with tenors and counter tenors, who may even have equal open quotient in both mechanisms.

The open quotient seems to be strongly dependent on the laryngeal mechanism used by the speaker or the singer during vocal production. A study of the open quotient's relationship to parameters such as vocal intensity or fundamental frequency therefore needs to take laryngeal mechanism into account. Unfortunately, few studies have done so. This could partly explain the lack of convergence between studies on this issue.

To date, no study has been specifically devoted to a detailed exploration of the variations of open quotient in western operatic singing. Our purpose is to provide an overview of the variations of open quotient with vocal intensity and fundamental frequency for all the main tessituras in classical singing voice production, taking into account the laryngeal mechanisms involved. We hope that our results will help to make sense of the seemingly conflicting results of other studies. The voice database recorded for the purpose of this study and the analysis method are presented in Sec. II. In order to avoid the problems and limitations inherent in inverse filtering (Henrich, 2001; Henrich *et al.*, 2001), electroglottography was chosen as a noninvasive technique to measure the glottal vibratory movement, and the differentiated EGG signal (DEGG) was used for open-quotient measurements. In Sec. III, the results of the database analysis will be presented, pointing out the relations of open quotient with laryngeal mechanisms, vocal intensity, and fundamental frequency. These relations will be discussed in Sec. IV, and the main results will be summarized in Sec. V.

II. MATERIAL AND METHOD

A. Recording procedure

Recording sessions took place in a soundproof booth. The acoustic signal was recorded using a 1/2 in. condenser microphone (Brüel & Kjær 4165) placed 50 cm from the singer's mouth, a preamplifier (Brüel & Kjær 2669), and a conditioning amplifier (Brüel & Kjær NEXUS 2690). The electroglottographic signal was recorded by the use of a two-channel electroglottograph (EG2, Rothenberg, 1992). Both signals were recorded simultaneously on the two channels of a DAT recorder (PORTADAT PDR1000). A calibration for absolute sound-pressure level (SPL) measurement was carried out in each recording session by using the NEXUS amplifier to generate a reference tone, which was sent through the acquisition chain and recorded on the DAT recorder. In addition, an analog sound-level meter was placed close to the microphone, while the singers were asked to produce a sustained sound at a relatively steady loudness of their choice. The SPL recorded by the sound-level meter was noted and used later for assessing the validity of the reference-tone calibration procedure.

The singers were asked to stand still during the whole recording session, and their position was marked on the floor. As they were not physically constrained in terms of bodily movements, head movements could occur during the recording, which should have a second-order effect on the SPL measurements at 50 cm.

B. Subjects

Eighteen trained singers were recorded for this study: seven baritones (subjects B1 to B7), two tenors (T1,T2), three counter tenors (CT1 to CT3), three mezzo-sopranos (MS1 to MS3), and three sopranos (S1 to S3). Most of them were professional singers, earning their living from singing. Among other questions, the singers were asked to indicate in what range of pitch they used mechanisms 1 and 2. The answers are given in Fig. 1. Notice the presence of a frequency band where both mechanisms can be used. To produce these frequencies, the singer can thus choose to phonate either in mechanism 1 or in mechanism 2.

C. Protocol

For each singer, the recording session lasted about 20 min. The singer was asked to go through a precise protocol

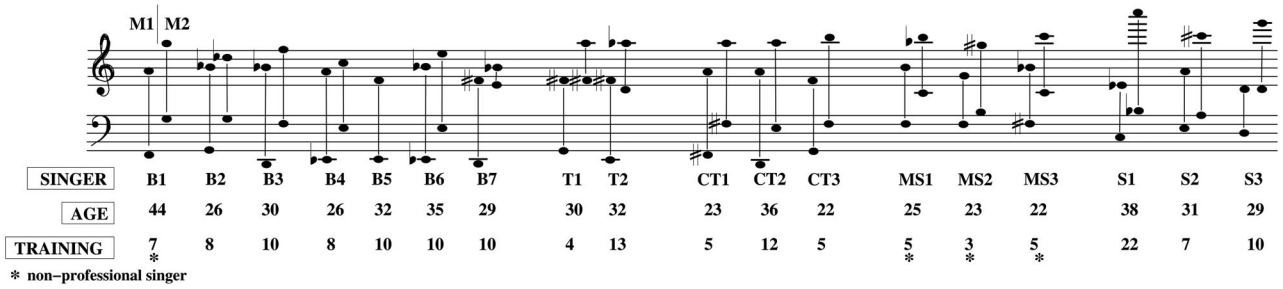


FIG. 1. Tessitura of the 18 singers recorded for this study, as given by the singer. For each singer, the pitch range is given in mechanism 1 (M1) and mechanism 2 (M2). In the case of baritone B5, no range is given in mechanism 2, as this singer did not use it at all and so had no idea about his range. Their age and the years of training are also given. Four of them are not professional singers (B1, MS1, MS2, and MS3).

(see below), and, if necessary, to describe his/her vocal production in terms of voice registers, or voice quality. Indeed, singers are apparently able to identify the laryngeal mechanism in which they phonate, information worth eliciting. We chose to use a noninvasive method for measuring glottal activity and only gathered information on vocal intensity (I), fundamental frequency (f_0), and open quotient (O_q). The protocol for this experiment is thus devised to study open-quotient variations as a function of fundamental frequency and vocal intensity for singing exercises (e.g., sustained vowels and crescendos) and for musical sentences (sung sentences). The subjects were asked to reduce the amount of vibrato if possible. The tasks were as follows:

(1) Speech/singing/shouting: A sentence in French, chosen by the subject, was first spoken, then sung, and finally shouted.
(2) Sung sentence: The first bars of Gounod's *Ave Maria* were performed with various degrees of loudness (piano, mezzo-forte, forte).
(3) Sustained vowels and crescendos/decrescendos: Three selected vowels [a], [e], and [u] were performed at different pitches depending on the singer's pitch range (see Table I) and at three degrees of loudness: piano, mezzo-forte, and forte. The subject was asked to maintain vowel color, pitch, and loudness during production. The sounds were 4 to 8 s long, and in cases when the laryngeal mechanism could not be straightforwardly identified, the subject indicated which laryngeal mechanism he or she was using. Following these tasks, the singers were asked to perform crescendos and decrescendos on the selected pitches and vowels.

TABLE I. Pitches sung by the singers for the sustained-vowels and crescendo-decrescendo tasks. When only the lower and upper pitches are given, the ascending scale is diatonic.

B1	C3, G3, C4, E4	CT1	D3, A3, D4, A4, D5
B2	C3, G3, C4, E4	CT2	D3, A3, D4, A4, D5
B3	C3, G3, C4, E4	CT3	B3 to E5
B4	C3, G3, C4, E4	MS1	C4, E4, G4, C5, E5
B5	A2 to C4	MS2	F3, G3, B3, C4, E4, G4, C5, E5
B6	C3 to C4	MS3	G3 to A4
B7	A2 to B3	S1	G4, C5, E5, G5
T1	F3 to C5	S2	G4, A4, C5, E5
T2	B2 to F4	S3	G4, C5, E5, G5, C6

(4) Glissandos: Rising and falling glissandos, mezzo-forte, continuous, and (if possible) without vibrato, were requested at the end of the session.

Each task was only performed once, apart from the glissandos which could be repeated several times. The use of both mechanisms was not mandatory in this study, and so the singers were not asked to repeat a task while singing in another laryngeal mechanism. However, some singers sometimes volunteered to do so. Singer CT1 repeated the sustained-vowels task on pitch D4 (293 Hz), mezzo-forte, in both mechanisms. Singers B2, B3, B4, B7, T2, CT1, CT2, and CT3 repeated either the spoken, the shouted, or the sung French sentence in both mechanisms. Singers T1, T2, CT1, CT2, MS1, and S2 sang the *Ave Maria* air in both mechanisms.

Another part of the protocol was dedicated to the exploration of various voice qualities. In the first task, the choice of voice quality was left to the singer. In the second task, voice qualities were specified, such as natural versus lyrical voice production. The exploration of these voice qualities from a perceptual and acoustical point of view is underway (Garnier *et al.*, 2004).

D. Analysis method

The fundamental frequency and the open quotient are measured from the DEGG signal, by using the DECOM (DEgg Correlation-based Open quotient Measurement) method as described in a previous paper (Henrich *et al.*, 2004). The method will be summarized here, and we refer the reader to that paper for more detail.

An EGG signal gives information about the vocal-fold contact area. A sudden variation in the contact will lead to noticeable peaks in the derivative (DEGG signal). These peaks can accurately be related to the glottal opening and closing instants, which are, respectively, defined as the instants at which the glottal flow starts to increase greatly from or decrease greatly toward the baseline (Childers *et al.*, 1990, 1983). The fundamental period can thus be derived from a DEGG signal by measuring the duration between two consecutive glottal closing instants. The duration between a glottal opening instant and the consecutive glottal closing instant corresponds to the open time. The open quotient can be de-

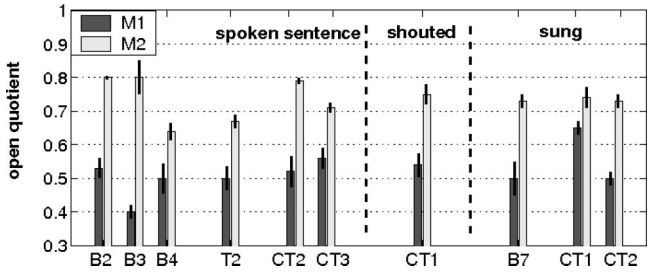


FIG. 2. Mean values and standard deviations of open quotient measured on the French sentences spoken, sung or shouted in mechanisms 1 (M1) and 2 (M2).

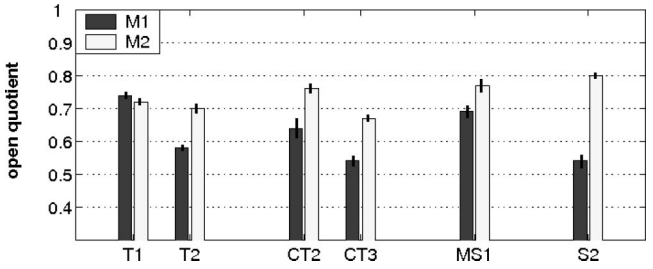


FIG. 3. Mean values and standard deviations of open quotient measured on the musical phrase *Ave Maria* of Gounod, sung in mechanisms 1 (M1) and 2 (M2).

rived from these two measures as the ratio between open time and fundamental period.

The DECOM method is applied to a four-period windowed DEGG signal which is separated into two parts: its positive part, which shows strong peaks related to glottal closing instants, and its negative part, which shows weaker peaks related to glottal opening instants. The fundamental period duration is derived from the autocorrelation function calculated on the positive part of the DEGG signal. The open time is derived from the intercorrelation function calculated between the positive part and the negative part.

These measures are accurate in the case where the glottal opening and/or closing peaks are single and precise. In some cases, however, the DEGG signal can present double or multiple peaks during the opening or the closing phase. Therefore, the DECOM method automatically detects double or undefined peaks, and only the measurements on glottal cycles for which the opening and closing peaks are unique are taken into account in this study.

III. RESULTS

This part presents the results of the open-quotient measurements. We will first deal with its relation to the laryngeal mechanisms and then explore the correlation with vocal intensity and fundamental frequency.

A. Open quotient and laryngeal mechanism

Several performing situations have been studied with regard to the open-quotient variation from one mechanism to another: on spoken and sung sentences, within the same pitch on sustained vowels, and during a glissando, i.e., a variation of fundamental frequency. As performing a task in both mechanisms was not obligatory, only a few cases are presented here, and they relate mainly to male voices.

1. Spoken and sung sentences

The results concerning the mean variation of open quotient during the French sentence, either spoken, shouted, or sung, are illustrated in Fig. 2. It should be noted that a given singer did not necessarily produce the three tasks (speech, singing, and shouting) using both mechanisms. For instance, it may be easier for a male singer to speak than to sing in mechanism 2 (M2). The three baritones and the tenor, whose results are plotted in Fig. 2, did not succeed in singing the sentences in both mechanisms. In most cases, vocal intensity is kept rather constant between M1 and M2, the differences in mean value being at most 10 dB with a mean difference of 4.2 dB. A great difference in the open-quotient mean values is observed in every case between mechanism 1 (M1) and mechanism 2 (M2). Whatever the vocal production is, the mean values of open quotient are between 0.4 and 0.65 in M1, and between 0.65 and 0.8 in M2. In the case of counter tenor CT1, who sang and shouted the sentence in both mechanisms, a difference between singing and shouting is found in M1. It goes along with a 10-dB increase in vocal intensity, which could be the effect of increased vocal effort in shouting. The relation between open quotient and vocal intensity will be developed in more detail in Sec. III B.

During the *Ave Maria* task, a few singers also managed to sing in both mechanisms, as shown in Fig. 3. Vocal intensity does not vary much between both productions in most cases, the differences in mean value between M1 and M2 being at most 8 dB with a mean difference of 3.8 dB. The pitch is quite different, allowing the singers to sing comfortably in one and the same mechanism over the whole sentence. Only counter tenor CT2 managed to sing the sentence in both mechanisms at the same pitch, with a mean vocal intensity of 77 dB in both cases. The differences in open quotient between the two laryngeal mechanisms are also obvious in these examples, except in the case of tenor T1. The greatest difference is found for soprano S2, with $O_q = 0.54$ in M1 and 0.8 in M2. In this case, mean vocal intensity is 81 dB in M1 and 88 dB in M2.

2. Sustained vowels

The use of one mechanism or another often goes with a change of fundamental frequency. However, counter tenor CT1 sang three sustained vowels [a], [e], and [u] in both mechanisms on the same pitch (D4, 293 Hz). The results are given in Table II and illustrated in Fig. 4 in the case of vowel [a].

A noticeable difference is found in open quotient between the two mechanisms. Laryngeal mechanism 1 is char-

TABLE II. Open quotient and vocal intensity [mean (standard deviation)] measured during sustained vowels sung by counter tenor CT1 in M1 or M2, at the same pitch (D4, 293 Hz).

	[a]		[e]		[u]	
	O_q	I	O_q	I	O_q	I
M1	0.64 (0.02)	88 (3)	0.64 (0.01)	83 (3)	0.65 (0.02)	82 (2)
M2	0.77 (0.02)	79 (3)	0.77 (0.04)	77 (3)	0.77 (0.03)	79 (3)

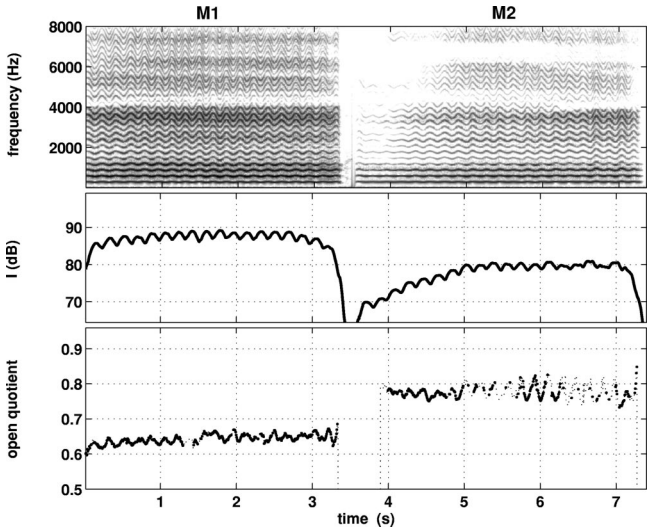


FIG. 4. Vowel [a] sung by counter tenor CT1 on the same pitch (D4) in mechanisms 1 and 2. From top to bottom panels: acoustic signal on a time-frequency space, vocal intensity, and open quotient.

acterized by a mean value of open quotient of 0.64 and mechanism 2 by a mean value of 0.77. These results are very similar to those obtained by this subject during the sung sentence (see Fig. 2). Mean vocal intensity is 88 dB in M1 and 79 dB in M2. The spectral analysis shows a noticeable decrease of energy in the high-frequency part of the spectrum in M2. Nevertheless, both phonations were judged as perceptually similar by the authors.

3. Glissandos

Transitions between laryngeal mechanisms have also been explored with regard to the laryngeal mechanisms. In the case of male singers, a transition from M1 to M2 usually occurs in the higher part of their vocal range: cases where the vocalis muscle is tensed and subglottal pressure is high (Miller, 2000), hence the probability of a noticeable frequency jump (Miller *et al.*, 2002; Roubeau, 1993; Svec *et al.*, 1999) and of a decrease of vocal intensity (Roubeau, 1993). The transition between laryngeal mechanisms can also be detected by an amplitude change in the envelope of the EGG and DEGG signal (Henrich *et al.*, 2003; Roubeau *et al.*, 1987).

A glissando sung by tenor T2 with noticeable frequency jumps is presented in Fig. 5. The transition M1→M2 goes with a frequency jump of 3 semitones (F4#→A4), and the transition M2→M1 goes with a frequency jump of 5–6 semitones (F4→C4). These results are in agreement with those obtained from three tenors by Miller *et al.* (2002) in studying the characteristic leap interval from chest register (M1) to falsetto (M2).

These pitch jumps go together with noticeable jumps of open quotient. During the transition M1→M2, open quotient varies from 0.4 to 0.62 within approximately 300 ms, and this slow variation precedes the jump in frequency. Similar observations were made by Miller *et al.* (2002), who found that the variation in open quotient lasted about 100 ms. The variation of open quotient for the transition M2→M1 is similar in amplitude, O_q decreasing from 0.75 to 0.53, but more sudden.

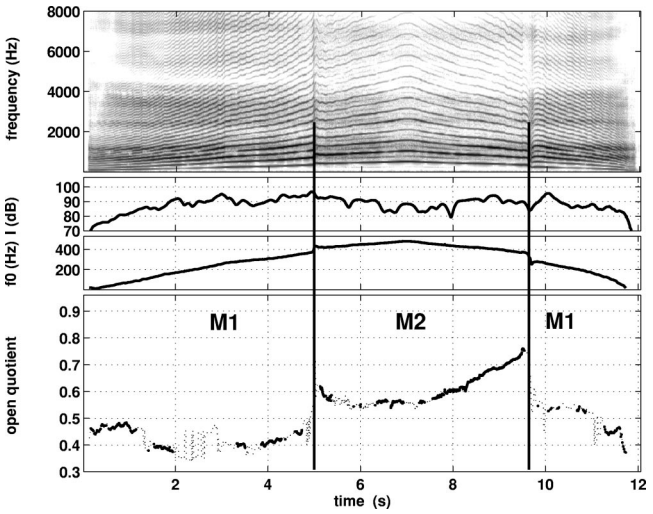


FIG. 5. Glissando sung by tenor T2. From top to bottom panels: acoustic signal in a time-frequency space, vocal intensity, fundamental frequency, and open quotient. The open-quotient measures for which glottal opening and closing peaks are unique are plotted with thick dots. Information about the measures obtained while the peaks are imprecise or double is given by the broken line. The two vertical lines indicate the transition between laryngeal mechanisms M1 and M2. Their placement is based on the pitch jump detection and the amplitude change in the EGG and DEGG signal.

In classical singing, singers who need to develop their vocal range over the two laryngeal mechanisms (such as male and female altos) learn how to ''smooth'' the transition from one mechanism to another, so as to avoid any noticeable timbre discontinuity in the melodic line. This is best illustrated by counter tenor CT1, whose glissando is presented in Fig. 6: the frequency jump is unnoticeable. Both transitions are characterized by a noticeable decrease in vocal intensity and high-frequency spectral energy, and by a

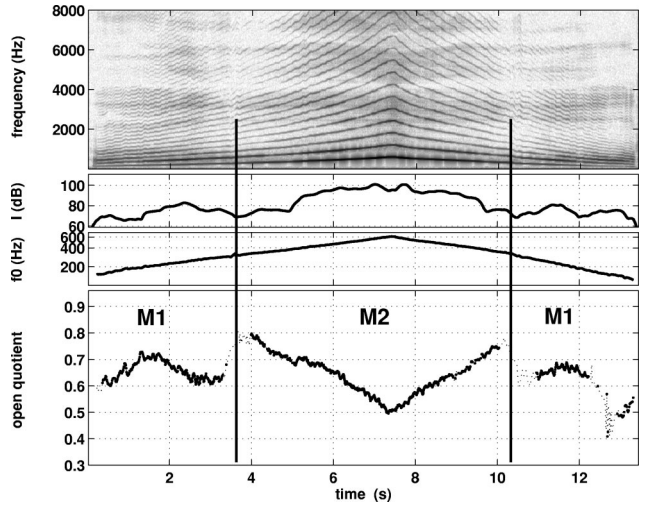


FIG. 6. Glissando sung by counter tenor CT1. From top to bottom panels: acoustic signal in a time-frequency space, vocal intensity, fundamental frequency, and open quotient. The open-quotient measures for which glottal opening and closing peaks are unique are plotted with thick dots. Information about the measures obtained while the peaks are imprecise or double is given by the broken line. The two vertical lines indicate the transition between laryngeal mechanisms M1 and M2. Their placement is based on the amplitude change in the EGG and DEGG signals.

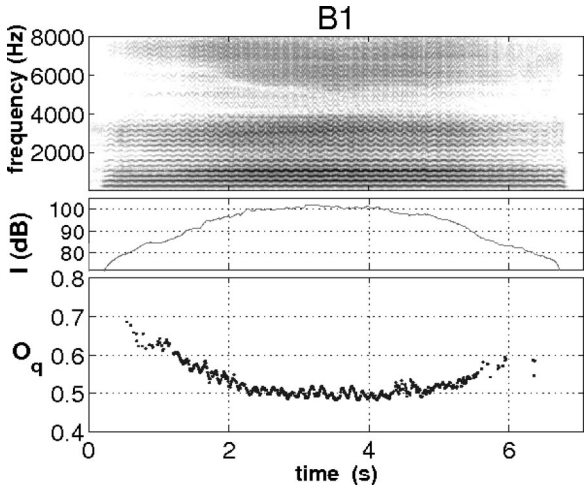


FIG. 7. Crescendo–decrescendo sung by the baritone B1 in mechanism 1, vowel [a], pitch C4.

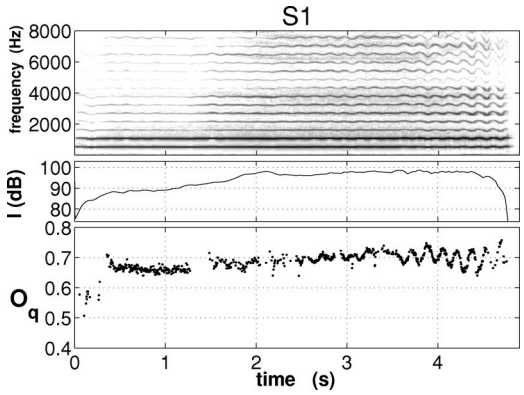


FIG. 8. Crescendo sung by the soprano S1 in mechanism 2, vowel [u], pitch C5.

change in the EGG and DEGG overall amplitude. Nevertheless, the transitions are barely audible at first listening. The open quotient varies from 0.62 to 0.78 at the transition M1→M2 and from 0.76 to 0.6 at the transition M2→M1. The open-quotient variations are reduced as compared to the case of the tenor T2, but they are still noticeable. A strong negative correlation can be observed between open quotient and fundamental frequency in M2. Such a tendency is found for all the glissandos sung by this singer, the other two counter tenors, the two tenors (see for instance Fig. 5), and the three sopranos. This point will be addressed later in Sec. III C.

The open-quotient variation close to a transition of laryngeal mechanisms is a common feature observed for all the singers, male and female. The amplitude of the jump ranges from 0.1 to 0.3. An extensive quantitative analysis of the open-quotient jump in glissandos is beyond the scope of this present study, and we will see in the following sections that many factors contribute to a variation of open quotient.

B. Open quotient and vocal intensity

The measures presented here result from the analysis of crescendos–decrescendos and sustained vowels sung at three degrees of loudness, performed at various pitches covering the singer's vocal range, and the analysis of *Ave Maria* musical phrases.

1. Crescendos–decrescendos and sustained vowels

A crescendo–descrescendo sung by baritone B1 on the vowel [a] at pitch C4 (260 Hz) is shown in Fig. 7. A strong correlation between open quotient and vocal intensity can be observed: the greater the vocal intensity, the lower the open quotient. A 20-dB increase of vocal intensity goes along with a decrease of open quotient from 0.7 to 0.5. It goes along with a spectral enhancement of the first formant region and an increase of the harmonic richness in the high-frequency part of the spectrum. This crescendo–decrescendo was produced in mechanism 1. A crescendo in mechanism 2 sung by soprano S1 is shown in Fig. 8. In this case, vocal intensity increases by about 10 dB and open quotient remains stable at about 0.7.

In both cases, pitch is kept constant, suppressing possible dependency between vocal intensity and fundamental frequency. As compared to ordinary speech, the singing technique comprises the ability to change from soft to loud while keeping the pitch constant, and to change pitch while maintaining loudness under control. But, this is only true within a given pitch range. On a voice range profile or phonetogram, a strong correlation exists between the variations of vocal intensity and fundamental frequency (Gramming *et al.*, 1988; Liénard and Di Benedetto, 1999; Titze and Sundberg, 1992), though these two parameters can be modified locally in an independent way. This general tendency is observed here as well, as illustrated in Table III: the data resulting from the analysis of crescendos–decrescendos and sustained vowels have been pooled across vocal intensity and fundamental frequency, and Pearson's correlation coefficient between these two variables has been calculated, with the laryngeal mechanism and the vowel as additional parameters. The correlation between vocal intensity and fundamental frequency seems even stronger in mechanism 2 than in mechanism 1. A few exceptions to this general trend can, however,

TABLE III. Pearson's correlation coefficient between vocal intensity and fundamental frequency, in the case of sustained vowels and crescendos sung in M1 and M2, for the three vowels [a], [e], and [u]. A strong correlation is indicated in bold ($r>0.70$). A nonsignificant correlation is indicated by n.s. ($p\geqslant 0.001$). The corresponding degrees of freedom are given in Table VIII.

M1				M2			
Singer	[a]	[e]	[u]	Singer	[a]	[e]	[u]
B1	**0.73**	0.47	**0.75**				
B2	0.68	0.69	**0.84**	CT1	**0.90**	**0.83**	0.63
B3	0.52	0.55	0.47	CT2	0.60	0.68	0.64
B4	**0.81**	**0.75**	0.69	CT3	**0.75**	**0.86**	**0.84**
B5	0.50	0.62	**0.73**				
B6	0.11	0.18	n.s.	MS1	**0.75**	**0.79**	**0.75**
B7	**0.72**	0.22	0.23	MS2	0.70	**0.93**	**0.92**
				MS3	**0.86**	**0.80**	**0.86**
T1	0.69	**0.88**	**0.94**				
T2	0.45	0.60	0.59	S1	0.65	**0.83**	**0.74**
				S2	**0.78**	**0.85**	**0.87**
CT1	0.64	0.49	0.66	S3	**0.83**	**0.78**	**0.78**
CT2	0.28	0.18	0.47				

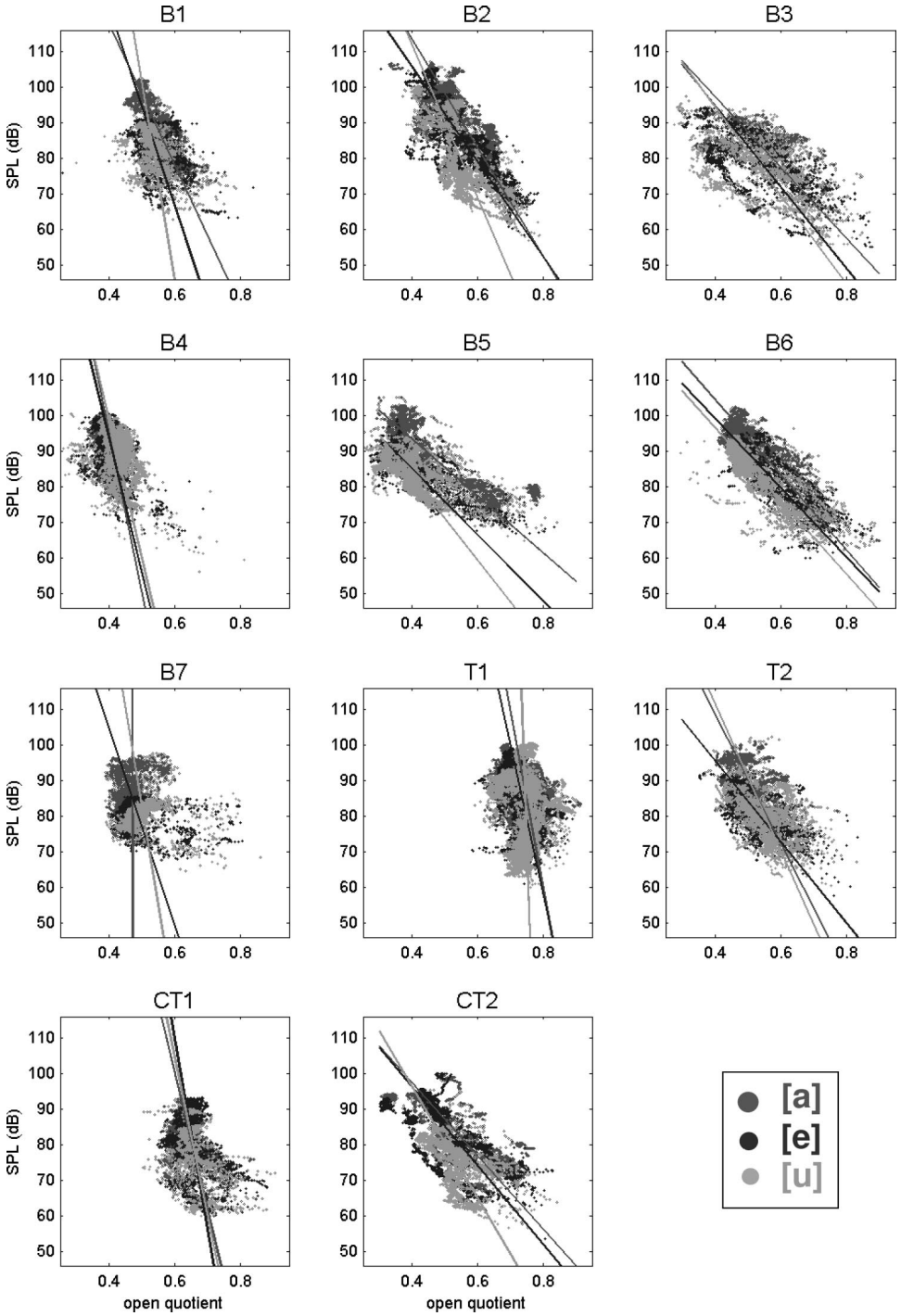


FIG. 9. Vocal intensity as a function of open quotient in the case of mechanism 1 for the seven baritones, the two tenors, and two counter tenors. The vowels have been distinguished by using a gray scale. The lines correspond to the major axes for a given vowel.

be noticed: baritone B6, baritone B7 for vowels [e] and [u], and counter tenor CT2 in mechanism 1.

While comparing open quotient and vocal intensity at different pitches, we should thus always keep in mind that the variation of vocal intensity may be due to a variation of fundamental frequency. So as to take this possible underlying variation into account in the statistical analysis of our results, we shall introduce a partial correlation coefficient (see the Appendix).

a. Vocal production in mechanism 1. Vocal intensity has been plotted as a function of open quotient for each male singer in mechanism 1, as shown in Fig. 9. The three vowels [a], [e], and [u] have been processed separately, but the resulting measurements are plotted on the same figure. For each vowel, a regression line (or major axis) between open quotient and vocal intensity is plotted on the figure. Table IV gives the corresponding Pearson correlation coefficients and partial correlation coefficients.

Vocal intensity ranges from 60–70 dB for *piano* sounds, to 95–105 dB for *forte* sounds. The open quotient ranges from 0.3 to 0.9. It seldom goes below 0.5 for the singers T1 and CT1, and it seldom goes beyond 0.7 in the case of the

TABLE IV. Pearson's correlation coefficient between open quotient and vocal intensity, in the case of sustained vowels and crescendos sung in mechanism 1 and for the 3 vowels [a], [e], and [u]. The partial correlation coefficient is given in parentheses. A strong correlation is indicated in bold ($r>0.70$). A nonsignificant correlation is indicated by n.s. ($p\geqslant 0.001$). The corresponding degrees of freedom are given in Table VIII.

Singer	[a]	[e]	[u]
B1	**−0.78** (**−0.72**)	−0.43 (−0.48)	−0.25 (−0.20)
B2	**−0.76** (**−0.79**)	**−0.83** (**−0.76**)	−0.68 (−0.46)
B3	−0.60 (**−0.78**)	−0.62 (**−0.83**)	−0.57 (**−0.76**)
B4	−0.44 (−0.30)	−0.49 (−0.41)	−0.43 (−0.24)
B5	**−0.84** (**−0.90**)	**−0.76** (−0.65)	−0.66 (−0.45)
B6	**−0.87** (**−0.86**)	**−0.84** (**−0.83**)	**−0.85** (**−0.85**)
B7	n.s. (n.s.)	−0.21 (−0.43)	−0.09 (−0.34)
T1	−0.30 (0.07)	−0.43 (n.s.)	−0.08 (−0.12)
T2	−0.60 (**−0.71**)	−0.69 (−0.55)	−0.49 (−0.31)
CT1	−0.39 (−0.55)	−0.33 (−0.26)	−0.39 (−0.37)
CT2	**−0.75** (**−0.92**)	**−0.75** (**−0.94**)	−0.57 (**−0.77**)

baritone singers. The trend is similar for the six baritones (B1 to B6), the tenor T2, and the counter tenors CT1 and CT2: the open quotient decreases as vocal intensity increases. A strong partial correlation ($r>0.70$) is found for singers B1, B2, B3, B5, B6, T2, and CT2 in the case of vowel [a], for singers B2, B3, B6, and CT2 in the case of vowel [e], and for singers B3, B6, and CT2 in the case of vowel [u]. The nature of the vowel may have an effect on glottal open phase, but this effect is not further explored in the present study.

The results of the two baritones B4 and B7 and of the tenor T1 differ from those of the other singers. In the case of B4, the variations in open quotient are limited, ranging from 0.35 to 0.5 whatever the pitch and vocal intensity. In the case of B7, the open quotient ranges mainly from 0.4 to 0.55 and it seldom goes beyond 0.6. For a given vocal intensity, the results of these two baritones are similar. In the case of T1, the variations of open quotient are limited, with values ranging from 0.65 to 0.85. Such values are much more often found in the case of vocal production in mechanism 2.

b. Vocal production in mechanism 2. The measurements

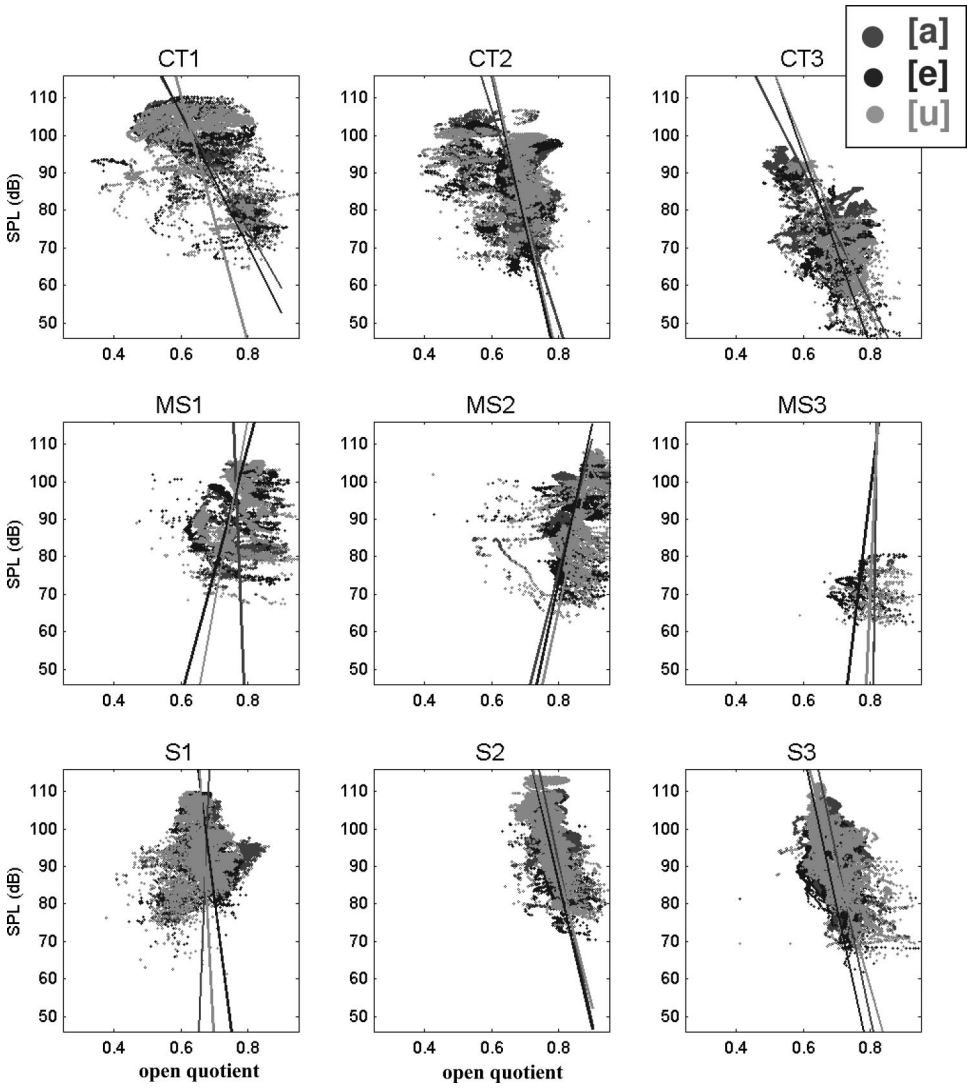


FIG. 10. Vocal intensity as a function of open quotient in the case of mechanism 2 for the three counter tenors, the three mezzo-sopranos, and the three sopranos. The vowels have been distinguished by using a gray scale. The lines correspond to the major axes for a given vowel.

TABLE V. Pearson's correlation coefficient between open quotient and vocal intensity, in the case of sustained vowels and crescendos sung in mechanism 2 and for the three vowels [a], [e], and [u]. The partial correlation coefficient is given in parentheses. A strong correlation is indicated in bold ($r>0.70$). A nonsignificant correlation is indicated by n.s. ($p\geqslant 0.001$). The corresponding degrees of freedom are given in Table VIII.

Singer	[a]	[e]	[u]
CT1	−0.63(0.42)	−0.55(0.27)	−0.26(0.40)
CT2	−0.44(0.07)	−0.33(0.33)	−0.29(0.29)
CT3	−0.56(0.21)	−0.50(0.44)	−0.61(−0.09)
MS1	n.s. (−0.54)	0.33(−0.31)	0.24(−0.46)
MS2	0.44(0.22)	0.47(−0.09)	0.49(0.23)
MS3	n.s. (−0.41)	n.s. (−0.30)	n.s. (−0.29)
S1	0.06(0.68)	−0.26(0.53)	−0.10(0.64)
S2	−0.42(0.07)	−0.60(−0.15)	−0.64(n.s.)
S3	−0.54(−0.14)	−0.39(−0.37)	−0.53(−0.41)

of vocal intensity as a function of open quotient are shown in Fig. 10 for each singer in mechanism 2 (the three counter tenors, the three mezzo-sopranos, and the three sopranos). Table V gives the corresponding Pearson correlation coefficients and partial correlation coefficients.

No strong correlation is observed. A change of sign is even observed between Pearson's correlation coefficient and the partial correlation coefficient in the case of the three counter tenors, the mezzo-soprano MS1, and the soprano S1. In the case of the three counter tenors, for instance, the trend is a decrease of open quotient with an increase of vocal intensity, with regard to the Pearson correlation coefficient. The opposite trend is found if the underlying variation of fundamental frequency is taken into account with the partial correlation coefficient. It could be explained by the strong correlation between fundamental frequency and vocal intensity (a summary of the data is found in Table III).

2. Analysis of sung phrases

The singers sang the first bars of Gounod's *Ave Maria*, using various degrees of vocal loudness (piano, mezzo-forte, and forte). The mean and standard values of open quotient measured over the whole length of the musical sentence are given in Fig. 11 for mechanism 1 and Fig. 12 for mechanism 2. In mechanism 1, the trend is similar to the one previously observed in the case of crescendos–decrescendos and sustained vowels: the open quotient decreases when the vocal intensity increases. Similarly to what was observed in the previous section, this trend is strong in the case of singers B1, B2, B6, T2, and CT2, and it is not found in the case of singers B4, B7, and T1.

A paired samples t-test (Daudin *et al.*, 1999) conducted on the means shows that the decrease of open quotient is very significant between the piano and mezzo-forte productions [$t(7)=3.74$, $p<0.01$] as well as between the mezzo-forte and forte productions [$t(7)=3.80$, $p<0.01$], if the results of singer T1 are not included in the test. If his results are included in the statistical analysis, the decrease of open quotient with an increase of vocal intensity remains significant [$t(8)=3.41$, $p<0.01$ for piano/mezzo-forte; $t(8)=2.91$, $p<0.05$ for mezzo-forte/forte].

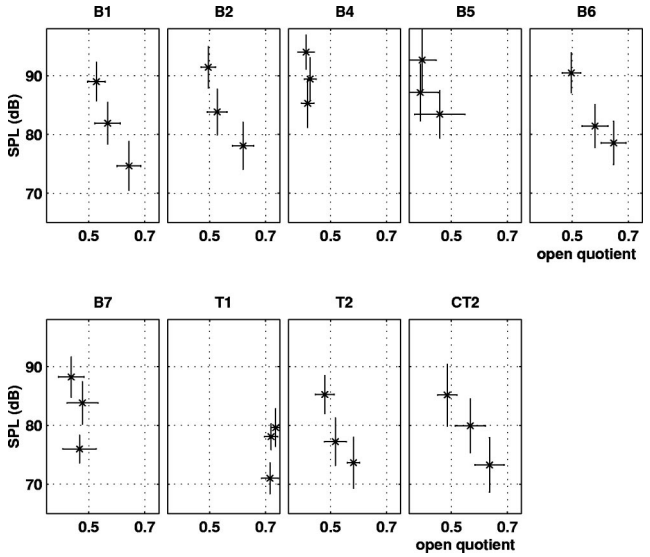


FIG. 11. Vocal intensity as a function of open quotient measured from the phrase *Ave Maria* of Gounod, sung with three different degrees of vocal loudness (piano, mezzo-forte, and forte) in the case of mechanism 1 for six baritones, the two tenors, and one counter tenor. The bars give the standard deviations for both parameters. No result can be given in the case of baritone B3, for whom the three corresponding DEGG signals present a double-peak feature at glottal opening almost throughout the phrase.

In mechanism 2, great differences of behavior are observed across singers, and, in each case, these behaviors agree with the trend pointed out by the partial correlation coefficient in the case of crescendos–decrescendos and sustained vowels. A decrease of open quotient with an increase of vocal intensity is found for singers MS1, MS3, and S3. On the contrary, an increase of open quotient with vocal intensity is found for singers CT3 and S1. No noticeable trend comes out for singers CT2 in mechanism 2, MS2, and S2. No statistical difference is found for open quotient between the three degrees of loudness [$t(8)=0.53$, $p>0.5$ for piano/

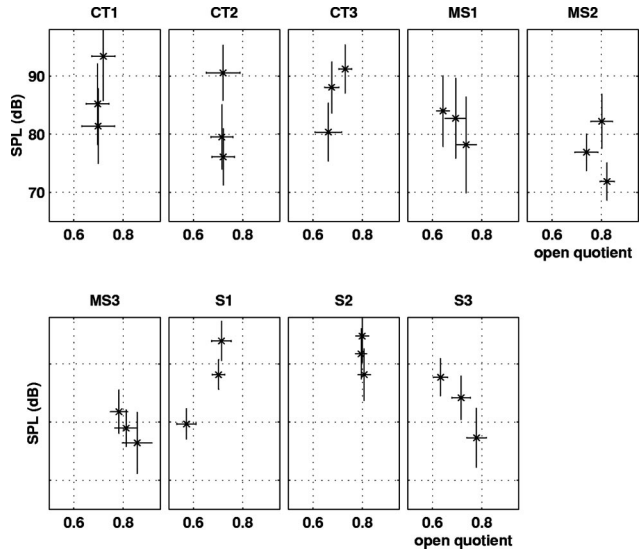


FIG. 12. Vocal intensity as a function of open quotient measured on the sentence *Ave Maria* of Gounod, sung with three different degrees of vocal loudness (piano, mezzo-forte, and forte) in the case of mechanism 2 for the three counter tenors, the three mezzo-sopranos, and the three sopranos. The bars give the standard deviations for both parameters.

TABLE VI. Pearson's correlation coefficient between open quotient and fundamental frequency, in the case of sustained vowels and crescendos sung in mechanism 1 and for the three vowels [a], [e], and [u]. The partial correlation coefficient is given in parentheses. A strong correlation is indicated in bold ($r>0.70$). A nonsignificant correlation is indicated by n.s. ($p \geqslant 0.001$). The corresponding degrees of freedom are given in Table VIII.

Singer	[a]	[e]	[u]
B1	−0.47(0.22)	n.s. (0.24)	−0.15(n.s.)
B2	−0.29(0.45)	−0.51(0.15)	−0.57(n.s.)
B3	0.13(0.64)	0.12(**0.71**)	0.19(0.63)
B4	−0.34(n.s.)	−0.30(0.11)	−0.39(−0.15)
B5	−0.12(0.62)	−0.53(−0.11)	−0.56(−0.15)
B6	−0.11(n.s.)	−0.15(n.s.)	−0.11(−0.13)
B7	n.s. (n.s.)	0.60(0.68)	0.67(**0.71**)
T1	−0.50(−0.42)	−0.49(−0.27)	n.s. (0.10)
T2	n.s. (0.49)	−0.51(−0.17)	−0.44(−0.22)
CT1	n.s. (0.42)	−0.22(n.s.)	−0.18(n.s.)
CT2	0.32(**0.83**)	0.46(**0.90**)	0.20(0.65)

TABLE VII. Pearson's correlation coefficient between open quotient and fundamental frequency, in the case of sustained vowels and crescendos sung in mechanism 2 and for the three vowels [a], [e], and [u]. The partial correlation coefficient is given in parentheses. A strong correlation is indicated in bold ($r>0.70$). A nonsignificant correlation is indicated by n.s. ($p \geqslant 0.001$). The corresponding degrees of freedom are given in Table VIII.

Singer	[a]	[e]	[u]
CT1	−**0.82**(−**0.74**)	−**0.78**(−0.69)	−**0.74**(−**0.77**)
CT2	−**0.79**(−**0.73**)	−**0.73**(−**0.73**)	−**0.71**(−**0.71**)
CT3	−**0.85**(−**0.78**)	−**0.75**(−**0.73**)	−0.69(−0.41)
MS1	0.36(0.62)	0.61(0.61)	0.63(**0.71**)
MS2	0.42(0.18)	0.53(0.31)	0.44(n.s.)
MS3	0.26(0.47)	0.37(0.45)	0.22(0.35)
S1	−0.57(−**0.80**)	−0.60(−**0.71**)	−0.60(−**0.79**)
S2	−0.59(−0.46)	−0.64(−0.30)	−**0.73**(−0.46)
S3	−0.57(−0.26)	−0.20(0.17)	−0.37(0.08)

mezzo-forte; $t(8)=0.21$, $p>0.5$ for mezzo-forte/forte], whereas vocal intensity varies significantly [$t(8)=6.12$, $p<0.001$ for piano/mezzo-forte; $t(8)=4.60$, $p<0.01$ for mezzo-forte/forte].

C. Open quotient and fundamental frequency

As shown in Fig. 6, a strong correlation between open quotient and fundamental frequency can be observed on the glissando sung by counter tenor CT1 in mechanism 2. Although less obvious, such a correlation can also be observed in the case of the glissando sung by tenor T2 in mechanism 2 (see Fig. 5). This correlation seems to depend on the laryngeal mechanism, as it is not found in the parts of the glissandos sung in mechanism 1. We will now try to characterize this correlation between open quotient and fundamental frequency with regard to the laryngeal mechanisms, in analyzing the measurements made on the sustained vowels and the crescendos–decrescendos sung at different pitches. It should be mentioned that, in the present study, the glissandos were only used to illustrate the relation between open quotient and fundamental frequency, and that the corresponding data have not been used for statistical analysis. Indeed, the glissandos have not been recorded for the purpose of proper statistical analysis, and in particular, the frequency range and the vowel have been left to the singer's choice, which implies that the glissandos do not cover the whole frequency range of a given singer in a given laryngeal mechanism. In addition, we were interested to see whether different loudness conditions would affect the relation, and this point could not be studied on glissandos, where the loudness is less easily controlled by the singer.

1. Crescendos–decrescendos and sustained vowels

a. Vocal production in mechanism 1. Table VI gives Pearson's correlation coefficients and partial correlation coefficients between open quotient and fundamental frequency in the case of the male singers in mechanism 1. Generally speaking, no strong correlation is found between O_q and f_0. Nevertheless, ''singer-dependent'' behavior can be noticed. A high positive partial correlation is found between O_q and f_0 in the case of baritones B3 and B7, and counter tenor CT2, for whom an increase of fundamental frequency goes together with an increase in open quotient. In the case of singers B3 and CT2, this effect of an increase in open quotient together with an increase of fundamental frequency is compensated for by the effect of an underlying increase of vocal intensity (related to a decrease in open quotient in M1), as Pearson's correlation is low whereas partial correlation is high.

b. Vocal production in mechanism 2. Table VII gives the Pearson correlation coefficients and partial correlation coefficients between open quotient and fundamental frequency in the case of the counter tenors in mechanism 2, the mezzo-sopranos, and the sopranos. A strong correlation between O_q and f_0 is observed in the case of the three counter tenors and the sopranos S1 and S2, which confirms the observations made on the glissandos. An increase of fundamental frequency goes along with a decrease of open quotient. No correlation is found in the case of soprano S3. The mezzo-sopranos present an inverse correlation, i.e., an increase of open quotient with an increase of fundamental frequency, but this effect is not strong.

IV. DISCUSSION

A. A strong dependency on laryngeal mechanisms

The results converge to confirm that the open quotient is dependent on the laryngeal mechanism used during the vocal production: open quotient values are lower in mechanism 1 than in mechanism 2. These differences may reflect the physiological differences between both laryngeal mechanisms, with regard to the thickness, the vibratory length, and the tension of the vocal folds. Thus, one can infer that, in those studies where the open quotient is a parameter of interest, it is of great importance to specify the laryngeal mechanism in which the voiced sound is produced.

The results of tenor T1 are surprising, as no difference is found between M1 and M2 (see Fig. 3). His vocal production is often considered as ''*voix mixte*'' (mixed voice) by his singing teachers. The analysis of his vocal production shows that this singer is always using high values of open quotient

($O_q \geqslant 0.7$), even during the glissandos with noticeable voice breaks. This result suggests that, in the case of this tenor, the so-called *voix mixte* could be characterized by the use of mechanism 1 but with unusually high values of open quotient so as to mimic the voice quality of mechanism 2. Further investigation is needed to characterize this vocal production, which is presently under study (Castellengo *et al.*, 2004; Chuberre, 2000; Expert, 2003).

Can the open quotient be considered as an indicator of the laryngeal mechanism? In most cases, the laryngeal mechanism predicts the open-quotient range. Nevertheless, the results of tenor T1 show that the open-quotient measure alone does not suffice to determine which laryngeal mechanism is involved in a given vocal production. Indeed, there is a degree of overlap between the O_q ranges corresponding to laryngeal mechanisms M1 and M2. For this purpose, open-quotient measurement should be combined with other means of characterization, such as listening to and analysis of the acoustic signal, and visualization of EGG or DEGG signals.

B. Correlation between open quotient and vocal intensity in M1

The results show that vocal intensity tends to be negatively correlated with open quotient in mechanism 1, for singing exercises (sustained vowels and crescendos–decrescendos), as well as for one musical sentence. This trend is not found in the case of mechanism 2, where the opposite trend can even be observed for some singers.

These results are in agreement with the observations of Dromey *et al.* (1992); Holmberg *et al.* (1988); Kitzing and Sonesson (1974); Orlikoff (1991); Sundberg *et al.* (1999b); and Timcke *et al.* (1958). As these previous studies were conducted with different exploratory methods (high-speed cinematography, photoglottography, electroglottography, inverse-filtered glottal flow), it underlines the agreement found between these methods and the measurement method based on the derivative of the EGG signal in the case of singing.

An increase in vocal intensity results from many factors, and these results suggest that the strategies used for increasing vocal intensity in mechanism 1 differ in some aspects from the ones used in mechanism 2. Whichever laryngeal mechanism is considered, an increase in vocal intensity generally results from an increase of subglottal pressure (Gauffin and Sundberg, 1989; Holmberg *et al.*, 1988; Isshiki, 1964; Karlsson, 1986; Ladefoged and McKinney, 1963; Lecuit and Demolin, 1998a,b; Schutte, 1980; Sundberg *et al.*, 1993; Tanaka and Gould, 1983; Titze and Sundberg, 1992). On the other hand, the activity of the vocalis muscle is strongly dependent on the laryngeal mechanism involved (Hirano, 1982; Roubeau, 1993). In mechanism 1, the vocalis contraction directly affects the glottal vibratory movement, and thus can have an impact on vocal intensity. In the highest part of mechanism 1, this muscle reaches its physiological limit of contraction. Thus, in mechanism 2, the tension of this muscle is reduced, whereas the crico-thyroid muscles are more activated (Hirano, 1982). The decrease in open quotient observed in mechanism 1 may thus be induced by the contraction of the vocalis muscle when vocal intensity is increased. This hypothesis accounts for the results obtained in mechanism 2, where the action of the vocalis muscle is limited and no correlation is found between open quotient and vocal intensity.

C. Correlation between open quotient and fundamental frequency in M2

In mechanism 2, a strong correlation between open quotient and fundamental frequency is found in the case of the counter tenors and the sopranos: an increase of fundamental frequency goes along with a decrease of open quotient. In mechanism 1, the open quotient seems not to be correlated to fundamental frequency. These results are in agreement with the observations made by Howard (1995); Howard *et al.* (1990), who noticed a decrease of open quotient with an increase of fundamental frequency in the case of trained female singers, and no correlation between these two parameters in the case of male singers. A comparison between previous studies and the present findings suggests that the gender differences reported in those studies can in fact be ascribed to a difference in the laryngeal mechanism involved, as also suggested by Holmberg *et al.* (1989): male subjects generally phonate in mechanism 1; female subjects more often phonate in mechanism 2.

Howard (1995) found that the correlation between open quotient and fundamental frequency observed on female singers is dependent on vocal training. The correlation is weak for the untrained female singers and it increases with the years of training. This could explain the results obtained in the case of the three mezzo-sopranos, who are not professional singers and had fewer years of training than the other singers recorded for this study.

In a few cases for male singers in M1, an increase of open quotient goes with an increase of fundamental frequency. This trend has already been observed by Childers *et al.* (1990); Cookman and Verdolini (1999); Hanson *et al.* (1990); Holmberg *et al.* (1989); and Kitzing and Sonesson (1974). However, the fundamental frequency is often highly correlated with the vocal intensity. As the open quotient decreases with an increase of vocal intensity in mechanism 1, this effect may compensate for an increase due to fundamental frequency and reduce in most cases the correlation between O_q and f_0.

D. Smoothing the transition between laryngeal mechanisms

As shown in Fig. 6, counter tenor CT1 managed to smooth the transition between laryngeal mechanisms, whereas the glissando sung by tenor T2 (see Fig. 5) is a good illustration of an abrupt transition. Prior to the transition ($M1 \rightarrow M2$), the major difference between the two cases is that the counter tenor sings at a lower vocal intensity ($I \simeq 72$ dB) and has higher open-quotient values (between 0.6 $\leqslant O_q \leqslant 0.7$) than the tenor ($I \simeq 95$ dB and $0.4 \leqslant O_q \leqslant 0.5$). The correlation between vocal intensity and open quotient in M1 implies that a decrease of vocal intensity in M1 close to the laryngeal mechanism transition goes along with an increase of open quotient. If the open-quotient values between

M1 and M2 are similar, the jump in open quotient at the transition is then reduced. This may help the singer to avoid the jump in fundamental frequency, and thus any perceptual voice break. Therefore, it seems that the smoothing of a transition between laryngeal mechanisms results from a lowering of vocal intensity prior to the transition and an increase of open quotient, so as to reach the value range which is common to both mechanisms.

This ''smoothing'' technique can be used by any skilled singer, e.g., tenor T2, who can also sing the glissandos without any voice breaks.

V. CONCLUSION

In this study, 18 trained male and female singers were recorded and their vocal production was analyzed with regard to the open quotient and its correlation with vocal intensity and fundamental frequency. This study shows the importance of taking into account the laryngeal mechanism in which the vocal sound is produced. The open quotient can be seen as an indicator of the laryngeal mechanism within a given voice production. Lower values of open quotient are usually found in mechanism 1, as compared to mechanism 2: the open quotient ranges from 0.3 to 0.8 in mechanism 1 and from 0.5 to 0.95 in mechanism 2. Yet, one should not rely solely on the values of open quotient to determine which laryngeal mechanism is used as there is a degree of overlap between the O_q ranges corresponding to laryngeal mechanisms M1 and M2. Listening to the corresponding sound samples can provide additional information and help to characterize a given vocal production. Nevertheless, the ear can sometimes be tricked by the vocal technique of the singer. For this reason, a combination of analysis, listening, and measuring of open quotient and other acoustical and EGG parameters is usually required to determine which laryngeal mechanism is being used.

The link between open quotient and vocal intensity depends on the laryngeal mechanism. In mechanism 1, they tend to be correlated: the greater the vocal intensity, the lower the open quotient. No correlation is found in mechanism 2. This effect could result from the activity of the vocalis muscle, which is reduced in mechanism 2 as compared to mechanism 1.

In the same way, the link between open quotient and fundamental frequency is also dependent on the laryngeal mechanisms. There tends to be a correlation in mechanism 2: the higher the fundamental frequency, the lower the open quotient. This correlation varies a bit from subject to subject. It is especially strong in the case of the counter tenors. No correlation was found in mechanism 1.

Another repeated observation is the large amount of variation across subjects. The same observation had already been made by Schutte (1980) when studying vocal efficiency. Differences in behavior are noticeable among singers within the same tessitura as well as between tessituras. In particular, only two tenors were recorded and their behaviors diverged markedly. It would be of great interest to record more singers within a given tessitura, in order to generalize the trends observed in the present study.

TABLE VIII. Degrees of freedom corresponding to the measures made on sustained vowels and crescendos sung in mechanisms 1 and 2 for the three vowels [a], [e], and [u].

M1				M2			
Singer	[a]	[e]	[u]	Singer	[a]	[e]	[u]
B1	4 323	2 548	1 867				
B2	5 314	5 578	4 672	CT1	3 790	3 361	4 155
B3	1 723	1 957	1 617	CT2	9 720	11 110	7 034
B4	4 864	4 937	4 089	CT3	9 173	8 462	5 893
B5	6 142	3 995	4 484				
B6	3 698	2 902	3 709	MS1	4 185	4 186	6 171
B7	3 215	2 435	2 239	MS2	5 615	6 838	5 336
				MS3	392	660	382
T1	15 020	11 808	12 729				
T2	3 771	3 237	2 985	S1	9 236	9 364	10 030
				S2	13 022	12 742	14 421
CT1	3 236	3 131	1 635	S3	15 856	9 306	11 104
CT2	4 805	4 805	4 807				

APPENDIX: PARTIAL CORRELATION COEFFICIENT

Given the variables x and y measured for n samples and their mean values $\overline{x}$ and $\overline{y}$, Pearson's correlation coefficient r_{xy} is given by

$$r_{xy}=\frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2\right)\left(\frac{1}{n}\sum_{i=1}^{n}(y_i-\overline{y})^2\right)}}.$$

It happens that the measured correlation between two variables x and y is due to the underlying variation of a third variable z (Saporta, 1990). In this case, the calculation of a partial correlation coefficient should help to eliminate this underlying variation. The partial correlation coefficient between the variables x and y which takes into account the underlying variation of the variable z is given by Dagnelie (1975) and Jolicoeur (1991)

$$r_{xy\cdot z}=\frac{r_{xy}-r_{xz}r_{yz}}{\sqrt{(1-r_{xz}^2)(1-r_{yz}^2)}}.$$

In our study, fundamental frequency and vocal intensity are strongly correlated, which could bias the correlation calculated between open quotient and these two variables. Thus, the calculation of a partial correlation coefficient helps to factor out the effect of this strong correlation.

The degrees of freedom corresponding to the correlation coefficients are given in Table VIII. In each case, the degree of freedom is very high, being greater than 200.

ACKNOWLEDGMENTS

The authors would like to thank the 18 singers who kindly participated in these experiments. This work is part of a Ph.D. thesis defended at the Université Pierre & Marie Curie (Paris, France) in November 2001. All the figures related to this work and the corresponding sound samples can be found in the Ph.D. thesis manuscript (Henrich, 2001). The authors are deeply grateful to Evelyn Abberton and Alexis

Michaud for their help with the English text, and to Daniel Ambroise for his help with statistical analyses. They also wish to acknowledge the contribution of the three anonymous reviewers and the Associate Editor Stephen McAdams, whose comments were very useful for improving the manuscript.

Alku, P., and Vilkman, E. (**1996**). ''A comparison of glottal voice source quantification parameters in breathy, normal, and pressed phonation of female and male speakers,'' Folia Phoniatr (Basel) **48**, 240–254.

Castellengo, M., Chuberre, B., and Henrich, N. (**2004**). ''Is '*voix mixte*,' the vocal technique used to smoothe the transition across the two main laryngeal mechanisms, an independent mechanism?,'' in International Symposium on Musical Acoustics, Nara, Japan.

Childers, D. G., Hicks, D. M., Moore, G. P., Eskenazi, L., and Lalwani, A. L. (**1990**). ''Electroglottography and vocal fold physiology,'' J. Speech Hear. Res. **33**, 245–254.

Childers, D. G., Naik, J. M., Larar, J. N., Krishnamurthy, A. K., and Moore, G. P. (**1983**). ''Electroglottography, speech and ultra-high speed cinematography,'' in *Vocal Fold Physiology and Biophysics of Voice*, edited by I. Titze and R. Scherer (Denver Center for the Performing Arts, Denver), pp. 202–220.

Chuberre, B. (**2000**). ''Les registres et passages dans la voix chantée (Registers and transitions in singing),'' Mémoire de phoniatrie, Université de Nantes.

Cookman, S., and Verdolini, K. (**1999**). ''Interrelation of mandibular laryngeal functions,'' J. Voice **13**, 11–24.

Dagnelie, P. (**1975**). *Théorie et Méthodes Statistiques (Theory and Methods in Statistics)* (Les Presses Agronomiques de Gembloux, Belgique), Vol. 2.

Daudin, J. J., Robin, S., and Vuillet, C. (**1999**). *Statistique Inférentielle. Idées, Démarches, Exemples. (Inferential Statistics. Ideas, Methods, Examples)* (Pratique de la statistique—Société Franaise de Statistique et Presses Universitaires de Rennes, Rennes).

Dejonckere, P. (**1981**). ''Comparison of two methods of photoglottography in relation to electroglottography,'' Folia Phoniatr. **33**, 338.

Dromey, C., Stathopoulos, E. T., and Sapienza, C. M. (**1992**). ''Glottal airflow and electroglotto-graphic measures of vocal function at multiple intensities,'' J. Voice **6**, 44–54.

Expert, R. (**2003**). ''Les voix d'alto (Alto voices),'' Mémoire de la classe d'acoustique musicale, Conservatoire National Supérieur de Musique et de Danse de Paris.

Garnier, M., Dubois, D., Poitevineau, J., Henrich, N., and Castellengo, M. (**2004**). ''Perception et description verbale de la qualité vocale dans le chant lyrique: Une approche cognitive (voice quality perception and description in western operatic singing),'' in *Journées d'Etude sur la Parole 2004 & Traitement Automatique du Langage Naturel 2004* (Fez, Maroc), pp. 133–136.

Gauffin, J., and Sundberg, J. (**1989**). ''Spectral correlates of glottal voice source waveform characteristics,'' J. Speech Hear. Res. **32**, 556–565.

Gramming, P., Sundberg, J., Ternstrom, S., Leanderson, R., and Perkins, W. (**1988**). ''Relationship between changes in voice pitch and loudness,'' J. Voice **2**, 118–126.

Hanson, H. M. (**1995**). ''Glottal characteristics of female speakers,'' Ph.D. thesis, Harvard University.

Hanson, H. M. (**1997**). ''Glottal characteristics of female speakers: Acoustic correlates,'' J. Acoust. Soc. Am. **101**, 466–481.

Hanson, D. G., Gerratt, B. R., and Berke, G. S. (**1990**). ''Frequency, intensity and target matching effects on photoglottographic measures of open quotient and speed quotient,'' J. Speech Hear. Res. **33**, 45–50.

Henrich, N. (**2001**). ''Etude de la source glottique en voix parlée et chantée: Modélisation et estimation, mesures acoustiques et électroglottographiques, perception (Study of the glottal source in speech and singing: Modeling and estimation, acoustic and electroglottographic measurements, perception),'' Ph.D. thesis, Université Paris 6, France.

Henrich, N., d'Alessandro, C., and Doval, B. (**2001**). ''Spectral correlates of voice open quotient and glottal flow asymmetry: Theory, limits and experimental data,'' in Eurospeech 2001, Aalborg, Denmark.

Henrich, N., Roubeau, B., and Castellengo, M. (**2003**). ''On the use of electroglottography for characterisation of the laryngeal mechanisms,'' in Stockholm Music Acoustics Conference, Stockholm, Sweden.

Henrich, N., d'Alessandro, C., Castellengo, M., and Doval, B. (**2004**). ''On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation,'' J. Acoust. Soc. Am. **115**, 1321–1332.

Hirano, M. (**1982**). ''The role of the layer structure of the vocal fold in register control,'' Vox Humana, University of Jyvaskyla, pp. 50–62.

Hollien, H. (**1974**). ''On vocal registers,'' J. Phonetics **2**, 125–143.

Holmberg, E. B., Hillman, R. E., and Perkell, J. S. (**1988**). ''Glottal air flow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice,'' J. Acoust. Soc. Am. **84**, 511–529.

Holmberg, E. B., Hillman, R. E., and Perkell, J. S. (**1989**). ''Glottal air flow and transglottal air pressure measurements for male and female speakers in low, normal, and high pitch,'' J. Voice **3**, 294–305.

Holmberg, E. B., Hillman, R. E., Perkell, J. S., Guiod, P. C., and Goldman, S. L. (**1995**). ''Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice,'' J. Speech Hear. Res. **38**, 1212–1223.

Howard, D. M. (**1995**). ''Variation of electrolaryngographically derived closed quotient for trained and untrained adult female singers,'' J. Voice **9**, 163–172.

Howard, D. M., Lindsey, G. A., and Allen, B. (**1990**). ''Toward the quantification of vocal efficiency,'' J. Voice **4**, 205–212.

Isshiki, N. (**1964**). ''Regulatory mechanism of voice intensity variation,'' J. Speech Hear. Res. **7**, 17–29.

Jolicoeur, P. (**1991**). *Introduction à la Biométrie (An Introduction to Biometry*) (Decarie/Masson, Montréal/Paris).

Karlsson, I. (**1986**). ''Glottal waveforms for normal female speakers,'' STL-QPSR **1**, 31–36.

Kitzing, P. (**1982**). ''Photo- and electroglottographical recording of the laryngeal vibratory pattern during different registers,'' Folia Phoniatr. **34**, 234–241.

Kitzing, P. (**1983**). ''Simultaneous photo- and electroglottographic measurements of voice strain,'' in *Vocal Fold Physiology*, edited by I. R. Titze and R. C. Scherer (The Denver Center for the Performing Arts, Denver), pp. 221–229.

Kitzing, P., and Sonesson, B. (**1974**). ''A photoglottographical study of the female vocal folds during phonation,'' Folia Phoniatr (Basel) **26**, 138–149.

Kitzing, P., Carlborg, B., and Löfqvist, A. (**1982**). ''Aerodynamic and glottographic studies of the laryngeal vibratory cycle,'' Folia Phoniatr. **34**, 216–224.

Klatt, D., and Klatt, L. (**1990**). ''Analysis, synthesis, and perception of voice quality variations among female and male talkers,'' J. Acoust. Soc. Am. **87**, 820–857.

Ladefoged, P., and McKinney, N. P. (**1963**). ''Loudness, sound pressure and subglottal pressure in speech,'' J. Acoust. Soc. Am. **35**, 454–460.

Lecluse, F. (**1977**). ''Elektroglottografie (Electroglottography),'' Thesis Rotterdam, Drukkerig Elinkwijk, Utrecht.

Lecluse, F., and Brocaar, M. (**1977**). ''Quantitative measurements in the electroglottogram,'' 17th International Congress of Logopedics and Phoniatrics.

Lecuit, V., and Demolin, D. (**1998a**). ''Relation entre pression sous-glottique et intensité: Étude des voyelles du français (Relation between subglottal pressure and intensity: Study of the French vowels),'' in *XXIIèmes Journées d'Etude sur la Parole* (Martigny, Suisse), pp. 299–302.

Lecuit, V., and Demolin, D. (**1998b**). ''The relationship between intensity and subglottal pressure with controlled pitch,'' in ICSLP, Sydney, pp. 3083–3086.

Liénard, J. S., and Di Benedetto, M. G. (**1999**). ''Effect of vocal effort on spectral properties of vowels,'' J. Acoust. Soc. Am. **106**, 411–422.

Miller, D. G. (**2000**). ''Registers in singing,'' Ph.D. thesis, Rijksuniversiteit Groningen, the Netherlands.

Miller, D. G., Svec, J. G., and Schutte, H. K. (**2002**). ''Measurement of characteristic leap interval between chest and falsetto registers,'' J. Voice **16**, 8–19.

Orlikoff, R. F. (**1991**). ''Assessment of the dynamics of vocal fold contact from the electroglottogram: Data from normal male subjects,'' J. Speech Hear. Res. **34**, 1066–1072.

Rothenberg, M. (**1992**). ''A multichannel electroglottograph,'' J. Voice **6**, 36–43.

Roubeau, B. (**1993**). ''Mécanismes vibratoires laryngés et contrôle neuromusculaire de la fréquence fondamentale (Laryngeal vibratory mechanisms and neuro-muscular control of fundamental frequency),'' thesis, Université Paris XI, Orsay, France.

Roubeau, B., Chevrie-Muller, C., and Arabia-Guidet, C. (**1987**). ''Electroglottographic study of the changes of voice registers,'' Folia Phoniatr. **39**, 280–289.

Roubeau, B., Chevrie-Muller, C., and Arabia, C. (**1991**). ''Control of laryngeal vibration in register change,'' in *Vocal Fold Physiology: Acoustic, Perceptual, and Physiological Aspects of Voice Mechanisms*, edited by J. Gauffin and B. Hammarberg (Singular, San Diego), pp. 279–286.
Saporta, G. (**1990**). *Probabilités, Analyse des Données et Statistique (Probabilities, Data Analysis and Statistics*) (Editions Technip, Paris).
Schutte, H. K. (**1980**). ''The efficiency of voice production,'' Ph.D. thesis, Rijksuniversiteit Groningen, the Netherlands.
Sundberg, J., and Högset, C. (**1999**). ''Voice source differences between falsetto and modal registers in counter tenors, tenors and baritones,'' TMH-QPSR **3–4**, 65–74.
Sundberg, J., Titze, I. R., and Scherer, R. C. (**1993**). ''Phonatory control in male singing: A study of the effects of subglottal pressure, fundamental frequency, and mode of phonation on the voice source,'' J. Voice **7**, 15–29.
Sundberg, J., Andersson, M., and Hultqvist, C. (**1999a**). ''Effects of subglottal pressure variation on professional baritone singers' voice sources,'' J. Acoust. Soc. Am. **105**, 1965–1971.
Sundberg, J., Cleveland, T. F., Stone, R. E., and Iwarsson, J. (**1999b**). ''Voice source characteristics in six premier country singers,'' J. Voice **13**, 168–183.
Svec, J. G., Schutte, H. K., and Miller, D. G. (**1999**). ''On pitch jumps between chest and falsetto registers in voice: Data from living and excised human larynges,'' J. Acoust. Soc. Am. **106**, 1523–1531.
Tanaka, S., and Gould, W. J. (**1983**). ''Relationships between vocal intensity and noninvasively obtained aerodynamic parameters in normal subjects,'' J. Acoust. Soc. Am. **73**, 1316–1321.
Timcke, R., von Leden, H., and Moore, P. (**1958**). ''Laryngeal vibrations: Measurements of the glottic wave,'' AMA Arch. Otolaryngol. **68**, 1–19.
Titze, I. R., and Sundberg, J. (**1992**). ''Vocal intensity in speakers and singers,'' J. Acoust. Soc. Am. **91**, 2936–2946.

Estimation of total attenuation and scatterer size from backscattered ultrasound waveforms

Timothy A. Bigelow, Michael L. Oelze, and William D. O'Brien, Jr.[a)]
Bioacoustics Research Laboratory, Department of Electrical and Computer Engineering, University of Illinois, 405 North Mathews, Urbana, Illinois 61801

(Received 7 September 2004; revised 15 December 2004; accepted 22 December 2004)

Quantitative ultrasound techniques using backscattered echoes have had limited success *in vivo* due to the frequency-dependent attenuation along the entire propagation path masking the frequency dependence of the backscatter. Herein, total attenuation and scatterer size are estimated simultaneously by an analysis of the *in vivo* backscattered power spectrum using two approaches. The simulations used to evaluate the two approaches used frequencies between 4 and 11 MHz with an effective scatterer radius of 25 μm. The first approach was based on approximations of the *in vivo* backscattered power spectrum (i.e., assumed Gaussian function), wherein attenuation and size were estimated by assuming each was a Gaussian transformation performed on Gaussian power spectra. The approach had poor accuracy due to the backscattered power spectra not being sufficiently modeled by a Gaussian function. The second approach estimated attenuation and size by fitting a modified reference spectrum to the *in vivo* backscattered power spectrum without any assumptions about the shape of the spectrum. The accuracy of the size estimate was better than 20% for signal-to-noise ratio >6 dB, window lengths greater than 4 mm, and attenuation between 0 and 1 dB/cm-MHz. However, the precision quickly degraded with increasing noise, increasing attenuation, and decreasing window length.
[DOI: 10.1121/1.1858192]

PACS numbers: 43.80.Ev, 43.80.Qf, 43.80.Vj [FD] Pages: 1431–1439

LIST OF SYMBOLS

A	term for form factor written as a power law [i.e., $F_\gamma(f,a_{\text{eff}})\cong e^{-Af^n}$]
a_{eff}	effective radius of scatterer
$a_{\text{eff}\,j}$	estimated effective radius of scatterer found from one set (i.e., 25 averaged rf echoes) of simulated backscatter waveforms
$\overline{a}_{\text{eff}}$	mean value of estimated effective radius from all sets of backscattered waveforms (i.e., $\overline{a}_{\text{eff}}=\Sigma_{\forall j}a_{\text{eff}\,j}/\Sigma_{\forall j}j$)
ASD	average squared difference term minimized by spectral fit algorithm
c	speed of sound in the tissue
c_o	speed of sound in water
$E[\ ]$	expected value with respect to scattering random process
$E_N[\ ]$	expected value with respect to electronic noise random process
f	frequency
$f\#$	f number of a focused transducer
f_o	the frequency corresponding to the spectral peak of the Gaussian spectrum [i.e., $\exp(-([(f-f_o)/\sqrt{2}\sigma_\omega])^2)$]
$\tilde{f}_o$	f_o for backscattered spectrum modified by scatterer size
$\tilde{f}'_o$	f_o for backscattered spectrum modified by scatterer size and attenuation along propagation path
$F_\gamma(\omega,a_{\text{eff}})$	form factor related to the scatterer geometry and size
g_{win}	windowing function used to gate the time-domain waveforms
$H(f)$	one-way passband characteristics of ultrasound source
k	wave number in tissue
k_o	wave number in water
L	length of the windowing function
n	power of frequency when form factor is written as a power law [i.e., $F_\gamma(f,a_{\text{eff}})\cong e^{-Af^n}$]
$N_{\text{space}}(f)$	spectral fluctuations from random scatterer spacing
$N(f)$	spectrum of additive electronic noise
P_{ref}	reference spectrum. [i.e., $P_{\text{ref}}(f)=k_o^4\lvert V_{\text{inc}}(\omega)\rvert^2\lvert H(\omega)\rvert^4$]
P_{scat}	$E[\lvert V_{\text{refl}}\rvert^2]$ estimated from set of waveforms ($\neq E[\lvert V_{\text{refl}}\rvert^2]$ due to the finite number of waveforms available to generate the estimate)
SNR	signal-to-noise ratio
t	time
T_{win}	length windowing function applied to time-domain waveform (i.e., $T_{\text{win}}=2L/c$)
V_{inc}	voltage spectrum applied to ultrasound source
V_j	backscattered voltage spectrum for a single rf echo
v_{noise}	noise signal voltage in time domain (i.e., no signal transmitted by source)
V_{plane}	backscattered voltage spectrum from rigid plane placed at the focal plane [i.e., proportional to $V_{\text{inc}}(\omega)(H(\omega))^2e^{i2k_oz_T}$]

[a)]Electronic mail: wdo@uiuc.edu

V_{refl} backscattered voltage spectrum from tissue containing scatterers

$v_{\text{refl}\,j}$ voltage of a rf echo in time domain

w_x, w_y, w_z equivalent Gaussian dimensions of velocity potential field in focal region

X, $\bar{X}$ terms used in minimization scheme for spectral fit algorithm

z_T distance from aperture plane to focal plane of ultrasound source

α total effective attenuation coefficient for tissue between ultrasound source and the region containing the scatterers (i.e., assumed linear frequency dependence of the form $\alpha = \alpha_o f + \alpha_b$)

α_b intercept of total attenuation coefficient at zero frequency

α_o slope of total attenuation coefficient versus frequency

$(\alpha_o z_T)_j$ estimated attenuation along the propagation path for single data set

$\overline{(\alpha_o z_T)}$ mean value for attenuation along the propagation path from all sets of backscattered waveforms [i.e., $\overline{(\alpha_o z_T)} = \Sigma_{\forall j}(\alpha_o z_T)_j / \Sigma_{\forall j} j$]

$\sigma_{a_{\text{lower}}}$ percent deviation in values of scatterer size for sizes smaller than the mean size (i.e., $a_{\text{eff}\,j} < \overline{a_{\text{eff}}}$)

$\sigma_{a_{\text{upper}}}$ percent deviation in values of scatterer size for sizes larger than the mean size (i.e., $a_{\text{eff}\,j} > \overline{a_{\text{eff}}}$)

$\sigma_{\alpha_{\text{lower}}}$ deviation in dB/MHz in values of attenuation for attenuation values smaller than the mean attenuation [i.e., $(\alpha_o z_T)_j < \overline{(\alpha_o z_T)}$]

$\sigma_{\alpha_{\text{upper}}}$ deviation in dB/MHz in values of attenuation for attenuation values greater than the mean attenuation [i.e., $(\alpha_o z_T)_j > \overline{(\alpha_o z_T)}$]

σ_ω the bandwidth term for Gaussian function [i.e., $\exp(-([(f-f_o)/\sqrt{2}\sigma_\omega])^2)$]

$\tilde{\sigma}_\omega$ σ_ω for backscattered spectrum modified by scatterer size

ω, λ radian frequency, wavelength

I. INTRODUCTION

For years many investigators have attempted to quantify the ultrasound backscatter from tissue by analyzing the power spectra of the rf echoes to estimate the characteristic size of the tissue microstructure (Lizzi *et al.*, 1983; Nassiri and Hill, 1986; Insana *et al.*, 1990; Hall *et al.*, 1996; Oelze *et al.*, 2004). The hypothesis is that the size of the tissue microstructure (i.e., scattering sources) could be used to assess the pathology of the imaged region. For example, researchers have verified that the backscattered power spectrum can be used to distinguish between fibroadenomas and carcinomas (Oelze *et al.*, 2004), as well as to monitor tissue necrosis in ultrasound therapy applications (Lizzi *et al.*, 1997). In light of these successes, it is suggested that the quantitative ultrasound (QUS) estimate of the characteristic size of the tissue microstructure could assist both diagnosis and treatment of pathological tissue.

Unfortunately, the scatterer size estimate has had limited success in clinical practice. The inability to apply laboratory measurements to a clinical setting is due primarily to a false assumption in the traditional estimation routine. Namely, it has been assumed that the total attenuation along the entire propagation path is the same for every patient and is known *a priori* (Oelze and O'Brien, 2002a). If known, the *in vivo* backscattered power spectrum can be compensated for the frequency-dependent attenuation, and the scatterer size can be accurately estimated assuming the scattering process is accurately modeled. However, the attenuation coefficient and thickness of different intervening tissue layers (i.e., fat, muscle, etc.) vary between patients. Lu *et al.* (1999) and Hall *et al.* (1996) attempted to improve the *in vivo* estimate by measuring the thickness of the different intervening tissue layers from the backscattered signals and assigning typical attenuation values to each layer. However, they did not account for the natural variability in attenuation for tissues of the same type. For example, Goss *et al.* (1978) report many different attenuation coefficients of fat taken at different frequencies. In particular, a study by Dussik and Fritch (1956) gives attenuation values of 0.6 ± 0.2 dB/cm at 1 MHz, 1.6 ± 0.2 dB/cm at 3 MHz, and 2.3 ± 0.7 dB/cm at 5 MHz. Hence, the slope of the attenuation versus frequency would vary from 0.2 to 0.65 dB/cm-MHz with a mean value of 0.43 dB/cm-MHz for that study alone. Lu *et al.* (1999) used a value of 0.6 dB/cm-MHz for the attenuation coefficient of fat while referring to Goss *et al.* (1978). However, it is not clear how this number was obtained from the data presented in Goss *et al.* (1978) because the attenuation slope was not measured directly over the diagnostic range in any of the studies presented. Clearly, there is significant variability in the attenuation slope. As a result, it is not feasible to determine accurately the characteristic size of the tissue microstructure of embedded tissues using the Lu *et al.* (1999) or Hall *et al.* (1996) technique.

The best successes clinically for scatterer size estimates were with tumors of the eye where the attenuation along the propagation path is negligible (Lizzi *et al.* 1983; 1997). Other investigators used scatterer size estimates to measure structural changes in the kidney (Insana *et al.*, 1995), where the exteriorized kidney reduced the impact of the attenuation along the propagation path.

Therefore, the frequency-dependent total attenuation along the propagation path must be determined on a patient-specific basis before the full potential of the estimation of scatterer size can be realized, as was recognized over 20 years ago by Lizzi *et al.* (1983). In addition to estimating the size of the tissue microstructure of embedded tumors and other pathologies, a knowledge of the attenuation along the propagation path could improve other medical ultrasound applications including acoustic radiation force impulse imaging (Nightingale *et al.*, 2000) and patient-specific dosimetry (Siddiqi *et al.*, 1992).

Due to the need and significant benefit of an accurate estimate of the total attenuation along the propagation path when quantifying tissue microstructure using ultrasound, as well as other applications, two different algorithms are explored herein. The two algorithms provide an estimate of total attenuation and scatterer size simultaneously from the

in vivo backscattered power spectrum. The algorithms represent two approaches. The first algorithm, termed the Gaussian transformation (GT) algorithm, was based on an approximation regarding the shape of the *in vivo* backscattered power spectrum and determined the scatterer size and total attenuation from two independent properties of the backscattered spectrum (i.e., center frequency and bandwidth). The second algorithm, termed the spectral fit (SF) algorithm, made no approximations about the shape of the *in vivo* backscattered power spectrum and solved for the size and attenuation by performing a fit over the entire frequency spectrum, similar to the established method for determining just the scatterer size when the attenuation is known (Insana *et al.*, 1990). Both algorithms are first described and then evaluated using computer simulations where the scatterers have Gaussian impedance distributions (i.e., Gaussian scatterers, a form factor). However, the SF algorithm, and possibly the GT algorithm, can be implemented for scatterers with any form factor (i.e., spherical shell, fluid-filled sphere, etc.) provided that the form factor is known *a priori* just like the established method for determining only the scatterer size when the attenuation is known (Insana *et al.*, 1990). In addition to the algorithms, a method to compensate for electronic system noise is developed. After presenting and evaluating each of the algorithms, some conclusions are drawn based on the results.

II. GAUSSIAN TRANSFORMATION ALGORITHM

The goal of the GT algorithm was to estimate the scatterer size and total attenuation from the *in vivo* backscattered power spectrum. The algorithm assumes that the backscattered signal has a Gaussian distributed spectrum where the scatterer size and total attenuation are independent Gaussian transformations. The analysis is similar to that presented by Wear (2002), who used Gaussian transformations to predict the *in vivo* backscattered power spectrum for given scatterers and attenuation.

A. Overview of algorithm

The expected backscattered voltage, $E[|V_{\text{refl}}|^2]$, from a tissue region where the window length is small compared to the depth of focus (i.e., local plane-wave approximation), and the attenuation in the focal region does not significantly alter the spectrum over the length of the window given by (Bigelow and O'Brien, 2004)

$$E[|V_{\text{refl}}|^2] \propto k_o^4 |V_{\text{plane}}(\omega)|^2 e^{-4\alpha z_T} F_\gamma(\omega, a_{\text{eff}}). \tag{1}$$

Equation (1) can be further simplified by assuming that the total attenuation along the propagation path has a linear dependence on frequency (i.e., $\alpha = \alpha_o f + \alpha_b$), a reasonable assumption over the bandwidth of most ultrasound sources (Jongen *et al.*, 1986). Then, the frequency-independent component of the attenuation, α_b, only influences the magnitude of the backscattered spectrum and hence can be absorbed by the proportionality, and the attenuation has a strictly linear dependence on frequency.

For Gaussian scatterers, as well as possibly for other form factors over a limited frequency range, $F_\gamma(\omega, a_{\text{eff}})$ can be reasonably approximated as e^{-Af^n}. Therefore, if the source spectrum and diffraction characteristics can be assumed Gaussian (i.e., $k_o^4 |V_{\text{plane}}(\omega)|^2 \propto \exp(-[(f - f_o)^2/2\sigma_\omega^2])$), then the backscattered voltage is given by

$$E[|V_{\text{refl}}|^2] \propto e^{[(f-f_o)^2/2\sigma_\omega^2]} e^{-Af^n} e^{-4z_T\alpha_o f}, \tag{2}$$

or a series of Gaussian transformations operating on a Gaussian function. Equation (2) can be simplifed by writing Af^n as a Taylor series about f_o, which is

$$Af^n = A\sum_{l=0}^{\infty} \frac{n! f_o^{n-l}}{(n-l)!\, l!} (f - f_o)^l. \tag{3}$$

Also, if $2 \leqslant n < 3$, then the terms in the Taylor series corresponding to $l \geqslant 3$ can be ignored due to the negative power on f_o. For Gaussian scatterers (i.e., $n = 2$), the $l \geqslant 3$ terms are already zero so there is no error introduced by making this assumption for these scatterers. After eliminating the $l \geqslant 3$ terms, Eq. (2) becomes

$$E[|V_{\text{refl}}|^2] \propto \exp\left(-\frac{(f - \tilde{f}_o')^2}{2\tilde{\sigma}_\omega^2} + \cdots \right), \tag{4}$$

where

$$\begin{aligned} \tilde{f}_o &= f_o - \tilde{\sigma}_\omega^2 A n f_o^{n-1}, \\ \tilde{\sigma}_\omega^2 &= \left[\frac{1}{\sigma_\omega^2} + An(n-1) f_o^{n-2} \right]^{-1}, \\ \tilde{f}_o' &= \tilde{f}_o - 4 z_T \alpha_o \tilde{\sigma}_\omega^2. \end{aligned} \tag{5}$$

Based on Eq. (4), the bandwidth of the backscattered spectrum is only affected by the scatterer size. Hence, the scatterer size can be determined independently of the attenuation by measuring changes in the bandwidth of the backscattered signals. Once the scatterer size has been determined, the backscattered spectrum can be multiplied by e^{+Af^n} to remove the effect of scatterer size. Then, the total attenuation can be determined from the downshift in the center frequency of the returned spectrum. Hence, the scatterer size and total attenuation can theoretically be determined from two different properties of the backscattered spectrum.

B. Simulation results

The performance of the GT algorithm was evaluated by simulations that were designed to test the algorithm's sensitivity as a function of attenuation. In all of the simulations, the ultrasound source was a spherically focused $f/4$ transducer with a focal length of 5 cm placed adjacent to a homogeneous semi-infinite half-space containing scatterers. The velocity potential field near the focus was assumed to follow a three-dimensional Gaussian function, an approximation shown to be valid for focused sources (Bigelow and O'Brien, 2004), with the equivalent Gaussian dimensions given by

$$\begin{aligned} w_x &= w_y = 0.87\lambda f\#, \\ w_z &= 6.01\lambda (f\#)^2. \end{aligned} \tag{6}$$

The source was excited by an impulse [i.e., $V_{\text{inc}}(f) = 1$], and had one-way passband characteristics given by (Bigelow and O'Brien, 2004)

$$H(f)=\frac{|f|\exp\left(-\left(\frac{f-8\ \text{MHz}}{6\ \text{MHz}}\right)^2\right)}{\max_{\forall f}\left(|f|\exp\left(-\left(\frac{f-8\ \text{MHz}}{6\ \text{MHz}}\right)^2\right)\right)}, \tag{7}$$

comparable to that measured for a PZT transducer that also goes to zero at zero frequency. The voltage spectrum returned from a plane placed at the focal plane would be given by

$$V_{\text{plane}}(\omega)\propto V_{\text{inc}}(\omega)(H(\omega))^2 e^{i2k_o z_T}. \tag{8}$$

The simulated source spectra and diffraction characteristics [i.e., $k_o^4|V_{\text{plane}}(\omega)|^2$] were not perfectly Gaussian, but instead went to zero at zero frequency, just like a real source. This was done so that reasonable deviations from the ideal theory could be assessed.

The backscattered voltage for the simulations was generated by solving for the scattered field from a single scatterer with a Gaussian impedance distribution [i.e., Gaussian scatterer with $F_\gamma(f,a_{\text{eff}})=\exp(-0.827(ka_{\text{eff}})^2)$] at an arbitrary location in the focal region. A Gaussian scatterer was selected because the microstructure of real tissue is often modeled using this form factor (Chaturvedi and Insana, 1996). The backscatter for many scatterers in the half-space was then obtained by adding together the backscatter from many different randomly positioned scatterers. In the simulations, the scatterers were positioned according to a uniform probability distribution throughout the focal region at a density of 35/mm^3 ($\sim$4.8 scatterers per resolution cell) and each scatterer had an effective radius, a_{eff}, of 25 μm. The sound speed for the half-space was 1532 m/s, a measured sound speed for liver (characteristic of tissue) (Goss *et al.*, 1978, 1980). The attenuation was uniform throughout the half-space and was varied between 0 and 1 dB/cm-MHz in order to test the algorithm's sensitivity to attenuation. The reflection off of the rigid plane at the focus was also simulated using a sound speed of 1540 m/s for the water path. The code used a ''digitizing'' sampling rate of 53 MHz with no added electronic noise.

For each value of attenuation, the backscattered voltage from 1000 independent random scatterer distributions was generated. The waveforms were then separated into 40 independent sets with 25 waveforms per set and windowed in the time domain using a Hamming gating function centered at the focus, thus yielding 40 independent estimates of the effective scatterer radius and the total attenuation. The use of 25 rf echoes is arbitrary, and increasing the number of waveforms may improve the accuracy and precision of the resulting estimates. The width of the Hamming window was varied from 1 to 8 mm in steps of 0.25 mm (i.e., 0.057 to 0.45 by 0.014 when normalized with respect to depth of focus at the center frequency of 9.83 MHz) with corresponding time gates determined from $T_{\text{win}}=2L/c$. The same length Hamming window relative to the wavelength (i.e., $T_{\text{win}}=2L/c_o$) was also used to window the signal returned from the rigid plane placed at the focus when obtaining the reference spectrum. No correction was made for the resulting convolution that would distort the estimates at small window length. The

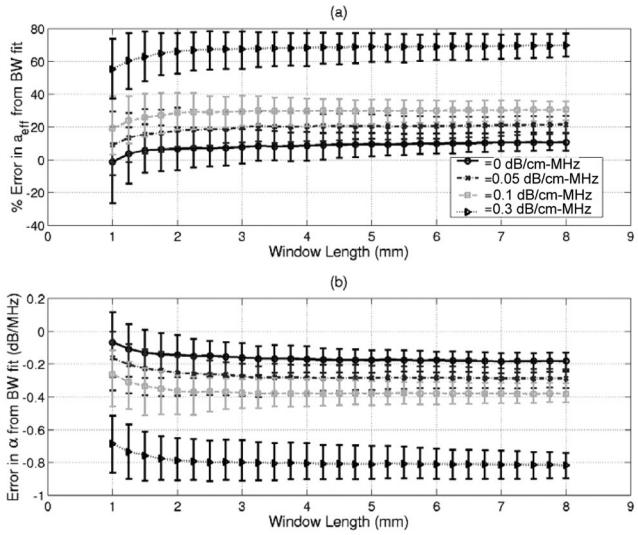


FIG. 1. (a) Percent error in scatterer radius and (b) error in total attenuation for semi-infinite half-spaces with attenuation values of 0, 0.05, 0.1, and 0.3 dB/cm-MHz using the Gaussian transformation algorithm.

resulting estimates from the GT algorithm for attenuation values of 0, 0.05, 0.1, and 0.3 dB/cm-MHz are shown in Fig. 1. The error bars in each case correspond to

$$\sigma_{a_{\text{apper}}}=\frac{100}{a_{\text{eff}|\text{Theory}}}\sqrt{\frac{\Sigma_{\forall a_{\text{eff}\,j}>\overline{a}_{\text{eff}}}(a_{\text{eff}\,j}-\overline{a}_{\text{eff}})^2}{\Sigma_{\forall a_{\text{eff}\,j}>\overline{a}_{\text{eff}}}j}},$$

$$\sigma_{a_{\text{lower}}}=\frac{100}{a_{\text{eff}|\text{Theory}}}\sqrt{\frac{\Sigma_{\forall a_{\text{eff}\,j}<\overline{a}_{\text{eff}}}(a_{\text{eff}\,j}-\overline{a}_{\text{eff}})^2}{\Sigma_{\forall a_{\text{eff}\,j}<\overline{a}_{\text{eff}}}j}}, \tag{9}$$

for the plots of error in a_{eff} and

$$\sigma_{\alpha_{\text{apper}}}=\sqrt{\frac{\Sigma_{\forall(\alpha_o z_T)_j>\overline{(\alpha_o z_T)}}((\alpha_o z_T)_j-\overline{(\alpha_o z_T)})^2}{\Sigma_{\forall(\alpha_o z_T)_j>\overline{(\alpha_o z_T)}}j}},$$

$$\sigma_{\alpha_{\text{lower}}}=\sqrt{\frac{\Sigma_{\forall(\alpha_o z_T)_j<\overline{(\alpha_o z_T)}}((\alpha_o z_T)_j-\overline{(\alpha_o z_T)})^2}{\Sigma_{\forall(\alpha_o z_T)_j<\overline{(\alpha_o z_T)}}j}}, \tag{10}$$

for the plots of error in total attenuation.

For small attenuation values, the scatterer size was estimated with reasonable accuracy (i.e., errors of $\sim$20%). However, as the attenuation was increased, the errors in the scatterer size estimate quickly grew and were $\sim$70% for an attenuation value of 0.3 dB/cm-MHz. The errors were even larger for the higher attenuation values (not shown in this plot). This increase in error with increasing attenuation resulted from a breakdown in the assumption that the source and diffraction characteristics [i.e., $k_o^4|V_{\text{plane}}(\omega)|^2$] followed a Gaussian function. For a Gaussian spectrum, the only change in the bandwidth of the Gaussian spectrum would be due to the size of the scatterer. However, the spectrum of a real transducer as well as our simulated spectrum is not a Gaussian function, going to zero at zero frequency. Hence, as the attenuation was increased and the scattered spectrum shifted to a lower frequency limit, the bandwidth of the Gaussian spectrum was also narrowed due to the attenuation

as the spectrum approached this lower limit. The algorithm did not include this possible change in the spectrum and thus failed.

One modification to the algorithm that was also considered to correct for its failure involved replacing the Gaussian spectrum approximation with a modified Rayleigh spectrum approximation that at least would go to zero at zero frequency (i.e., matching the theoretical analysis to the generated spectrum). The derivation leading to Eq. (4) would remain the same except for a multiplication of $|f|^4$ before the exponential. Then, the change in bandwidth would be found by fitting a modified Rayleigh distribution to the appropriate measured spectrum and solving for the scatterer size appropriately. Unfortunately, fitting Rayleigh functions to the measured spectrum was not a robust operation. Hence, changes in the attenuation still resulted in inappropriate changes in the bandwidth, and the algorithm still failed.

III. SPECTRAL FIT ALGORITHM

The goal of the SF algorithm was the same as that for the GT algorithm, that is, estimate the scatterer size and total attenuation from the *in vivo* backscattered power spectrum. Unlike the GT algorithm approach for which the *in vivo* backscattered power spectrum was assumed to be Gaussian distributed, there was no such assumption with the SF algorithm. Instead, a single minimization routine was used to fit a curve to the *in vivo* backscattered power spectrum, hence its name. The analysis is similar to the traditional method by Insana *et al.* (1990), who fitted the spectrum of the backscattered rf echoes from the region of interest (ROI) to an appropriate reference spectrum. And, unlike the approach by He and Greenleaf (1986), who used the entire backscattered rf echo where tissue heterogeneities along the propagation path would introduce errors, only the echoes from the ROI were needed to obtain the estimates of scatterer size and attenuation. Hence, tissue heterogeneity along the propagation path would not affect the estimates.

A. Overview of algorithm

The first step in the implementation of the SF algorithm is to estimate the theoretical value of $E[|V_{\text{refl}}|^2]$ from measured backscattererd rf echoes. In this work, $E[|V_{\text{refl}}|^2]$ was estimated by averaging the power spectra from 25 rf echoes in the log domain, that is

$$E[|V_{\text{refl}}(f)|^2] \approx P_{\text{scat}}(f) = \exp\left(\frac{1}{25} \sum_{j=1}^{25} \ln(|V_j|^2) \right). \quad (11)$$

The averaging removes the spectral fluctuation that results from the random scatterer spacing that masks the true value of $E[|V_{\text{refl}}|^2]$. The averaging was done in the log domain because it was assumed that the transmitted pulse was convolved with the random medium to generate the reflected spectrum. Hence, the spectral fluctuation from the random scatterer spacing should be included as a multiplicative random impulse train in the frequency domain [i.e., $|V_{\text{refl}}(f)|^2 = E[|V_{\text{refl}}|^2]|N_{\text{space}}(f)|^2$]. Multiplicative random variations can be removed by taking the log of the measured signals, so that the noise becomes additive, and then averaging.

In addition to estimating $E[|V_{\text{refl}}|^2]$, the SF algorithm requires a reference spectrum from a known reflector. In this work, the reference spectrum was obtained using the echo signal from a plane positioned at the focus. The echo from the plane was windowed by the same Hamming window used to window the simulated backscattered signal, and multiplied by k_o^4, yielding

$$P_{\text{ref}}(f) = k_o^4 |V_{\text{inc}}(\omega)|^2 |H(\omega)|^4 \propto k_o^4 |V_{\text{plane}}(\omega)|^2. \quad (12)$$

Once P_{scat} and P_{ref} had been determined, the scatterer size, a_{eff}, and total attenuation along the propagation path, α_o, were then found by estimating the values of a_{eff} and α_o that minimized the error given by

$$\text{ASD} = \underset{f}{\text{mean}}[(X(f, a_{\text{eff}}, \alpha_o) - \overline{X}(a_{\text{eff}}, \alpha_o))^2], \quad (13)$$

where

$$X(f, a_{\text{eff}}, \alpha_o) = \ln\left(\frac{P_{\text{scat}}(f)}{\max_f(P_{\text{scat}}(f))} \right) - \ln\left(\frac{P_{\text{ref}}(f) F_\gamma(f, a_{\text{eff}}) e^{-4\alpha_o f z_T}}{\max_f(P_{\text{ref}}(f) F_\gamma(f, a_{\text{eff}}) e^{-4\alpha_o f z_T})} \right),$$
$$\overline{X}(a_{\text{eff}}, \alpha_o) = \underset{f}{\text{mean}}[X(f, a_{\text{eff}}, \alpha_o)]. \quad (14)$$

Hence, the algorithm involved directly fitting a modified reference spectrum to the measured backscattered spectrum from the tissue region (i.e., spectral fitting). The fitting procedure is very similar to that described by Insana *et al.* (1990), when the attenuation is known. However, as implemented herein, two parameters are estimated instead of only a_{eff}. Fortunately, Eq. (13) has one minimum, as was confirmed by plotting the minimization surface for many different cases, so a simple minimization routine can be utilized.

B. Simulation results without electronic noise

The performance of the SF algorithm was also evaluated by simulations that were designed to test its performance as a function of attenuation. The same simulated scattered signals used to test the GT algorithm were used to test the SF algorithm. Hence, 40 independent estimates of scatterer size and attenuation were found per case for cases with homogeneous half-spaces with attenuation values between 0 and 1 dB/cm-MHz. When fitting the spectra [Eq. (13)], the SF algorithm used all spectral values greater than -30 dB relative to the value at the spectral peak. The choice of -30 dB for the limiting bandwidth is somewhat arbitrary; however, smaller bandwidths would reduce the amount of data available when performing the fit. The results for an attenuation of 0.3 dB/cm-MHz for the Hamming window lengths in the absence of electronic noise are shown in Fig. 2. The error bars are the same as those defined by Eq. (9) and Eq. (10). For the larger window lengths, the algorithm provided a reasonable estimate for scatterer size and total attenuation. However, as the window length was decreased, both the accuracy and precision degraded. The loss of accuracy was due to the convolution effects of the windowing function (Akita and Ueda, 1988). The loss of precision at smaller window

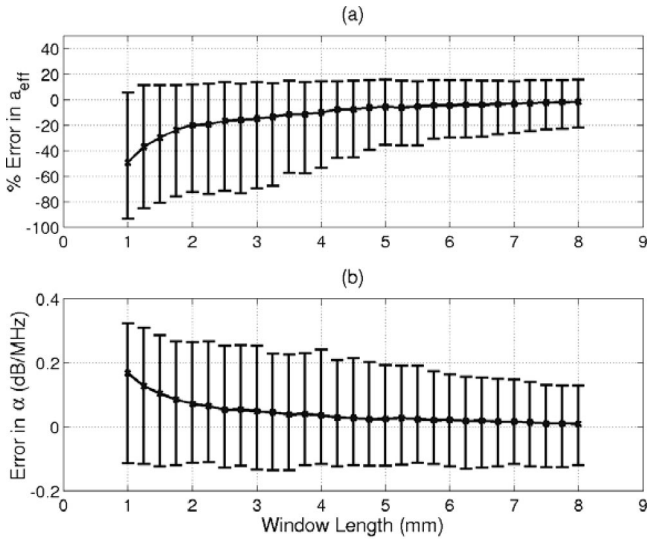


FIG. 2. Errors in estimates of (a) scatterer size and (b) total attenuation for different Hamming window lengths for an attenuation of 0.3 dB/cm-MHz in the absence of any electronic noise using the spectral fit algorithm.

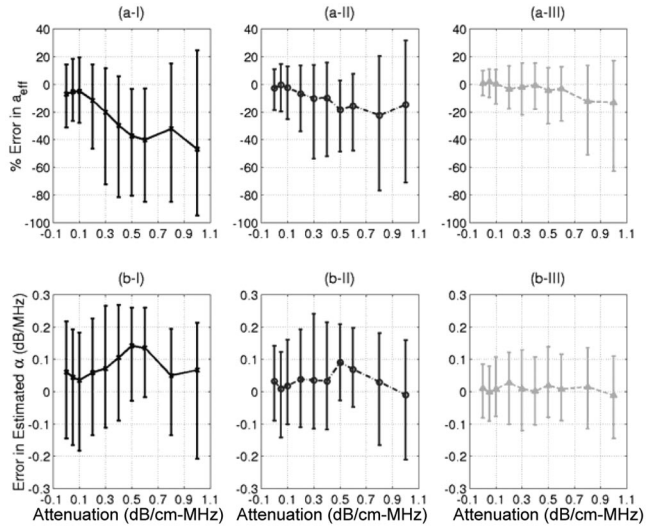


FIG. 3. Errors in estimates of (a) scatterer size and (b) total attenuation for different values of attenuation with Hamming window lengths of (I) 2 mm; (II) 4 mm; and (III) 8 mm in the absence of any electronic noise using the spectral fit algorithm.

lengths was probably due to a reduction in the number of independent samples in the least-squares fit [Eq. (13)] due to a loss in the frequency resolution of the spectrum (Wear 2001). The degradation with window length was also observed for other attenuation values (Fig. 3). The precision of the attenuation estimate, as well as the accuracy and precision of the scatterer size estimate, degraded with increasing attenuation. However, the average of all 40 estimates (i.e., accuracy) had errors smaller than 20% for the estimate of scatterer size and errors less than 0.2 dB/MHz for the attenuation estimate for all attenuation values and window lengths greater than 4 mm. An error of 0.2 dB/MHz translates to an error of about 20% when estimating the pressure amplitude at 8 MHz.

C. Electronic noise compensation

Although the accuracy results (Figs. 2 and 3) appear promising, the SF algorithm still needs to be evaluated in the presence of electronic system noise because real measurement systems will have noise. The effect of noise is twofold. First, the noise will reduce the usable bandwidth of the backscattered spectrum. The selection of the usable frequency range in the presence of noise for our work will be discussed in the next section. The second impact of electronic noise is a slight broadening of the backscattered spectrum. This broadening must be compensated before accurate estimates can be obtained from the noisy waveform.

In order to understand the broadening, assume that the electronic noise is additive, white, and with zero mean. Then, the expected spectrum over the set of possible additive noise for a given scatterer distribution is given by

$$E_N[|V_{\text{refl}}(f)+N(f)|^2]=|V_{\text{refl}}(f)|^2E_N\left[\left|1+\frac{N(f)}{V_{\text{refl}}(f)}\right|^2\right]$$
$$=|V_{\text{refl}}(f)|^2\left(1+\frac{E_N[|N(f)|^2]}{|V_{\text{refl}}(f)|^2}\right). \tag{15}$$

The values near the peak of $|V_{\text{refl}}|^2$ are not affected by the noise while the $|V_{\text{refl}}|^2$ values near the edge of the usable spectrum are increased resulting in a broadening of the spectrum.

In order to compensate for the spectral broadening from the noise, consider $P_{\text{scat}}(f)$ obtained from averaging the spectra in the log domain in the presence of noise given by

$$P_{\text{scat}}(f)=\exp\left(\frac{1}{25}\sum_{i=1}^{25}\ln\left(|V_i(f)|^2\left(1+\frac{E_N[|N(f)|^2]}{|V_i(f)|^2}\right)\right)\right)=P_{\text{scat}}(f)|_{\text{ideal}}\exp\left(\frac{1}{25}\sum_{i=1}^{25}\ln\left(1+\frac{E_N[|N(f)|^2]}{|V_i(f)|^2}\right)\right)$$
$$\cong P_{\text{scat}}(f)|_{\text{ideal}}\exp\left(\ln\left(1+\frac{1}{25}\sum_{i=1}^{25}\frac{E_N[|N(f)|^2]}{|V_i(f)|^2}\right)\right)\cong P_{\text{scat}}(f)|_{\text{ideal}}\exp\left(\ln\left(1+\frac{E_N[|N(f)|^2]}{\frac{1}{25}\sum_{i=1}^{25}|V_i(f)|^2}\right)\right)$$
$$=P_{\text{scat}}(f)|_{\text{ideal}}\left(1+\frac{E_N[|N(f)|^2]}{\frac{1}{25}\sum_{i=1}^{25}|V_i(f)|^2}\right)\cong P_{\text{scat}}(f)|_{\text{ideal}}\left(1+\frac{E_N[|N(f)|^2]}{P_{\text{scat}}(f)|_{\text{ideal}}}\right), \tag{16}$$

where $P_{\text{scat}}(f)|_{\text{ideal}}$ equals $P_{\text{scat}}(f)$ in the absence of noise. $E_N[|N(f)|^2]$ can be determined for the experimental system by recording the noise in the absence of a transmitted signal and then taking the mean value of $|N(f)|^2$. The effect of electronic noise can then be reduced by dividing $P_{\text{scat}}(f)$ by $(1+E_N[|N(f)|^2]/P_{\text{scat}}(f))$ before solving for the scatterer size and attenuation assuming that $P_{\text{scat}}(f)|_{\text{ideal}} \cong P_{\text{scat}}(f)$. Compensating for the electronic noise using this term is similar to the SNR weighting described previously by Oelze and O'Brien (2002b) only without their accompanying loss of bandwidth. Comparisons of the accuracy of the estimates (i.e., average of all 40 estimates) both with and without the electronic noise compensation are provided in Fig. 4 for three SNR values (9, 23, and 36 dB) calculated from

$$\text{SNR} = \frac{1}{25}\sum_{j=1}^{25}\left(10\cdot\log\left(\frac{\int(g_{\text{win}}(t)v_{\text{refl}_j}(t))^2dt}{\int(g_{\text{win}}(t)v_{\text{noise}}(t))^2dt}\right)\right), \quad (17)$$

where v_{refl_j} are from each group of 25 rf echoes used in the estimate before the noise was added, and v_{noise} is the noise signal used to obtain the estimate for $E_N[|N(f)|^2]$. The SNR values in the legend are the average values from all 40 data sets [Eq. (17)]. Accurate estimates of both scatterer size (~10%) and attenuation (~0.15 dB/MHz) were obtained using the noise compensation for all three levels of electronic noise, while the estimates without the noise compensation quickly degraded with increasing noise. Even SNRs as high as 23 dB had errors of 20% in the scatterer size estimate in the absence of noise compensation.

D. Simulation results with electronic noise

The effect of electronic noise on the SF algorithm was evaluated by adding white Gaussian noise of varying levels to the simulated rf echoes in the time domain. The power of the random noise signal was set relative to the reference signal from the rigid plane placed at the focal plane. Also, one of the randomly generated noise signals in the time domain for each noise level was used to calculate $E_N[|N(f)|^2]$ for the noise compensation. In order to not bias the estimates by using frequencies dominated by noise, the fit used all frequencies corresponding to signal values satisfying

$$\left\{10\log\left(\frac{P_{\text{scat}}(f)}{\max_{\forall f}(P_{\text{scat}}(f))}\right)\right\}_{\substack{\text{polynomial}\\ \text{fit}}} > \max\left(\left[-30\ \ \text{mean}\left(\left\{10\log\left(\frac{P_{\text{scat}}(f)}{\max_{\forall f}(P_{\text{scat}}(f))}\right)\right\}_{\substack{\text{polynomial}\\ \text{fit}}}\right)\right]\right), \quad (18)$$

where the spectra were fit by a polynomial of large degree to reduce the impact of spectral variations on the selected frequency range. The backscattered signal was significantly oversampled (i.e., sampling frequency of 53 MHz was ~4 times the Nyquist frequency), so that the mean of the spectrum would yield a value slightly larger than the level of the white zero-mean random noise. Therefore, Eq. (18) only selected signal values greater than the noise floor of the system. Although the degree of the polynomial is arbitrary, a value of 50 was used. Other values considered did not have a significant impact on the final results.

For each noise level and for half-space attenuation values between 0.05 and 1 dB/cm-MHz, the scatterer size and total attenuation were estimated for Hamming window lengths from 1 to 8 mm. The simulation results with half-space attenuation values of 0.3, 0.6, and 1 dB/cm-MHz are shown in Figs. 5, 6, and 7, respectively. The vertical error

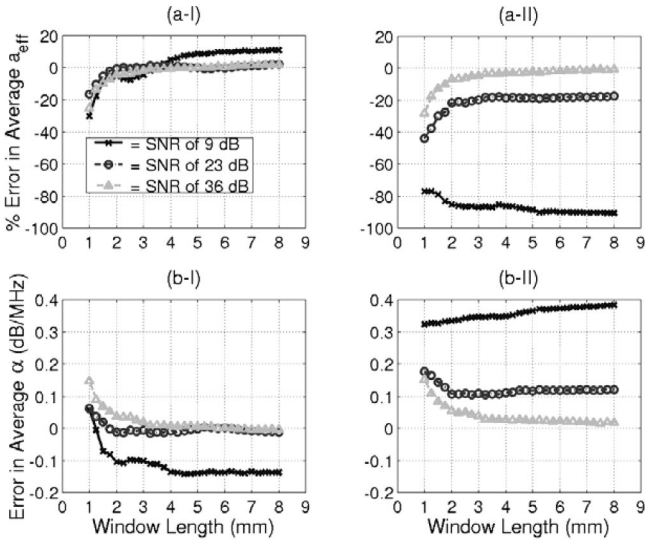


FIG. 4. Errors in estimates of (a) scatterer size and (b) total attenuation both (I) with and (II) without noise compensation for a half-space with an attenuation of 0.05 dB/cm-MHz using the spectral fit algorithm.

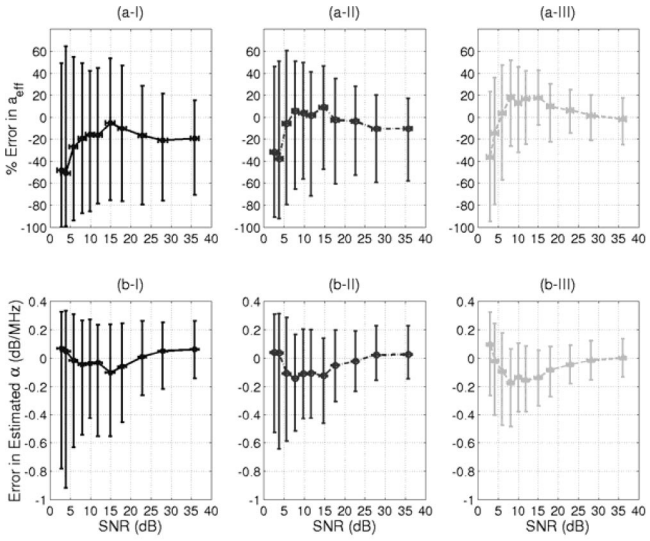


FIG. 5. Errors in estimates of (a) scatterer size and (b) total attenuation for different levels of electronic noise for a half-space attenuation of 0.3 dB/cm-MHz and Hamming window lengths of (I) 2 mm; (II) 4 mm; and (III) 8 mm using the spectral fit algorithm.

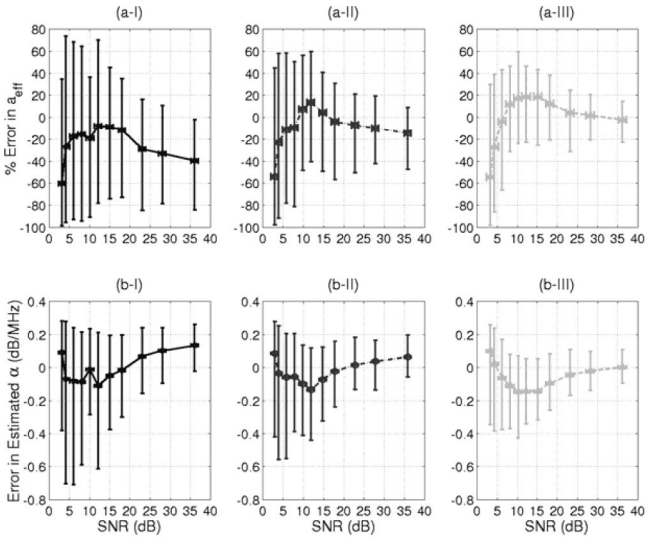


FIG. 6. Errors in estimates of (a) scatterer size and (b) total attenuation for different levels of electronic noise for a half-space attenuation of 0.6 dB/cm-MHz and Hamming window lengths of (I) 2 mm; (II) 4 mm; and (III) 8 mm using the spectral fit algorithm.

bars are the same as those given in Eq. (9) and Eq. (10), and the horizontal error bars correspond to plus and minus 1 standard deviation of the calculated SNR values from Eq. (17) from each of the 40 estimates.

The SF algorithm gives reasonably accurate average estimates of the scatterer size with errors less than 20% for SNRs between 6 and 36 dB for half-space attenuation values between 0.05 and 1 dB/cm-MHz and window lengths greater than 4 mm. Likewise, the average estimate for the total attenuation had errors less than 0.2 dB/MHz for SNRs between 3 and 36 dB. Unfortunately, the algorithm is not very precise as can be seen from the large error bars about the average estimates. As a result, consistently precise estimates (i.e., error bars $<$20% for size estimates and $<$0.2 dB/MHz for attenuation estimates) can only be obtained for SNRs better than 23 dB for half-space attenuation values between 0.05 and 0.3 dB/cm-MHz and better than 28 dB for half-space attenuation values between 0.4 and 0.6 dB/cm-MHz for the window length of 8 mm. The higher values of half-space attenuation never meet this precision limit. Furthermore, the precision of the smaller window lengths is worse.

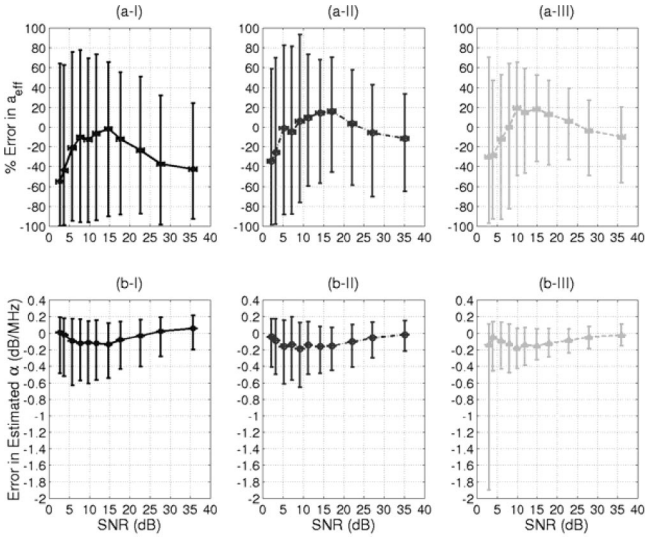


FIG. 7. Errors in estimates of (a) scatterer size and (b) total attenuation for different levels of electronic noise for a half-space attenuation of 1 dB/cm-MHz and Hamming window lengths of (I) 2 mm; (II) 4 mm; and (III) 8 mm using the spectral fit algorithm.

IV. DISCUSSION

The performance of two algorithms for estimating the scatterer size and total attenuation simultaneously from the same set of backscattered waveforms has been assessed. The first was the Gaussian transformation algorithm, which treated the scatterer size and attenuation as two different Gaussian transformations operating on the Gaussian-distributed backscattered spectrum. Although the algorithm's performance was reasonable for very small attenuation values, the accuracy of the scatterer size estimate quickly degraded with increasing attenuation. The failure was a result of the algorithm assuming that the bandwidth of the spectrum was only affected by the scatterer size, an approximation that would be true for a purely Gaussian spectrum. However, the spectrum of a real transducer is also narrowed due to the attenuation as the spectrum is shifted towards zero frequency. It is thus concluded that it is unlikely that any algorithm that makes approximations regarding the shape of the spectra will yield accurate results when estimating both the scatterer size and total attenuation from the *in vivo* backscattered power spectrum, even though the approximations are valid in other applications (Wear, 2002). Algorithms that make an assumption about the shape of the spectrum restrict its applicability by limiting any type of spectral coding that may be added to improve noise performance. Hence, the other algorithm considered did not make any assumption about the spectrum.

The second algorithm investigated was the spectral fit algorithm. Although the accuracy of the algorithm (i.e., average value of estimates) was reasonable for SNRs between 6 and 36 dB and half-space attenuation values between 0.05 and 1 dB/cm-MHz for Hamming window lengths greater than 4 mm, the precision was only reasonable for SNRs between 23 and 36 dB for half-space attenuation values of 0.05 and 0.3 dB/cm-MHz and SNRs between 28 and 36 dB for half-space attenuation values from 0.4–0.6 dB/cm-MHz at window lengths of 8 mm. The precision degraded with increasing noise, decreasing window length, and increasing attenuation. Although the SF algorithm is very similar to the established method for determining just the scatterer size when the attenuation is known (Insana *et al.*, 1990), the accuracy and precision of the estimates have been degraded by the need to estimate two parameters instead of just one.

Even though the SF algorithm is not as precise as the traditional method where the attenuation is known, the amount of data (i.e., window length and number of rf echoes) needed to obtain reasonable estimates can be compared to the amount of data typically used to obtain estimates of the scatterer size and attenuation independently. The window length typically used when estimating scatterer size in this frequency range has varied from 2.5 to 16 mm (Oelze *et al.*, 2002), while Insana *et al.* (1995) obtained consistent results using a window length of 7.9 mm. Hence, using a window

length of 4 to 8 mm is reasonable over this frequency range. Similarly, a typical number of rf echoes used in the estimate has been 14 (Insana *et al.*, 1995) and 20 (Oelze *et al.*, 2002), on the same order as the 25 echoes used herein. When estimating the attenuation independently, the concept of window length does not apply because typically a very short window is used throughout the ROI (Baldeweck *et al.*, 1994). However, the number of rf echoes required to obtain precise attenuation estimates can still be compared. In the past, 256 echoes have been used when estimating attenuation (Baldeweck *et al.*, 1994; D'Astous and Foster, 1986), considerably more than the 25 echoes used for each estimate herein. The precision of the SF algorithm would be improved if 256 echoes were used per estimate instead of the 25. We decided to only use 25 echoes so that the results would be more comparable to what is currently done in QUS imaging.

Although the precision of the SF algorithm still has challenges, the robustness of the average estimates (i.e., accuracy) allows for improved quantification of the tissue properties. Consider the estimate of scatterer size for a tumor. Typically, estimates are obtained throughout the tumor and then averaged together in order to obtain a final size estimate for classifying the tumor (Oelze *et al.*, 2004). Hence, if the tumor was large enough to allow for many different estimates to be obtained from the SF algorithm and averaged, the final estimate would be very close to the true characteristic size for the tumor and would not be affected by any intervening tissue layers.

The situation is similar when quantifying acoustic radiation force impulse imaging (Nightingale *et al.*, 2000) or patient-specific dosimetry (Siddiqi *et al.*, 1992). In both applications, the intensity of the ultrasound, obtainable from the total attenuation, is the only quantity that cannot be estimated using established methods. However, the total attenuation at a particular tissue depth will not normally change drastically for neighboring tissue regions. Therefore, estimates from the SF algorithm could be averaged from neighboring tissue regions to obtain a more accurate estimate of the total attenuation. The size of the region over which the total attenuation estimates could be averaged without biasing the results is a subject of further study beyond the scope of this paper. As an added benefit, it may be possible to combine the scatterer size estimate also obtained from the SF algorithm with acoustic radiation force impulse imaging to develop a multiparameter, noninvasive diagnostic tool.

ACKNOWLEDGMENTS

This work was supported by the University of Illinois Research Board, by a NDSEG Fellowship awarded to T. A. Bigelow, and by a Beckman Institute Graduate Fellowship awarded to T. A. Bigelow.

Akita, M., and Ueda, M. (**1988**). "The effects of windowing on spectral estimation of echoes scattered by a random medium," J. Acoust. Soc. Am. **107**, 1243–1248.

Baldeweck, T., Herment, A., Laugier, P., and Berger, G. (**1994**). "Attenuation estimation in highly attenuating media using high frequencies: A comparison study between different meant frequency estimators," IEEE Ultrason. Symp., 1783–1786.

Bigelow, T. A., and O'Brien, Jr., W. D. (**2004**). "Scatterer size estimation in pulse–echo ultrasound using focused sources: Theoretical approximations and simulation analysis," J. Acoust. Soc. Am. **116**, 578–593.

Chaturvedi, P., and Insana, M. F. (**1996**). "Error bounds on ultrasonic scatterer size estimates," J. Acoust. Soc. Am. **100**, 392–399.

D'Astous, F. T., and Foster, F. S. (**1986**). "Frequency dependence of ultrasound attenuation and backscatter in breast tissue," Ultrasound Med. Biol. **12**, 795–808.

Dussik, K. T., and Fritch, D. J. (**1956**). "Determination of sound attenuation and sound velocity in the structures constituting the joints, and of the ultrasonic field distribution within the joints on living tissues and anatomical preparations, both in normal and pathological conditions," Public Health Service, Nat. Inst. Health Project A454, Progr. Report.

Goss, S. A., Johnston, R. L., and Dunn, F. (**1978**). "Compilation of empirical ultrasonic properties of mammalian tissues," J. Acoust. Soc. Am. **64**, 423–457.

Goss, S. A., Johnston, R. L., and Dunn, F. (**1980**). "Compilation of empirical ultrasonic properties of mammalian tissues. II," J. Acoust. Soc. Am. **68**, 93–108.

Hall, T. J., Insana, M. F., Harrison, L. A., and Cox, G. G. (**1996**). "Ultrasonic measurement of glomerular diameters in normal adult humans," Ultrasound Med. Biol. **22**, 987–997.

He, P., and Greenleaf, J. F. (**1986**). "Application of stochastic analysis to ultrasonic echoes—estimation of attenuation and tissue heterogeneity from peaks of echo envelope," J. Acoust. Soc. Am. **79**, 526–534.

Insana, M. F., Wagner, R. F., Brown, D. G., and Hall, T. J. (**1990**). "Describing small-scale structure in random media using pulse–echo ultrasound," J. Acoust. Soc. Am. **87**, 179–192.

Insana, M. F., Wood, J. G., Hall, T. J., Cox, G. G., and Harrison, L. A. (**1995**). "Effects of endothelin-1 on renal microvasculature measured using quantitative ultrasound," Ultrasound Med. Biol. **21**, 1143–1151.

Jongen, H. A. H., Thijssen, J. M., Aarssen, M., and Verhoef, W. A. (**1986**). "A general model for the absorption of ultrasound by biological tissues and experimental verification," J. Acoust. Soc. Am. **79**, 535–540.

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

Lizzi, F. L., Astor, M., Liu, T., Deng, C., Coleman, D. J., and Silverman, R. H. (**1997**). "Ultrasonic spectrum analysis for tissue assays and therapy evaluation," Int. J. Imaging Syst. Technol. **8**, 3–10.

Lu, Z. F., Zagzebski, J. A., and Lee, F. T. (**1999**). "Ultrasound backscatter and attenuation in human liver with diffuse disease," Ultrasound Med. Biol. **25**, 1047–1054.

Nassiri, D. K., and Hill, C. R. (**1986**). "The use of angular acoustic scattering measurements to estimate structural parameters of human and animal tissues," J. Acoust. Soc. Am. **79**, 2048–2054.

Nightingale, K. L., Nightingale, R. W., Palmer, M. L., and Trahey, G. E. (**2000**). "A finite element model of remote palpation of breast lesions using radiation force: Factors affecting tissue displacement," Ultrason. Imaging **22**, 35–54.

Oelze, M. L., and O'Brien, Jr., W. D. (**2002a**). "Frequency-dependent attenuation-compensation functions for ultrasonic signals backscattered from random media," J. Acoust. Soc. Am. **111**, 2308–2319.

Oelze, M. L., and O'Brien, Jr., W. D. (**2002b**). "Method of improved scatterer size estimation and application to parametric imaging using ultrasound," J. Acoust. Soc. Am. **112**, 3053–3063.

Oelze, M. L., Zachary, J. F., and O'Brien, Jr., W. D. (**2002**). "Characterization of tissue microstructure using ultrasonic backscatter: Theory and technique for optimization using a Gaussian form factor," J. Acoust. Soc. Am. **112**, 1202–1211.

Oelze, M. L., O'Brien, Jr., W. D., Blue, J. P., and Zachary, J. F. (**2004**). "Differentiation and characterization of rat mammary fibroadenomas and 4T1 mouse carcinomas using quatitative ultrasound imaging," IEEE Trans. Med. Imaging **23**, 764–771.

Siddiqi, T. A., O'Brien, Jr., W. D., Meyer, R. A., Sullivan, J. M., and Miodovnik, M. (**1992**). "Human *in situ* dosimetry: Differential insertion loss during passage through abdominal wall and myometrium," Ultrasound Med. Biol. **18**, 681–689.

Wear, K. A. (**2001**). "Fundamental precision limitations for measurements of frequency dependence of backscatter: Applications in tissue mimicking phantoms and trabecular bone," J. Acoust. Soc. Am. **110**, 3275–3282.

Wear, K. A. (**2002**). "A Gaussian framework for modeling effects of frequency-dependent attenuation, frequency-dependent scattering, and gating," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **49**, 1572–1582.

Ultrasonic excitation of a bubble near a rigid or deformable sphere: Implications for ultrasonically induced hemolysis

Sheryl M. Gracewski[a)]
Mechanical Engineering, Biomedical Engineering, and Rochester Center for Biomedical Ultrasound, University of Rochester, Rochester, New York 14627

Hongyu Miao
Mechanical Engineering, University of Rochester, Rochester, New York 14627

Diane Dalecki
Biomedical Engineering, and Rochester Center for Biomedical Ultrasound, University of Rochester, Rochester, New York 14627

(Received 5 October 2004; revised 15 December 2004; accepted 22 December 2004)

A number of independent studies have reported increased ultrasound bioeffects, such as hemolysis and hemorrhage, when ultrasound contrast agents are present. To better understand the role of cavitation in these bioeffects, one- and two-dimensional models have been developed to investigate the interactions between ultrasonically excited bubbles and model ''cells.'' First, a simple one-dimensional model based on the Rayleigh–Plesset equation was developed to estimate upper bounds for strain, strain rate, and areal expansion of a simulated red blood cell. Then, two-dimensional boundary element models were developed (with DynaFlow Inc.) to obtain simulations of asymmetric bubble dynamics in the presence of rigid and deformable spheres. The deformable spherical ''cell'' was modeled using Tait's equation of state for water, with a membrane approximated by surface tension that increases linearly with areal expansion. The presence of a rigid or deformable sphere had little effect on the bubble expansion, but caused an asymmetric collapse and jetting for the conditions considered. Predicted membrane areal expansions were found to be below critical values for hemolysis reported in the literature for the cases considered near the inertial cavitation threshold. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1858211]

PACS numbers: 43.80.Gx, 43.35.Ei, 43.35.Wa [FD] Pages: 1440–1447

I. INTRODUCTION

A number of studies have suggested that cavitational activity may be involved in producing some ultrasound bioeffects such as hemolysis and hemorrhage (e.g., Chen *et al.*, 2003; Carstensen *et al.*, 2000; Dalecki *et al.*, 1997; M. W. Miller *et al.*, 1996). Bioeffects increase significantly when bubbles in the form of echo contrast agents such as Albunex® or Optison® are present (Feril *et al.*, 2003; Kobayashi *et al.*, 2002; M. W. Miller *et al.*, 2001; Dalecki *et al.*, 1997; D. L. Miller and Thomas, 1993a; Dalecki *et al.*, 2000; Miller and Quddus, 2000; Wible *et al.*, 2003; Miller and Gies, 1998; 2000). In addition, thresholds of some bioeffects were observed to increase with increasing ultrasonic frequency, consistent with a cavitation mechanism (Brayman *et al.*, 1997; Dalecki *et al.*, 1997; D. L. Miller and Thomas, 1993b). However, classical cavitation theory for spherically symmetric bubble response cannot explain all observed phenomena [e.g., the dependence of hemolysis on hematocrit (M. W. Miller *et al.*, 1996)], and it is likely that bubbles collapse asymmetrically when in the proximity of cells or vessel walls. Therefore, in this paper, models are developed to investigate the response of a bubble in the vicinity of a deformable or rigid sphere to acoustic excitation so that potential mechanisms of ultrasonically induced hemolysis can be explored.

An ultrasonically excited bubble in a free field can expand and collapse violently, creating local pressures of higher amplitude than the pressures of the excitation field. If there is a nearby interface, the bubble motion will deviate from spherically symmetric and a high-speed liquid jet can be produced. The localized high pressures, strains, strain rates, streaming, and liquid jetting may all potentially contribute to observed ultrasound-induced bioeffects. To identify the most probable mechanisms of ultrasonically produced hemolysis, these ultrasonically generated fields need to be compared with measurements of red cell fragility.

Red blood cell fragility has been measured over a wide range of loading conditions from quasistatic to time harmonic. In quasistatic micropipette experiments, hemolysis was observed to occur when the cell areal expansion exceeds approximately 5% (Rand, 1964; Evans *et al.*, 1976). The corresponding critical membrane tension for lysis in the micropipette experiments was on the order of 10 dyne/cm. In shear flow experiments, critical values of shear stress for hemolysis were in the range 10^3–10^5 dyne/cm^2; lower values were obtained from cone-plate or Couette flow viscometer shear flows, while higher values were measured for nonuniform shear flows created by a liquid jet (Williams, 1973; Blackshear and Blackshear, 1987; Rooney, 1974). Critical shear stress was observed to depend on a number of factors including exposure time, liquid viscosity, temperature, tonicity, and red blood cell size. An estimate of the critical mem-

a)Electronic mail: grace@me.rochester.edu

brane tension from these shear flow experiments is of the same order of the magnitude 1–10 dyne/cm as that obtained from the micropipette experiments. Hemolysis near an ultrasonically excited stable hemispherical bubble oscillation or near an oscillating wire, both at 20 kHz, occurred when the vibration displacement amplitude exceeded a critical value of approximately 20 μm (Rooney, 1970; Williams *et al.*, 1970). It was hypothesized that the observed lysis was caused by the steady streaming flows generated near the pulsating objects. The estimated critical shear stresses were in the same range as for the steady shear flow experiments.

Much work has been published on spherical bubble behavior in ultrasonic fields and on asymmetric expansion and collapse of laser- or spark-generated bubbles near boundaries. However, fewer reports exist on the response of ultrasonically excited bubbles near boundaries. For example, Vogel *et al.* (1989) and Shima *et al.* (1981) experimentally observed the effects of distance from a solid planar boundary on bubble collapse for laser- and spark-generated bubbles, respectively. Generation of jets directed either toward or away from the wall, and maximum wall stress was observed to depend on the separation distance. If the surrounding liquid is assumed to be inviscid, incompressible, and irrotational, the boundary integral method can be used to simulate the asymmetric bubble response near rigid boundaries (Sato *et al.*, 1994; Zhang *et al.*, 1994; Krasovitski and Kimmel, 2001; Brujan *et al.*, 2002; Tomita *et al.*, 2002; and Brujan, 2004). Tomita *et al.* (2002) reported that the curvature of the rigid boundary can significantly influence the collapse of a nearby laser-generated bubble, and predicted higher jet velocities for convex surfaces. For ultrasonically excited bubbles, the frequency of the ultrasound as well as the distance to a rigid boundary determine whether a jet will form, with the influence of the boundary on the bubble motion weakening with increasing frequency (Sato *et al.*, 1994; Krasovitski and Kimmel, 2001).

More complicated bubble behavior was observed for laser-induced cavitation near an elastic boundary (Brujan *et al.*, 2001a, 2001b). More recently, the complex interaction of bubbles and cells exposed to ultrasound has been observed with high-speed cameras (Wolfrum *et al.*, 2002; and Kudo *et al.*, 2002). To the authors' knowledge, the only numerical investigation of bubble interaction with simulated cells is that reported by Chahine (1994). In that simulation, a liquid jet directed away from the simulated cell was observed in most cases.

The goal of the current paper is to further investigate the interaction between an ultrasonically excited bubble and a model biological cell. A simple spherically symmetric model and two-dimensional axisymmetric models for the bubble dynamics are developed. The models are used to investigate the interaction of an ultrasonically excited bubble with rigid and deformable spheres, and to predict the resulting fields in the liquid. Spherically symmetric results are obtained by using the Rayleigh–Plesset model because the assumptions of liquid incompressibility and viscosity are similar to the assumptions used for the two-dimensional boundary element model. The Rayleigh–Plesset model adequately describes the expansion and initial collapse phases of the bubble response. In Sec. II, this simple model is used to determine the strain, strain rates, and deformation for a spherical ''cell'' that has exactly the same material properties as the surrounding liquid. In Sec. III, a two-dimensional boundary element model is used to investigate the opposite extreme, i.e., bubble response and generated fields when the bubble is near a *rigid* sphere. Finally, in Sec. IV, the two-dimensional boundary element model is used to explore the bubble response near a *deformable* sphere.

II. SIMPLE SPHERICALLY SYMMETRIC BUBBLE MODEL

In this simple model, the effect of the cell on the bubble dynamics is neglected. The spherically symmetric response of a bubble to an acoustic excitation is obtained. The stresses and strains in the liquid surrounding the bubble provide a first-order approximation for the stresses and strains that a cell next to the bubble would experience, if the cell's properties are similar to those of the liquid. The mechanical properties of red blood cells are of the same order of magnitude as the surrounding plasma. For instance, both of them have almost the same density slightly higher than water and their compressibilities are close to each other (about $35\times10^{-12}\,\mathrm{cm}^2/\mathrm{dyne}$ for red blood cells and $40\times10^{-12}\,\mathrm{cm}^2/\mathrm{dyne}$ for human plasma, Wang *et al.*, 2000). Therefore, in this section, this simple spherically symmetric bubble model is used to estimate the strains and strain rates in the medium surrounding the bubble. However, in general, the presence of a cell will affect the bubble response, potentially destroying the spherical symmetry of the bubble response. A two-dimensional axisymmetric simulation to investigate the interaction of a bubble and cell is therefore discussed in the following sections.

If an acoustically excited bubble is much smaller than both the acoustic wavelength and the distance to the nearest boundary, then the pressure gradient in the acoustic field can be neglected and the bubble response, if stable, can be assumed spherically symmetric. In this case, a variety of models can be used to obtain the bubble response to an acoustic excitation (Leighton, 1994). For the Rayleigh–Plesset model (Leighton, 1994), one of the simplest models of spherically symmetric bubble response, the time dependence of the bubble radius $R(t)$ is found by numerically integrating a nonlinear ordinary differential equation

$$R\ddot{R}+\frac{3}{2}\dot{R}^2=\frac{1}{\rho}\left[\left(p_0-p_v+\frac{2\sigma}{R_0}\right)\left(\frac{R_0}{R}\right)^{3\Gamma}-\frac{2\sigma}{R}-\frac{4\mu\dot{R}}{R}-p_\infty\right], \tag{1}$$

where ρ is the liquid density, p_v the vapor pressure, σ the surface tension of the bubble, μ the fluid shear viscosity, Γ the ratio of the specific heats of the gas, p_0 the initial ambient pressure, and p_∞ the acoustic pressure at infinity.

The Rayleigh–Plesset model assumes the surrounding liquid is incompressible, a reasonable assumption during the relatively slow expansion and initial collapse phases of the bubble. Once the motion of the bubble surface is obtained from the Rayleigh–Plesset equation, the motion of the surrounding liquid is defined by the incompressibility assumption. Specifically, when the bubble of original radius R_0 ex-

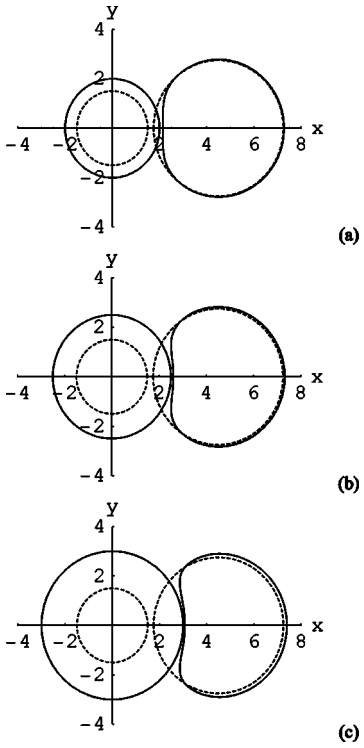


FIG. 1. Cross-sectional view of the deformation of a spherical surface moving with the surrounding incompressible fluid, as a bubble, centered at the origin, expands from $R_0=1.5\ \mu$m to (a) $R=2.0\ \mu$m; (b) 2.5 μm; and (c) 3.0 μm. The dashed lines indicate the positions of the original bubble and sphere (''cell'') with diameter 5.5 μm and center at 4.5 μm.

pands to a radius R, a fluid particle originally at a distance r_0 from the bubble center moves radially outward to a position r given by

$$r=[r_0^3+(R^3-R_0^3)]^{1/3}. \tag{2}$$

To illustrate the effect of bubble motion on the surrounding media, the deformation of a circular ring of fluid adjacent to the bubble was determined from Eq. (2) for various values of bubble expansion. This ring is a cross section of a sphere of fluid that represents a virtual cell. Plots for three values of bubble expansion are given in Fig. 1, for $R_0=1.5\ \mu$m, cell diameter $d_0=5.5\ \mu$m, and center-to-center distance of 4.5 μm. These dimensions were chosen to approximate the average Albunex® radius and typical human red blood cell volume. The spacing (0.25-μm gap) between the bubble and cell was based on average spacing assuming a 40% hematocrit and equally spaced spherical cells with a bubble centered in the interstitial space.

Even though the motion of the fluid is spherically symmetric about the origin, the off-center cell next to the bubble distorts as the bubble expands. Because a sphere has the minimum surface area for a given volume, the surface area must increase as the incompressible sphere deforms. A plot of percent areal expansion as a function of normalized bubble radius obtained by numerically integrating over the deformed sphere is given in Fig. 2. The solid curve in Fig. 2 is for the same parameters used for Fig. 1. Note that the percent areal expansion reaches the critical value for red cell lysis of ~5% when the bubble radius expands to just over twice its original value for the bubble-cell spacing shown in

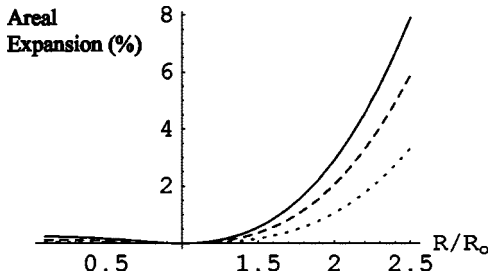


FIG. 2. Percent surface areal expansion of the deformed cell plotted versus normalized bubble radius for an initial gap between the bubble and cell equal to 0.25 μm (solid), 0.5 μm (dashed), and 1.0 μm (dotted). The parameters for the solid curve are the same as those of Fig. 1.

Fig. 1. However, the deformation, and therefore the surface areal expansion, decreases with distance from the bubble. In addition, a red blood cell will have a membrane tension that will resist the deformation. Therefore, these values only represent upper bounds on the areal expansion. High-speed photographs of bubble interactions with normal rat kidney fibroblast cells indicate that the cell stiffness can significantly reduce the cell deformation (Wolfrum *et al.*, 2002). In that paper, a fibroblast cell was estimated to experience only a 1.2% areal expansion in response to a 2-μm bubble expanding 30 times its original size. In this paper, an axisymmetric boundary element method will be used in Sec. IV to obtain a better approximation for the areal expansion and also to predict the induced membrane tension.

Lokhandwalla and Sturtevant (2001) also developed a model to predict membrane tension for a cell in a spherically symmetric, incompressible displacement field generated by an expanding bubble. They approximated the flow locally by a purely extensional flow and used a force balance based on Bernoulli's equation to estimate membrane tension for a disk-shaped cell. It is difficult to compare the present results to their results, however, because they considered a local field for a cell far from the bubble and the present results are for a simulated cell in the near field.

The strains and strain rates experienced by the liquid surrounding an acoustically excited bubble can be estimated from the bubble radius time curve $R(t)$ predicted from the Rayleigh–Plesset equation using the incompressibility equation [Eq. (2)]. For a spherically symmetric, incompressible deformation, in spherical coordinates, the nonzero strain rates at a radial position r are given by $\dot{e}_{\theta\theta}=\dot{e}_{\phi\phi}=u_r/r$ and $\dot{e}_{rr}=-2\dot{e}_{\theta\theta}$, where u_r is the radial velocity component at that position. For this strain rate tensor, the maximum shear strain rate is equal to $3/2*\dot{e}_{\theta\theta}$ and therefore the maximum shear stress can be approximated by $3/2*\mu*\dot{e}_{\theta\theta}$, where μ is the shear viscosity coefficient. A time derivative of the incompressibility condition given by Eq. (2), results in an expression for radial velocity given by

$$u_r\equiv\dot{r}=\frac{R^2\dot{R}}{r^2}. \tag{3}$$

Therefore, all strains and stresses decrease as $1/r^3$ with distance from the bubble. In terms of the original distance r_0 from the bubble when in equilibrium, this approximation for the shear stress can be expressed as

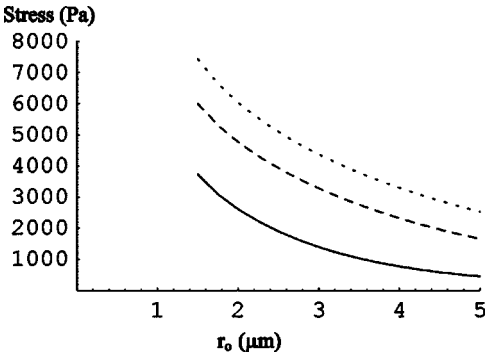


FIG. 3. Maximum shear stress during expansion as a function of initial radial position r_0 for an acoustic pressure of 0.2 MPa (solid), 0.4 MPa (dashed), and 0.6 MPa (dotted). Results are for a bubble with $R_0=1.5\ \mu$m in water exposed to a 1-MHz acoustic pressure.

$$\frac{3}{2}\mu\frac{R^2\dot{R}}{r_0^3+(R^3-R_0^3)}. \tag{4}$$

The maximum shear stress during bubble expansion is plotted in Fig. 3 as a function of r_0 for three acoustic pressure amplitudes. Calculations were done using parameters for an air bubble (specific heat ratio $\Gamma=1.4$) in water (density $\rho=1000$ kg/m^3, viscosity $\mu=10^{-3}$ kg/m·s, vapor pressure $p_v=2.3\times10^3$ Pa, and surface tension $\sigma=0.077$ N/m) and a 1-MHz excitation. As shown in Fig. 3, the maximum shear stress during bubble expansion occurs near the bubble surface and decays rapidly with distance from the bubble.

The Rayleigh–Plesset model predicts that a 1.5-μm air bubble expands to approximately twice its original radius in response to a 1-MHz, 0.2-MPa acoustic pressure wave in water, as shown in Fig. 4. A radial expansion of $R\approx2R_0$ is one criterion that has been used for the onset of inertial cavitation (Flynn and Church, 1988). Above the inertial cavitation threshold, the bubble collapses violently, generating localized pressures orders of magnitude larger than the acoustic excitation pressure. The following shear stress calculation is for an excitation pressure near the inertial cavitation threshold to determine whether stresses during bubble expansion are large enough to produce hemolysis.

The maximum tensile strain and shear stress were calculated for an excitation pressure amplitude of 0.2 MPa for a particle originally at $r_0=4.5\ \mu$m. This would represent the center of a 5.5-μm-diameter cell located 0.25 μm from a

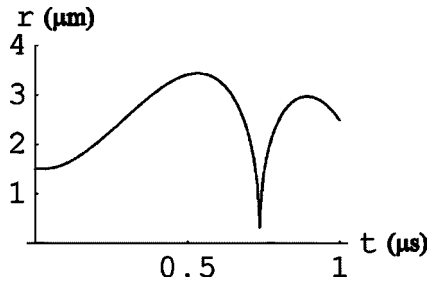


FIG. 4. Bubble radius versus time for an initial bubble radius $R_0=1.5\ \mu$m predicted by the Rayleigh–Plesset equation in response to a 1-MHz, 0.2-MPa acoustic wave. This acoustic pressure is just above the inertial cavitation threshold, and so the bubble expands to approximately twice its original size and collapses rapidly.

bubble with equilibrium radius 1.5 μm, as in Fig. 1. During the bubble expansion phase, the Rayleigh–Plesset model predicts a maximum tensile strain at $r_0=4.5\ \mu$m of 0.11 and maximum shear stress of 580 Pa. This is above the critical shear stress measured for hemolysis in steady shear flows but below the critical values measured in nonsteady flows. Therefore, it is unlikely this shear stress of duration less than 1 μs would cause hemolysis. Because the bubble collapses rapidly during inertial cavitation, the strain rates and stresses during the collapse phase will be even greater than those occurring during the expansion phase, though for a much shorter time. The Rayleigh–Plesset model predicts a maximum shear stress of 1700 Pa at $r_0=4.5\ \mu$m for the collapse phase. However, this value must be considered only as an approximation because many of the assumptions of the Rayleigh–Plesset model are no longer valid during the rapid inertial collapse of a bubble. Even so, a shear stress of this magnitude and short duration ($<0.1\ \mu$s) is unlikely to cause hemolysis.

III. BUBBLE NEAR A RIGID SPHERE

A two-dimensional axisymmetric boundary element method was used to investigate the effect of a nearby object, such as a cell, on a bubble's response to an acoustic field. In this section, the effect of a fixed rigid sphere on the pressure fields created by the bubble's motion is considered. In the next section, a more realistic deformable model for the cell is used to also predict the cell response to the bubble motion.

Here, the interaction of an acoustically excited bubble with initial radius $R_0=1.5\ \mu$m with a nearby fixed rigid sphere with diameter $d_0=5.5\ \mu$m was investigated using the boundary element method (developed with DynaFlow Inc., Jessup, MD) to reveal the details of the pressure field during the bubble expansion and collapse phases. These are the same dimensions as used in the previous section. As in the Rayleigh–Plesset model, the fluid is assumed incompressible and inviscid, and the state of the gas inside the bubble is assumed spatially uniform. Details of the boundary element formulation used for bubble dynamics simulations can be found in Kalumuck *et al.* (1995), Chahine and Kalumuck (1998). Unless otherwise stated, an acoustic sinusoidal pressure field with amplitude 0.2 MPa and frequency 1 MHz was imposed at infinity, beginning with the negative phase. The point on the rigid sphere surface which is the closest to the bubble will be the point which experiences the highest pressure during bubble collapse. If this point is denoted as the topmost point [as shown in Fig. 5(b)] then h is defined as the distance from the original bubble center to the topmost point.

Figure 5 presents an overall view of the pressure field around the bubble and the cell for two cases where $h=3.6\ \mu$m and $h=4.5\ \mu$m. The distances were chosen so that when the bubble reaches its maximum radius during expansion, the minimum distances from bubble boundary to sphere surface are 0.1 and 1.0 μm, respectively. Figures 5(a), (c), and (e) are the field pressure distributions for $h=3.6\ \mu$m; Figs. 5(b), (d), and (f) are for $h=4.5\ \mu$m. Figures 5(a) and (b) show the pressure distribution corresponding to the maximum bubble wall velocity during expansion; 5(c) and (d) correspond to the moment when the bubble reaches the

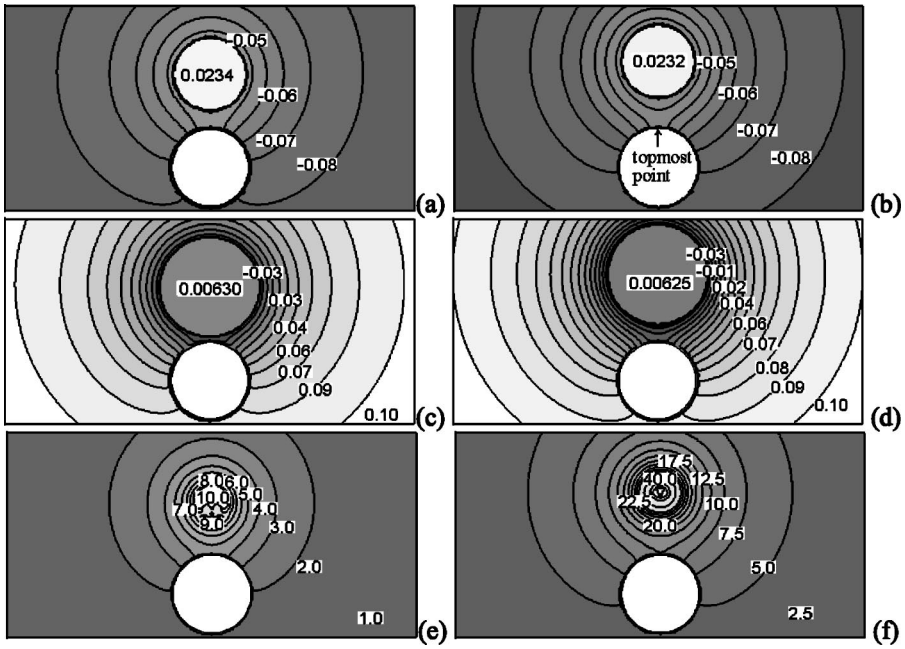


FIG. 5. Field pressure contours for $h=3.6\,\mu$m [panels (a), (c), (e)] and $h=4.5\,\mu$m [panels (b), (d), (f)] predicted using BEM model. h is the distance from the original bubble center to the topmost point of the rigid sphere. The fixed rigid sphere is the sphere near the bottom of each figure. Panels (a) and (b) show the pressure distribution corresponding to the maximum bubble wall velocity during expansion; (c) and (d) correspond to the moment when the bubble reaches its maximum radius; (e) and (f) show the moment of jet penetration (the time when the water jet from the upper side of the bubble touches the lower side of the bubble).

maximum bubble radius; 5(e) and (f) show the moment of jet penetration. Jet penetration is defined as the time when the liquid jet just reaches the far bubble surface and calculations are terminated.

Figures 5(a) through (d) indicate that during the bubble expansion phase the presence of the bubble mitigates the pressure field so the sphere experiences pressure amplitudes that are lower than the imposed acoustic pressure. However, upon bubble collapse, Figs. 5(e) and (f) show that the upper part of the rigid sphere experiences a pressure approximately 2 orders of magnitude higher than the applied acoustic pressure. In addition, these figures show that higher pressures are produced when the bubble is further from the sphere. At the moment of jet penetration, the pressure at the topmost point for $h=4.5\,\mu$m is about 10.3 MPa, which is 2.8 times the pressure at the topmost point for $h=3.6\,\mu$m. Coincidently, the bubble volume for $h=4.5\,\mu$m is smaller than the volume for $h=3.6\,\mu$m at the jet penetration, which indicates that for $h=4.5\,\mu$m the bubble contracts more during collapse. From these results we can hypothesize that the bubble collapse will be more spherically symmetric for a bubble further from the sphere, thereby generating a higher pressure inside the bubble. In addition, the maximum bubble radius is 0.2% larger for the bubble with $h=4.5\,\mu$m, which will also contribute to a more rapid collapse. This is consistent with results presented in Sato *et al.* (1994) that show an even larger decrease in maximum bubble radius for decreasing separation between a bubble and an infinite rigid planar boundary. The bubble shapes during the collapse phase are compared for the two cases in Fig. 6, illustrating that the bubble collapse is more aspherical and the jet forms earlier for $h=3.6\,\mu$m than for $h=4.5\,\mu$m. For comparison, the Rayleigh–Plesset model for a spherically symmetric bubble in an infinite liquid predicts peak pressures of 15.2 and 12.2 MPa, respectively, for field points 3.6- and 4.5 μm from the bubble center.

To investigate the influence of the presence of the fixed rigid sphere on the pressure field, the field pressure obtained from the Rayleigh–Plesset model is compared to that obtained by using the boundary element method (BEM) for $h=4.5\,\mu$m and $h=3.6\,\mu$m (Fig. 7). The dashed curve indicates the field pressure for the Rayleigh–Plesset model and the solid curve represents the field pressure at the topmost point on the sphere for $h=4.5\,\mu$m. During the bubble expansion and initial collapse phases, the pressures differ by less than 0.005 MPa. Even for $h=3.6\,\mu$m, as indicated by

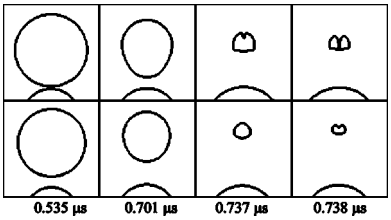


FIG. 6. Bubble shape comparison for $h=3.6\,\mu$m (upper row) and $h=4.5\,\mu$m (lower row) with time labeled at the bottom for a bubble near a rigid sphere.

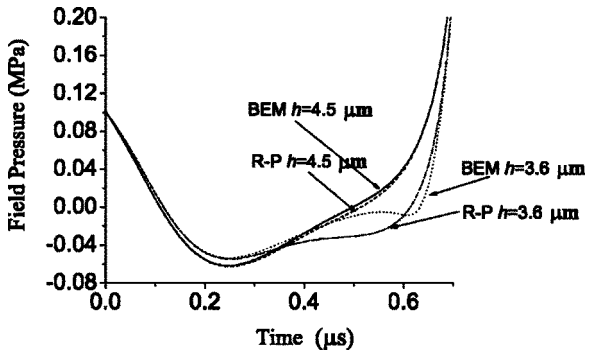


FIG. 7. Comparisons of field pressure versus time during bubble expansion. The dashed line is given by the Rayleigh–Plesset model for $h=4.5\,\mu$m and the dash-dotted line for $h=3.6\,\mu$m; the solid line for $h=4.5\,\mu$m and the dotted line for $h=3.6\,\mu$m are given by the BEM model at the topmost point of a fixed rigid sphere with diameter $d_0=5.5\,\mu$m.

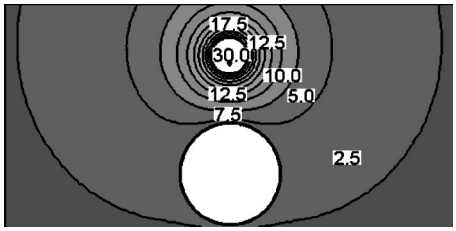


FIG. 8. Field pressure contour for $h=4.5\ \mu$m at the time of jet penetration predicted using BEM model.

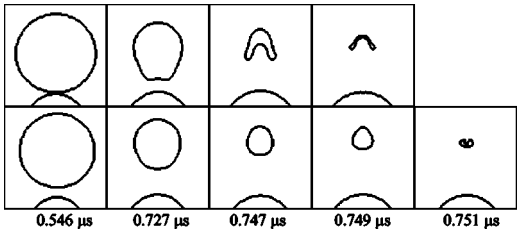


FIG. 9. Bubble shape comparison for $h=3.6\ \mu$m (upper row) and $h=4.5\ \mu$m (lower row) with time labeled at the bottom for a bubble near a deformable sphere.

the dotted and the dash-dotted curves, the maximum pressure difference is less than 0.05 MPa during this time range. So, even if the original distance between the bubble and the sphere is small such that the bubble almost touches the sphere when the bubble reaches its maximum radius, the influence of the rigid sphere on the bubble expansion phase is small.

IV. BUBBLE NEAR A DEFORMABLE SPHERE

In this section, a model is developed using the boundary element method to investigate the interaction of an acoustically excited bubble with a nearby movable and deformable sphere. The interior of the sphere is governed by the Tait's equation of state for water

$$P=A(\rho/\rho_0)^m-B, \tag{5}$$

where P and ρ are the time-varying pressure and density of the liquid, respectively, and ρ_0 is the equilibrium liquid density. The constants used here are $A=300.1$ MPa, $B=300.0$ MPa, $m=7$, and $\rho_0=1000\ \mathrm{kg/m^3}$.

The elasticity of the cell membrane is approximated using a surface tension that increases linearly with the areal expansion. The surface tension T was calculated from the equation

$$T=K\frac{\Delta A}{A_0}, \tag{6}$$

where A_0 is the original surface area of the cell, ΔA the current surface area minus the original surface area of the cell, and K the areal expansion modulus. In the simulations, $K=0.5$ N/m was used, based on the measured value for the red blood cell membrane areal expansion modulus at room temperature of $K=0.45$ N/m $\pm$ about 15 to 20 percent [Waugh and Evans (1979); Hochmuth (1987)]. The assumptions described previously for the surrounding liquid and bubble are also used in these simulations.

The deformability and mobility of the cell influence not only the pressure distribution but also bubble shape and jet formation. First, as a comparison, the pressure distribution around the cell at jet penetration is given in Fig. 8 for the case $h=4.5\ \mu$m. The pressure distributions when the bubble reaches its maximum bubble wall velocity and maximum bubble radius are not given because they are not significantly different from those for the rigid sphere case. A pressure value of 2.3 MPa at the topmost point of the cell is predicted at jet penetration, which is approximately 22% of the pressure near the top of the rigid sphere (about 10.3 MPa). For the case $h=3.6\ \mu$m, the pressure near the top of the cell at jet penetration is about 2.1 MPa, approximately 57% of the pressure for the case of rigid sphere at the same distance (about 3.7 MPa). So, during bubble collapse, the movable and deformable sphere experienced a peak pressure less than 60% of the peak pressure experienced by a fixed rigid sphere.

The jet was directed toward the deformable sphere for $h=4.5\ \mu$m, but away from it for $h=3.6\ \mu$m, as shown in Fig. 9. As mentioned in Sec. I, Chahine (1994) also predicted jet formation away from a deformable spherical globule. However, the model he used for the deformable sphere was much less stiff, governed by the gas-state equation with polytropic constant $k=10$, with a constant surface tension, either equal to 0.77 N/m or a much larger value of 10^4 N/m. Therefore, in his model the globule acted similar to a pressure release surface. In addition, Fig. 9 indicates that the collapse of a bubble farther from the deformable sphere will be more spherical than the collapse of a bubble closer to the sphere for a longer period of time.

To show the influence of the distance between a bubble and a deformable sphere on the response of the sphere, the sphere's areal expansion versus time is plotted in Fig. 10 for $h=4.5\ \mu$m and $h=3.6\ \mu$m. Even for $h=3.6\ \mu$m, the maximum cell areal expansion is only 0.007% and 0.022% during the bubble expansion and collapse phases, respectively. These values are well below the threshold for cell damage (on the order of 5%). The spherically symmetric model, developed as an upper bound for the deformation, significantly overestimates the resulting areal expansion. For $h=3.6\ \mu$m, the simple spherically symmetric model predicts an areal expansion of 0.28% when $R/R_0=2$. This is 40 times the value obtained using the BEM model of a bubble near a deformable sphere for the acoustic pressure amplitude 0.2 MPa. In the two-dimensional model, the bubble can move away from the cell and fluid can flow from between the bubble and cell. Both of these effects will tend to reduce the cell deformation and thus the areal expansion.

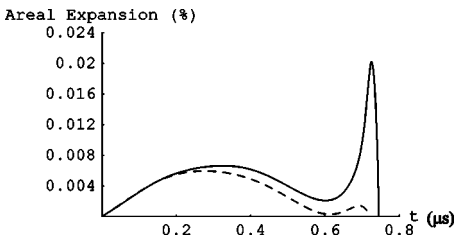


FIG. 10. Cell areal expansion (%) for two different distances. The solid line is for $h=3.6\ \mu$m; the dashed line is for $h=4.5\ \mu$m.

The results presented in this paper consider only the bubble expansion and collapse up until the jet reaches the far side of the bubble. Results indicate it is unlikely that bubble expansion could cause hemolysis at the cavitation threshold, but could potentially cause sufficient cell deformation to cause hemolysis at higher acoustic pressures. It is important to extend this analysis to the final collapse phase, e.g., to determine the fluid field after jet penetration, to more fully understand mechanisms of ultrasound-induced hemolysis. In addition, the effects of cell shapes, multiple cells, and media viscosity, viscoelelasticity, and compressibility will be considered in a future publication.

V. SUMMARY

Three models have been developed to explore the interaction of an acoustically driven bubble with a simulated biological cell. The simplest model assumes spherically symmetric bubble motion, neglecting the effect of the cell on the bubble, and can be used to predict upper bounds on the areal expansion, strains, and stresses for a cell near an expanding bubble. An axisymmetric boundary element model of an acoustically excited bubble near a rigid sphere indicates that the presence of the sphere on the bubble expansion is small, but may cause an asymmetric collapse that leads to jetting. Finally, a model for a deformable sphere was developed within the boundary element code. The interior of the cell was modeled using Tait's equation of state for water, and the elasticity of the cell membrane was represented by a surface tension that increased linearly with areal expansion. This model can be used to predict the pressure and velocity fields around simulated cells and areal expansion of their membranes. The predicted areal expansion during bubble expansion in response to a 1-MHz, 0.2-MPa ultrasonic wave was well below the threshold for cell lysis for the bubble to cell separations considered.

ACKNOWLEDGMENTS

The authors gratefully acknowledge helpful discussions with Edwin Carstensen and David Blackstock, and the encouragement and support of Morton Miller. This work was supported by NIH research grants R37CA39230-28, R37EB00213-28, and R01HL69824.

Blackshear, P. L., and Blackshear G. L. (**1987**). ''Mechanical Hemolysis,'' in *Handbook of Bioengineering*, edited by R. Skalak and S. Chien (McGraw-Hill, New York), pp. 15.1–15.19.

Brayman, A. A., Strickler, P. L., Luan, H., Barned, S. L., Raeman, C. H., Cox, C., and Miller, M. W. (**1997**). ''Hemolysis of 40% hematocrit, Albunex®-supplemented human erythrocytes by pulsed ultrasound: Frequency, acoustic pressure and pulse length dependence,'' Ultrasound Med. Biol. **23**, 1237–1250.

Brujan, E. A. (**2004**). ''The role of cavitation microjets in the therapeutic applications of ultrasound,'' Ultrasound Med. Biol. **30**, 381–387.

Brujan, E. A., Keen, G. S., Vogel, A., and Blake, J. R. (**2002**). ''The final stage of the collapse of a cavitation bubble close to a rigid boundary,'' Phys. Fluids **14**, 85–92.

Brujan, E. A., Nahen, K., Schmidt, P., and Vogel, A. (**2001a**). ''Dynamics of laser-induced cavitation bubbles near an elastic boundary,'' J. Fluid Mech. **433**, 251–281.

Brujan, E. A., Nahen, K., Schmidt, P., and Vogel, A. (**2001b**). ''Dynamics of laser-induced cavitation bubbles near elastic boundaries: Influence of the elastic modulus,'' J. Fluid Mech. **433**, 283–314.

Carstensen, E. L., Gracewski, S., and Dalecki, D. (**2000**). ''The search for cavitation *in vivo*,'' Ultrasound Med. Biol. **26**, 1377–1385.

Chahine, G. L. (**1994**). ''Cavitation dynamics at microscale level,'' J. Heart Valve Dis. **3**, S103–S116.

Chahine, G. L., and Kalumuck, K. M. (**1998**). ''BEM software for free surface flow simulation including fluid–structure interaction effects,'' Int. J. Comput. Applic. Technol. **11(3–5)**, 177–198.

Chen, W.-S., Brayman, A. A., Matula, T. J., and Crum, L. A. (**2003**). ''Inertial cavitation dose and hemolysis produced *in vitro* with or without Optison®,'' Ultrasound Med. Biol. **29**, 725–737.

Dalecki, D., Child, S. Z., Raeman, C. H., Xing, C., Gracewski, S., and Carstensen, E. L. (**2000**). ''Bioeffects of positive and negative acoustic pressures in mice infused with microbubbles,'' Ultrasound Med. Biol. **26**, 1327–1332.

Dalecki, D., Raeman, C. H., Child, S. Z., Cox, C., Francis, C. W., Meltzer, R. S., and Carstensen, E. L. (**1997**). ''Hemolysis *in vivo* from exposure to pulsed ultrasound,'' Ultrasound Med. Biol. **23**, 307–313.

Evans, E. A., Waugh, R., and Melnik, L. (**1976**). ''Elastic area compressibility modulus of red cell membrane,'' Biophys. J. **16**, 585–595.

Feril, Jr., L. B., Kondo, T., Zhao, Q.-L., Ogawa, R., Tachibana, K., Kudo, N., Fujimoto, S., and Nakamura, S. (**2003**). ''Enhancement of ultrasound-induced apoptosis and cell lysis by echo-contrast agents,'' Ultrasound Med. Biol. **29**, 331–337.

Flynn, H. G., and Church, C. C. (**1988**). ''Transient pulsations of small gas bubbles in water,'' J. Acoust. Soc. Am. **84**, 985–998.

Hochmuth, R. M. (**1987**). ''Properties of red blood cells,'' in *Handbook of Bioengineering*, edited by R. Skalak and S. Chien (McGraw-Hill, New York), pp. 12.1–12.17.

Kalumuck, K. M., Duraiswami, R., and Chahine, G. L. (**1995**). ''Bubble dynamics fluid–structure interaction simulation by coupling fluid BEM and structural FEM codes,'' J. Fluids Struct. **9**, 861–883.

Kobayashi, N., Yasu, T., Yamada, S., Kudo, N., Kuroki, M., Kawakami, M., Miyatake, K., and Saito, M. (**2002**). ''Endothelial cell injury in venule and capillary induced by contrast ultrasonography,'' Ultrasound Med. Biol. **28**, 949–956.

Krasovitski, B., and Kimmel, E. (**2001**). ''Gas bubble pulsation in a semi-confined space subjected to ultrasound,'' J. Acoust. Soc. Am. **109**, 891–898.

Kudo, N., Miyaoka, T., Okada, K., and Yamamoto, K. (**2002**). ''Study on mechanism of cell damage caused by microbubbles exposed to ultrasound,'' Ultrasonics Symposium Proceedings, pp. 1351–1354.

Leighton, T. G. (**1994**). *The Acoustic Bubble* (Academic, London).

Lokhandwalla, M., and Sturtevant, B. (**2001**). ''Mechanical haemolysis in shock wave lithotripsy (SWL). I. Analysis of cell deformation due to SWL flow-fields,'' Phys. Med. Biol. **46**, 413–437.

Miller, D. L., and Gies, R. A. (**1998**). ''Gas-body-based contrast agent enhances vascular bioeffects of 1.09 MHz ultrasound on mouse intestine,'' Ultrasound Med. Biol. **24**, 1201–1208.

Miller, D. L., and Gies, R. A. (**2000**). ''The influence of ultrasound frequency and gas-body composition on the contrast agent-mediated enhancement of vascular bioeffect in mouse intestine,'' Ultrasound Med. Biol. **26**, 307–313.

Miller, D. L., and Quddus, J. (**2000**). ''Diagnostic ultrasound activation of contrast agent gas bodies induces capillary rupture in mice,'' Proc. Natl. Acad. Sci. U.S.A. **97**, 10179–10184.

Miller, D. L., and Thomas, R. M. (**1993a**). ''Contrast agent gas bodies enhance hemolysis induced by lithotripter shockwaves and high-intensity focused ultrasound in whole blood,'' Ultrasound Med. Biol. **22**, 1089–1095.

Miller, D. L., and Thomas, R. M. (**1993b**). ''Frequency dependence of cavitation activity in a rotating tube exposure system compared to the mechanical index,'' J. Acoust. Soc. Am. **93**, 3475–3480.

Miller, M. W., Miller, D. L., and Brayman, A. A. (**1996**). ''A review of the *in vitro* bioeffects of inertial ultrasonic cavitation from a mechanistic perspective,'' Ultrasound Med. Biol. **22**, 1131–1154.

Miller, M. W., Everbach, E. C., Cox, C., Knapp, R. R., Brayman, A. A., and Sherman, T. A. (**2001**). ''A comparison of the hemolytic potential of Optison™ and Albunex® in whole human blood *in vitro*: Acoustic pressure, ultrasound frequency, donor and passive cavitation detection considerations,'' Ultrasound Med. Biol. **27**, 709–721.

Rand, R. P. (**1964**). ''Mechanical properties of the red cell membrane II. Viscoelastic breakdown of the membrane,'' Biophys. J. **4**, 303–316.

Rooney, J. A. (**1970**). ''Hemolysis near an ultrasonically pulsating gas bubble,'' Science **169**, 869–871.

Rooney, J. A. (**1974**). ''Hydrodynamic shearing of biological cells,'' J. Biol. Phys. **2**, 26–40.
Sato, K., Tomita, Y., and Shima, A. (**1994**). ''Numerical analysis of a gas bubble near a rigid boundary in an oscillatory pressure field,'' J. Acoust. Soc. Am. **95**, 2416–2424.
Shima, A., Takayama, K., Tomita, Y., and Miura, N. (**1981**). ''An experimental study on effects of a solid wall on the motion of bubbles and shock waves in bubble collapse,'' Acustica **48**, 293–301.
Tomita, Y., Robinson, P. B., Tong, R. P., and Blake, J. R. (**2002**). ''Growth and collapse of cavitation bubbles near a curved rigid boundary,'' J. Fluid Mech. **466**, 259–283.
Vogel, A., Lauterborn, W., and Timm, R. (**1989**). ''Optical and acoustic investigations of the dynamics of laser-produced cavitation bubbles near a solid boundary,'' J. Fluid Mech. **206**, 299–338.
Wang, S. H., Lee, L. P., and Lee, J. S. (**2000**). ''A linear relation between the compressibility and density of blood,'' J. Acoust. Soc. Am. **109**, 390–396.
Waugh, R., and Evans, E. A. (**1979**). ''Thermoelasticity of red blood cell membrane,'' Biophys. J. **26**, 115–131.
Wible, J. H., Galen, K. P., Wojdyla, J. K., Hughes, M. S., Klibanov, A. L., and Brandenburger, G. H. (**2002**). ''Microbubbles induce renal hemorrhage when exposed to diagnostic ultrasound in anesthetized rats,'' Ultrasound Med. Biol. **28**, 1535–1546.
Williams, A. R. (**1973**). ''Shear-induced fragmentation of human erythrocytes,'' Biorheology **10**, 303–311.
Williams, A. R., Hughes, D. E., and Nyborg, W. L. (**1970**). ''Hemolysis near a transversely oscillating wire,'' Science **169**, 871–873.
Wolfrum, B., Mettin, R., Kurz, T., and Lauterborn, W. (**2002**). ''Observations of pressure-wave-excited contrast agent bubbles in the vicinity of cells,'' Appl. Phys. Lett. **81**, 5060–5062.
Zhang, S., Duncan, J. H., and Chahine, G. L. (**1994**). ''The behavior of a cavitation bubble near a rigid wall,'' Bubble Dyn. Interface Phenom. 429–436.

Characteristics of the audio sound generated by ultrasound imaging systems

Mostafa Fatemi,[a] Azra Alizad, and James F. Greenleaf
Department of Physiology and Biomedical Engineering, Mayo Clinic College of Medicine, Rochester, Minnesota 55905

(Received 10 May 2004; revised 24 November 2004; accepted 4 December 2004)

Medical ultrasound scanners use high-energy pulses to probe the human body. The radiation force resulting from the impact of such pulses on an object can vibrate the object, producing a localized high-intensity sound in the audible range. Here, a theoretical model for the audio sound generated by ultrasound scanners is presented. This model describes the temporal and spectral characteristics of the sound. It has been shown that the sound has rich frequency components at the pulse repetition frequency and its harmonics. Experiments have been conducted in a water tank to measure the sound generated by a clinical ultrasound scanner in various operational modes. Results are in general agreement with the theory. It is shown that a typical ultrasound scanner with a typical spatial-peak pulse-average intensity value at 2 MHz may generate a localized sound-pressure level close to 100 dB relative to 20 μPa in the audible (<20 kHz) range under laboratory conditions. These findings suggest that fetuses may become exposed to a high-intensity audio sound during maternal ultrasound examinations. Therefore, contrary to common beliefs, ultrasound may not be considered a passive tool in fetal imaging.
[DOI: 10.1121/1.1852856]

PACS numbers: 43.80Gx, 43.25.Qp, 43.40.Ng [FD] Pages: 1448–1455

I. INTRODUCTION

Ultrasonography is virtually the only imaging method used today to monitor the fetus. Obstetricians use ultrasound scanners to monitor fetal activities and to conduct clinical tests for assessing fetal health (Manning, 1995). Some invasive procedures on the fetus are also performed under ultrasound monitoring (Berkowitz *et al.*, 1986; de Crespingy *et al.*, 1985). Ultrasound imaging devices are used frequently by numerous investigators to monitor fetal response in various types of stimulations, including acoustic signals (Querleu *et al.*, 1981; Birnholz and Benacerraf, 1983). The common assumption in all these applications is that ultrasound imaging is a ''quiet'' and nonstimulating monitoring method. Auditory effects of ultrasound have been investigated by several researchers. Electrical response to pulsed ultrasound in the cat auditory system has been demonstrated (Foster and Wiederhold, 1978). Human sensing of modulated or pulsed ultrasound applied to the middle ear has been studied by Gavrilov (1984a, 1984b), and a threshold for sensation of sinusoidally modulated ultrasound versus modulation frequency has been determined (Tsirulnikov *et al.*, 1988). The possibility of hearing pulsed ultrasound by bone conduction based on the acoustic radiation force has been considered (NCRP, 2002). There have been some indications that ultrasound may generate audible sound in the uterus (Arulkumaran *et al.*, 1991); however, it has been thought that such sound is not significant (Arulkumaran *et al.*, 1996). Fetal stimulation by ultrasound has been reported in the literature (Fatemi *et al.*, 2001; Saeian *et al.*, 1995). Such stimulation has been attributed to the sound generated by pulsed ultrasound (Fatemi *et al.*, 2001). The goal of this study is to investigate the audio sound generated by obstetric ultrasound scanners in terms of its spatial distribution, spectral composition, and relative intensities in various operational modes. In the following sections we present a theoretical model for the sound generated by pulsed ultrasound. In addition, we present the results of experiments demonstrating the sound generated by two clinical ultrasound-imaging systems in a water tank.

II. THEORY

Here, we briefly describe the radiation force phenomenon. When an acoustic wave is absorbed or reflected by an object, causing a change in the acoustic energy density, a steady force is exerted in the direction of the beam on that object. For normal incidence beams, the radiation force is given by (Westervelt, 1951; Beyer, 1978; Torr, 1984)

$$F=(1+R)W/c, \tag{1}$$

where W, c, and R are the total incident power on the object, speed of sound, and the power reflection coefficient of the object, respectively. ($R=0$ corresponds to a perfectly absorbing object). The force decreases for incidence angle less than 90 deg. When an ultrasound pulse impinges on an object, a radiation force is momentarily exerted on the object. Repeated pulse transmission produces an oscillating force on the object, causing vibrations at the pulse repetition frequency (PRF). This vibration is caused by direct conversion of ultrasound energy to low-frequency sound in the object.

Equation (1) describes the radiation force of an ultrasound beam. The radiation pressure or the stress, p_{rad}, on a target is related to the spatial averaged intensity (I) by

[a]Electronic mail: fatemi.mostafa@mayo.edu

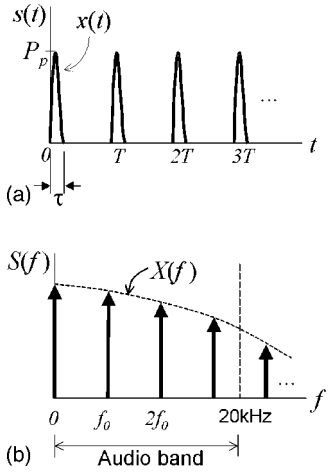


FIG. 1. (a) The radiation pressure pulse train on an object. (b) Frequency spectrum of the waveform in (a). Frequency components occur at the pulse repetition rate and its harmonics.

$$p_{\mathrm{rad}} = (1+R)I/c. \tag{2}$$

The spatial-averaged intensity for a single ultrasound pulse, or more conventionally, the spatial-average-pulse-average intensity (I_{SAPA}) depends on the beam shape. This is a parameter of the ultrasound probe, and can be written approximately in terms of the spatial-peak pulse-average intensity (I_{SPPA}), as $I_{\mathrm{SAPA}} = 0.5(I_{\mathrm{SPPA}})$ (Harris, 1985). The peak radiation pressure (P_p) produced during each incident pulse on an object being exposed to the ultrasound beam can be written as

$$P_p = 0.5(1+R)(I_{\mathrm{SPPA}})/c. \tag{3}$$

This pressure pulse is generated only during each pulse period (τ).

Let $x(t)$ be one such pressure pulse [Fig. 1(a)]. [In this notation, $x(t)$ is proportional to the envelope of the instantaneous intensity of the ultrasound pulse waveform.] The duration and peak of this pulse are τ and P_p, respectively. The radiation pressure waveform on the object, $s(t)$, is comprised of a series of pressure pulses $x(t)$ repeated at the pulse repetition frequency f_0

$$s(t) = \sum_m x(t - mT), \tag{4}$$

where m is an integer and $T = 1/f_0$, as shown in Fig. 1(a). The frequency spectrum of $s(t)$, denoted as $S(f)$, is comprised of a series of components, represented by a series of Dirac delta functions at frequencies $n\ f_0$, $n = 0, \pm 1, \pm 2, \ldots$, as seen in Fig. 1(b)

$$S(f) = \sum_n X(f)\,\delta(f - nf_0), \tag{5}$$

where $X(f)$ is the Fourier transform of $x(t)$ (Goodman, 1968). The value of the first component at $f = 0$ is the hydrostatic radiation pressure, and is denoted by P_0. If we approximate $x(t)$ as a narrow rectangular pulse, then $P_0 = \tau f_0 P_p$. In practice, $\tau \ll 1/f_0$; hence, we may approximate $x(t)$ by an impulse $P_0\delta(t)$. In such case, $X(f) = P_0$ is constant for all frequencies; hence, $S(f)$ may be written as

$$S(f) = P_0 \sum_n \delta(f - nf_0). \tag{6}$$

The audio portion of this signal is comprised of those components that fall within the audible range, i.e., $0 < nf_0 \leqslant 20$ kHz. Assuming the number of such component pairs is N, the time-domain representation of the audio signal, $s_a(t)$, is determined by taking the inverse Fourier transform of these N component pairs

$$s_a(t) = 2P_0 \sum_{n=1}^{N} \cos(2\pi n f_0 t). \tag{7}$$

The root-mean-square (rms) value of this portion of the signal, P_{rms}, can be written as

$$P_{\mathrm{rms}} = P_0\sqrt{2N}. \tag{8}$$

As an example, we calculate P_{rms} for a clinical ultrasound scanner (Acuson 128XP, Acuson Corp., Mountain View, CA). The parameters of this ultrasound scanner and its probe are as follows: $I_{\mathrm{SPPA}} = 273$ W/cm^2 for water (according to the manufacturer's manual), $\tau \cong 1\ \mu$s, $f_0 = 4$ kHz. Hence, for a perfectly absorbing object (i.e., $R = 0$), $P_0 = 2.84$ Pa. Given the value of f_0, we have $N = 5$, and finally $P_{\mathrm{rms}} = 8.98$ Pa, which corresponds to 113 dB relative to 20 μPa.[1] For a perfectly reflecting object, P_{rms} would be 6 dB higher.

It should be noted that P_{rms} represents the stress field exerted on the object by the ultrasound beam. This field is generated because the object changes the energy density of the impinging ultrasound through such mechanisms as reflection, absorption, or nonlinearity. The radiation stress would not be present without such changes in the energy density; therefore, it is object dependent. The radiation stress can vibrate the object and cause it to radiate a secondary acoustic field (acoustic emission) in the surrounding medium. The spatial distribution and the intensity of the acoustic emission field depend on the characteristics of the object and the medium. This is analogous to the sound generated in air by the vibration of a speaker membrane. The main objective in this paper is to investigate the stress field on the object, not the acoustic emission in the medium. We may interpret this stress field as a localized sound field, and we call it the ''ultrasound-generated sound'' (UGS). (In analogy with the speaker, we are interested in the sound pressure on the speaker membrane, not the sound in air.)

III. EXPERIMENTS

The following experiments were designed to measure the audio range ($\sim$15 Hz to 20 kHz) sound generated by the ultrasound scanner in a water tank 28 cm high, 48 cm wide, and 98 cm long. Considering the boundary conditions (walls and the water surface), the primary resonance frequencies of the water tank in height, width, and length are 1339, 1562, and 765 Hz, respectively. The experiment setup is shown in Fig. 2. In the first experiment, the ultrasound probe (model V2) of a clinical ultrasound scanner (Acuson 128XP) was secured in the vertical position with its tip slightly immersed in the water. The ultrasound system was tested in B mode, combined B mode, and cw Doppler, pulse Doppler, color

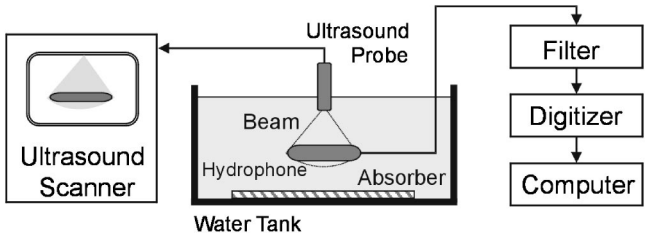


FIG. 2. Experimental setup for measuring the sound generated by the ultrasound beam of a clinical scanner on a hydrophone. The system includes a clinical ultrasound scanner with an array probe. The probe is slightly immersed in the water tank, projecting the ultrasound beam on the audio hydrophone. The sheet of ultrasound absorber is placed on the tank bottom to prevent possible reflections. Hydrophone output is bandpass filtered, then digitized and stored in the computer.

Doppler, and M mode. The pulse repetition rate ranged between 2 to 5 kHz. In the B mode setting, the pulse repetition frequency was 4 kHz, and the spatial-peak pulse-average intensity I_{SPPA} was 273 W/cm^2, which corresponded to the values used in the theory.

The ultrasound beam of the probe was projected on an audio hydrophone (ITC model 6080C). This hydrophone acts as an almost-perfect absorber. The power reflection of this hydrophone, measured independently by ultrasound pulse–echo method, is $R=0.09$. This hydrophone is approximately cylindrical with a diameter and length of 50 and 200 mm, respectively. The actual bandwidth of this hydrophone is about 100 kHz, but it is calibrated from 15 Hz to 55 kHz by the manufacturer, and its sensitivity in this range is -155 dB relative to 1 volt/μPa. The hydrophone includes a piezoelectric element that converts acoustic pressure to electrical signal, and a preamplifier to improve the signal-to-noise ratio. The radiation force of ultrasound on the hydrophone produced a localized sound that was measured by this hydrophone. The hydrophone signal was digitized at the rate of 1 MHz. The sound data were recorded in blocks of 40 000 samples for each experiment. To prevent the reflection of the ultrasound beam by the tank, a sheet of ultrasonic absorber was placed on the bottom of the tank. To measure the sound, the hydrophone was positioned horizontally, as shown in Fig. 2. The ultrasound probe was oriented to scan along the hydrophone's length. The vertical distance from the probe surface to the top surface of the hydrophone was 3.5 cm. The transmit focus of the probe was set at 4 cm and the depth control was set at 14 cm.

The second experiment was designed similarly to the first experiment except for the use of a different ultrasound scanner. The reason for this experiment was to demonstrate that sound generation phenomenon was not limited to a particular ultrasound scanner. However, because there are many differences in the design of these two scanners, it may not be fair to compare the results of experiments on one scanner with those of the other scanner in every aspect. In this experiment, a Toshiba ultrasound system, model SSA-250A, was used in a setting similar to Fig. 2. The experiment was conducted first with a 3.75-MHz curved-array probe and then with a 5.75-MHz curved-array probe. Scanner view depth was set at 15 cm, and the focus at 5.1 cm for the 3.75-MHz probe and 4.6 cm for the 5.75-MHz probe.

In the above experiments, the hydrophone received a localized radiation pressure (stress) field, as well as a secondary, and probably small, acoustic field that was generated in the surrounding water as a result of hydrophone vibrations. For this reason, the hydrophone signal in these experiments represents what we may call the ''effective sound,'' which is the combined result of vibrations due to the radiation force and the acoustic pressure in water.

IV. RESULTS

The sound measurement data are presented in two forms: wideband (received signal at its full bandwidth) and audio band (received signal filtered by a digital low-pass filter with 20-kHz cutoff frequency). In the first experiment (using the Acuson scanner), UGS signal is recorded at different operational modes of the scanner. Plots of the wideband sound and its spectrum are presented in Figs. 3(a) and (b), respectively. The silent region within the signal corresponds to the interframe interval. The frame rate was 35 per second. Uneven pressure amplitude during a frame is because of the change in the distance and the angle of beam incidence, which, in turn, changes the magnitude of radiation force on the hydrophone (Beyer, 1978). The frequency spectrum plot in Fig. 3(b) shows that the fundamental component of UGS is at 4 kHz, which corresponds to the pulse repetition rate of the scanner at the settings used for this experiment. The spectrum is rich in harmonics, which are present up to 100 kHz. Note that the frequency response of the hydrophone starts to attenuate the signal after about 55 kHz. The rms value of the wideband sound is found to be 19.35 Pa corresponding to 119.7 dB (relative to 20 μPa). The peak instantaneous sound level is 81.48 Pa (132.2 dB). To obtain the audio band portion of the sound, a digital low-pass filter with the cutoff frequency equal to 20 kHz is employed. The expanded version of the audio waveform is shown in Fig. 3(c). The rms value of this signal is 1.9 Pa (99.7 dB). Figure 3(d) shows the frequency spectrum of the audio band sound.

The ultrasound scanner produced different levels of sound at different operational modes. A list of the average and peak values of the wideband and audio range sound is presented in Table I. The continuous wave (cw) Doppler mode is not included because a cw ultrasound beam can only produce a static radiation pressure, which is not audible. Simultaneous B mode and cw Doppler is indicated by B/cw Doppler. The ultrasound power values in the case of combined pulse Doppler and B mode (where the B mode image and Doppler information can be seen simultaneously on the screen) are similar to those of pulse Doppler mode.

The results of the second experiment for the Toshiba scanner, with 3.75-MHz and 5.75-KHz probes, are presented in Fig. 4 and Tables II and III. As in the previous experiment, the sound levels are given for both the wideband and audible frequency spectra.

V. DISCUSSION

The theoretical results presented in this paper predict that the ultrasound beam from a clinical scanner can produce significant audio range radiation pressure on an object lo-

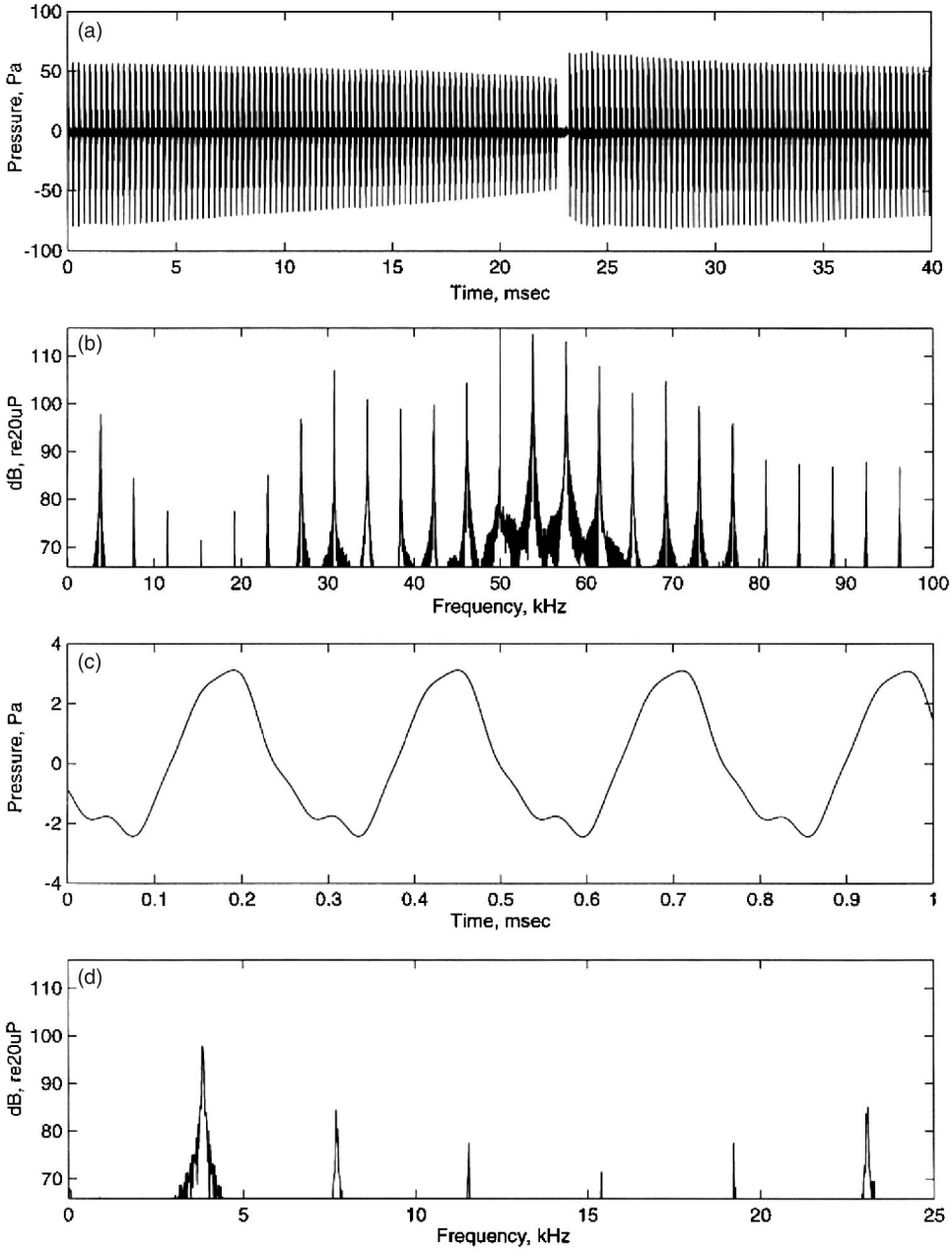


FIG. 3. Ultrasound-generated sound produced by Acuson 128XP scanner with a V2 probe on the hydrophone in water: (a) wideband (0–100-kHz) sound waveform; (b) power spectrum of the signal in (a); (c) audio band waveform; (d) audio band frequency spectrum corresponding to (c).

cated at the beam focal point. They also indicate that the frequency spectrum of this radiation pressure is composed of a main component at the PRF and a number of harmonics.

The experimental results demonstrate that the spectrum of the measured sound is in agreement with the theory. The effective sound measured in the experiment depends on the scanner type and mode settings. In B mode, the sound level measured in the experiment was 99.7 dB versus 113 dB in the theory. It appears that the color Doppler and pulse Doppler modes produce the highest average power values, which may be due to increased transmitted power levels at these modes. In the case of the Acuson scanner, Table I indicates that the B/cw Doppler mode has the lowest average power values in either the wideband or audio band cases. Comparing the results presented Tables II and III for the Toshiba scanner, the 5.75-MHz transducer appears to generate lower sound values than the 3.75-MHz transducer. This may be attributed to a change in the ultrasound output power of the transducer. The experimental results presented in Tables I, II, and III confirm that a clinical ultrasound scanner can produce significant sound on a hydrophone (80–100 dB) in the audible frequency range.

TABLE I. Ultrasound generated sound level (in dB relative to 20 μPa) at different imaging modes (Acuson 128XP scanner).

	Wideband ($f \leqslant 100$ kHz)		
Mode	Average (rms)	Peak	Audio band ($f \leqslant 20$ kHz) Average (rms)
B mode	119.7	132.2	99.7
B/cw Doppler	113.1	131.8	93.9
Pulse Doppler	119.6	131.9	98.7
Color Doppler	124.3	136.4	98.4
M mode	118.5	131.4	99.7

It should be noted that the experimental setup has some differences with the ideal conditions assumed in the theory.

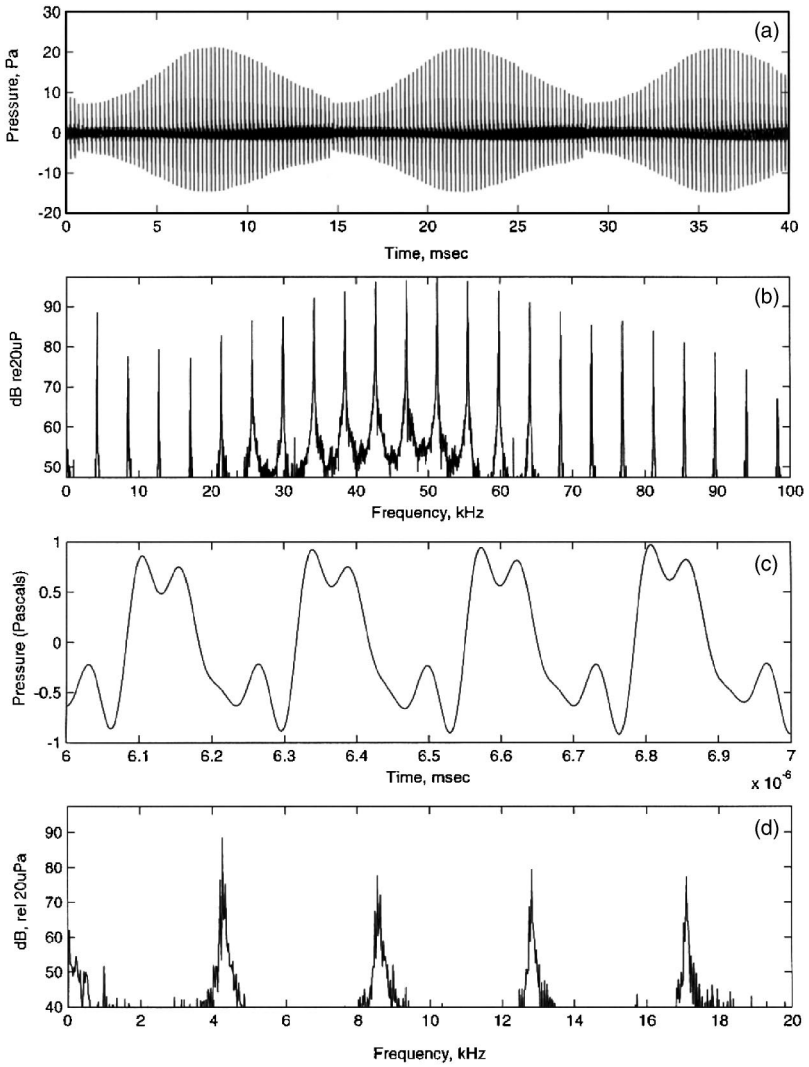


FIG. 4. Ultrasound-generated sound produced by Toshiba SSA-250A scanner with a 3.75-MHz phased-array probe on the hydrophone in water: (a) wideband (0–100-kHz) sound waveform; (b) power spectrum of the signal in (a); (c) power spectrum in the audio band; (c) audio band waveform; (d) audio band frequency spectrum corresponding to (c).

Thus, the theoretical and experimental results are not expected to match exactly. Our theoretical model assumes that ultrasound pulses are a series of idealized impulses, and hence, predicts the frequency spectrum to be in the form of equal-amplitude components at f_0 and its harmonics. In practice, however, ultrasound pulses have finite duration; therefore, the spectrum has variations in its amplitude as illustrated in Fig. 1(b). The spectrum of the sound generated by ultrasound scanners has a principal component at the PRF of the imaging system. The PRF values depend on the depth setting on the system. The PRF of the Acuson imaging system used in our experiment ranges approximately from 2 to 5 kHz for the depth setting from 24 to 4 cm, respectively. The spectrum presented in Fig. 3 is calculated from a portion of the signal within one frame. If the time window of the signal were chosen to be several times the frame duration, then a lower-frequency component at the frame rate of the system, which in this case is 35 Hz, would also appear in the spectrum. Modern ultrasound scanners include features that can produce complicated simultaneous image and Doppler display (for example, combined pulse Doppler and B mode). The ultrasound signal in these cases becomes more complex and additional low-frequency components can appear in the ultrasound-generated sound.

TABLE II. Ultrasound generated sound level (in dB relative to 20 μPa) at different imaging modes (Toshiba SSA-250A scanner, 3.75-MHz probe).

	Wideband ($f \leqslant 100$ kHz)		
Mode	Average (rms)	Peak	Audio band ($f \leqslant 20$ kHz) Average (rms)
B mode	104.1	120.5	87.3
B/cw Doppler	104.1	119.7	87.6
Pulse Doppler	117.9	131.1	99.7
Color Doppler	116.3	133.6	98.3
M mode	116.6	129.9	98.9

TABLE III. Ultrasound generated sound level (in dB relative to 20 μPa) at different imaging modes (Toshiba SSA-250A scanner, 5.75-MHz probe).

	Wideband ($f \leqslant 100$ kHz)		
Mode	Average (rms)	Peak	Audio band ($f \leqslant 20$ kHz) Average (rms)
B mode	100.3	113.3	83.9
B/cw Doppler	100.3	113.4	83.8
Pulse Doppler	109.5	121.4	94.9
Color Doppler	110.8	124.4	95.2
M mode	107.7	119.4	92.4

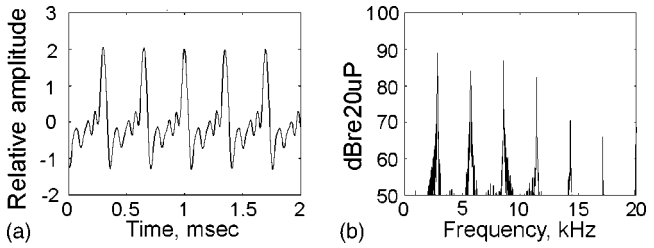


FIG. 5. (a) Probe case vibration waveform. (b) frequency spectrum of (a).

In the experiments we assumed that the measured sound was caused only by the vibration of the object in response to the radiation force of the impinging ultrasound. This assumption may not always be true. In fact, the ultrasound probe of the scanner vibrates at audio frequencies when activated in normal imaging mode. Here, we discuss possible sound emission from ultrasound probes. To investigate this possibility, we used Acuson model XP128P ultrasound scanner with a 2-MHz type V2 phased array probe and set the scanner at B mode operation. We held the probe in air and placed a hydrophone in contact with one side of the plastic case of the probe without touching the front face (piezoelectric array side). To ensure acoustic coupling, we applied some acoustic gel between the hydrophone and probe case. We recorded the hydrophone's output and subsequently filtered it using a 20-kHz low-pass filter. The resulting signal is shown in Fig. 5(a). The frequency spectrum of probe case vibrations, shown in Fig. 5(b), indicates energy concentration at the PRF (about 3 kHz at this setting) and its harmonics. We observed similar vibrations on all surfaces of the probe, including the active array side of the probe. This experiment shows that the probe body is a source of audio range vibrations that potentially can generate a sound in patient's body. However, we can show that the sound generated by the probe case is not significant compared to the sound generated directly on an object by the radiation force of the impinging ultrasound. To show this we conducted the following experiment. We measured the audio range sound generated by the ultrasound beam of the scanner (Acuson) using a small hydrophone. The scanner was set at B mode, depth of view of 22 cm, frame rate $=23$/s, and PRF $=2.9$ kHz. The ultrasound probe was secured in the vertical position with its tip slightly immersed in the water. We measured the sound with a wideband hydrophone (Bruel & Kjaer 8103). The frequency bandwidth of this hydrophone is from 0.1 Hz to 180 kHz. The hydrophone is shaped like a cylinder (diameter $=9.5$ mm, length $=15.8$ mm). The hydrophone was positioned at 19 cm below the probe, with its long axis perpendicular to the scan plane of the ultrasound probe. Thus, as the ultrasound beam scanned a frame from side to side, the time that it spent to pass over the diameter of the hydrophone was only a small portion of the frame period. We used a 20-kHz low-pass digital filter to obtain the audio band portion of the hydrophone signal. The resulting waveform is shown in Fig. 6. This figure shows that the audio signal is significant only when the ultrasound beam is impinging on the hydrophone. As demonstrated before, the entire probe case vibrates and produces a sound in the medium. Considering that the probe vibrates during the entire frame period, the absence of an appreciable audio sound during the time when the ultrasound beam is off the hydrophone indicates that the sound resulting from probe vibration is not significant compared to the sound generated by the radiation force of impinging ultrasound.

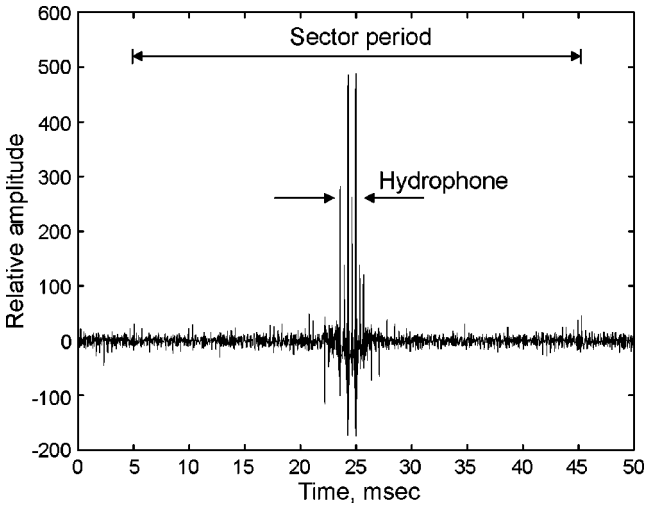


FIG. 6. Ultrasound-generated sound recorded by a small hydrophone in water. Period of the sector covered during a frame is shown by the long arrow. High-amplitude signals appear only when the ultrasound beam impinges on the hydrophone as indicated by small arrows.

The reader may be naturally intrigued to compare the ultrasound-induced sound level to the sound from a common sound source. In this paper, we mentioned that the sound generated through the radiation force mechanism at the focal point of an ultrasound beam might reach to 100 dB or higher. This sound level is high enough that we may compare it to what one may hear in a busy street or a concert hall. Such comparison immediately leads us to some questions, for example, ''if the ultrasound-generated sound is this high, why can't the patient or the physician hear it during an ultrasound exam?'' To answer this question, we need to realize that the pressure field produced by the radiation force of ultrasound is nonuniform, i.e., the radiation pressure of the ultrasound beam is concentrated at its focal point, and therefore we may model it as a dipole point source. The resulting pressure field propagates at every direction and decays due to diffraction (the so-called $1/r$ rule, where r is the radius from the point source) and attenuation by the medium. If the pressure value at the point source is finite, the field in the surrounding medium tends to diminish fast as the distance increases. This indicates that the high sound-pressure value (around 100 dB) that we estimated at the focal point of the ultrasound beam vanishes rapidly away from the focal point and hence will not be heard as a 100-dB sound in the surrounding medium. Another reason why the patient or the physician may not hear the sound produced by ultrasound radiation pressure is the loss of sound intensity at tissue–air interfaces. Due to huge impedance differences, the sound generated inside the human body is mostly reflected at such interfaces. As an illustrative example, consider the focal region of the ultrasound beam as a point (small sphere) source 1 mm in diameter, located 30 mm from the skin. Sound attenuation due to the $1/r$ rule amounts to 36 dB. The transmission coefficient of the sound from soft tissue to air is about -65 dB (Enderle

et al., 2000). Therefore, any sound generated at the focal region will be attenuated by 101 dB, making it impossible to be heard by persons near the patient.

In this paper, we did not elaborate on how the ultrasound-generated sound field propagates away from the focal point of ultrasound. That is because the local effects of such sound field may be more important than its effect on surroundings. For instance, the sound generated by ultrasound may have significant implications if the beam is directed toward the fetal hearing structures. In particular, ultrasound can exert a stronger radiation force on ossified middle-ear bones because they strongly reflect the impinging ultrasound. The resulting sound could be conducted to the cochlea efficiently (Fatemi, 2001). The hydrophone we used in the experiment was approximately a perfect absorber. The power reflection coefficient at bone-soft tissue interface is about 0.4 (McDicken, 1991). Hence, according to Eq. (2), the radiation pressure on the bone should be a factor of 1.4 (equivalent to 3 dB) higher than the value on a perfect absorber. Tissue attenuation is another factor to be considered when calculating the sound level in human body.[2] Another source of UGS is the interaction of ultrasound with attenuating soft tissue. The radiation force resulting from this interaction generates a sound in the soft tissue. The resulting sound is generated along the beam path and propagates through the body. However, this component is not expected to play a major role in fetal stimulation, as the fetal cochlea is not very sensitive to the sound traveling through body (Fatemi, 2001). Our experimental setup, where we recorded the effective sound produced by ultrasound on a hydrophone, to some extent approximates fetal hearing of the sound generated by ultrasound. However, it should be noted that the sound perceived by a fetus might vary because of the complexity and varieties of the pathways that transmit the sound to the fetal auditory system. One may also wonder what UGS sounds like. Hydrophone signals, which we recorded in our experiments, sounded like the noise from a road construction jackhammer. Multiple tones, including low frequencies (frame scanning rate), mid frequencies (pulse repetition frequency), and high frequencies (harmonics) were discernible in a rattling composition.

Finally, findings in this investigation suggest that, contrary to common assumption, ultrasound scanners cannot be considered as passive monitoring devices for fetal imaging. We should emphasize that the sound generated by ultrasound scanners is probably the source of fetal stimulation during clinical ultrasound examinations, as predicted in Fatemi *et al.,* 2001.

VI. CONCLUSIONS

Pulsed ultrasound from clinical scanners can produce a localized pressure field in the audible frequency range on targets intercepting the beam. The stress field produced on the target is highly localized, and its frequency spectrum includes components at the pulse repetition frequency and its harmonics, which are usually in the audible range of human hearing. Experimental results showed that the spectral characteristics of the sound measured in a water tank were in general agreement with the theoretical model. Because ultrasound scanners can produce high-intensity sound on bony structures in the human body, such scanners should not be considered to be passive imaging tools for fetal assessment.

ACKNOWLEDGMENTS

The authors are grateful to Dr. Paul L. Ogburn for his insightful comments throughout this study, Randall R. Kinnick for laboratory support, and Elaine C. Quarve for help with the manuscript.

[1]Sound-pressure level is defined as 10 times the logarithm to base 10 of the ratio of the time-mean-square pressure of a sound, to the square of the reference sound pressure. In gases, the reference pressure is normally 20 μPa. For sound in media other than gases, reference sound pressure is normally 1 μPa [refer to American National Standard acoustic terminology, Acoustical Society of America, New York, 1994]. Although the propagation medium considered in this paper is either water or soft tissue, we use 20 μPa as the reference pressure because this choice allows us to compare the ultrasound-generated sound level to the airborne sound in conventional human hearing. For example, 1 μPa would be equivalent to 120 dB relative to 1 μPa and 94 dB relative to 20 μPa.

[2]Even though soft tissue attenuates the incident ultrasound, the effect on the UGS generated on the target may not be very significant. For example, for a target at 30-mm depth from the skin, and considering a typical value of 0.7 dB/cm/MHz for tissue attenuation coefficient, the resulting attenuation for a 2-MHz ultrasound beam is 4.2 dB, which translate into 8.4-dB reduction in UGS pressure level.

Arulkumaran, S., Talbert, D. G., Talbert, D. G., Westgren, M., Su, H. S., and Ratnam, S. (**1991**). "Audible *in-utero* sound from ultrasound scanner," Lancet **338**, 704–705.

Arulkumaran, S., Talbert, D. G., Nyman, M., Westgren, M., Su, H. S., and Su, H. S., and Ratnam, S. (**1996**). "Audible *in utero* sound caused by the ultrasonic radiation force from a real-time scanner," J. Obstet. Gynaecol. Res. **22**(6), 523–527.

Berkowitz, R. L., Chitkara, U., Goldberg, J. D., Wilkins, I., Chervenak, F. A., and Lynch, L. (**1986**). "Intrauterine intravascular transfusions for severe red blood cell isoimmunization: Ultrasound-guided percutaneous approach," Am. J. Obstet. Gynecol. **155**(3), 574–581.

Beyer, R. T. (**1978**). "Radiation pressure—the history of a mislabeled tensor," J. Acoust. Soc. Am. **63**(4), 1025–1030.

Birnholz, J. C., and Benacerraf, B. R. (**1983**). "The development of human fetal hearing," Science **222**(4623), 516–518.

de Crespingy, L. C., Robinson, H. P., Quinn, M., Doyle, L., Ross, A., and Cauchi, M. (**1985**). "Ultrasound-guided fetal blood transfusion for severe rhesus isoimmunization," Obstet. Gynecol. (N.Y., NY, U. S.) **74**(6), 529–532.

Enderle, J. D., Blanchard, S. M., and Bronzino, J. D. (**2000**). *Introduction for Biomedical Engineering* (Academic, San Diego).

Fatemi, M., Ogburn, Jr., P. L., and Greenleaf, J. F. (**2001**). "Fetal stimulation by pulsed diagnostic ultrasound," J. Ultrasound Med. **20**, 883–889.

Foster, K. R., and Wiederhold, M. L. (**1978**). "Auditory responses in cats produced by pulsed ultrasound," J. Acoust. Soc. Am. **63**, 1199–1205.

Gavrilov, L. R. (**1984a**). "Use of focused ultrasound for stimulation of nerve structures," Ultrasonics **22**, 132–138.

Gavrilov, L. R. (**1984b**). "Focused ultrasound as a stimulator of the nerve structures. Archies of acoustics," Arch. Acoust. **9**, 67–74.

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

Harris, G. R. (**1985**). "A discussion of procedures for ultrasonic intensity and power calculations from miniature hydrophone measurements," Ultrasound Med. Biol. **11**(6), 803–817.

Manning, F. A. (**1995**). "Dynamic ultrasound based fetal assessment," Clin. Obstet. Gynecol. **38**, 26–44.

McDicken, W. N. (**1991**). *Diagnostic Ultrasonics. Principles and Use of Instruments*, 3rd ed. (Churchill Livingstone, Edinburgh).

NCRP Report No. 140 (**2002**). *Exposure Criteria for Medical Diagnostic Ultrasound: II. Criteria Based on All Known Mechanisms* (National Council on Radiation Protection and Measurements, Bethesda, MD).

Querleu, D., Renard, X., and Crepin, G. (**1981**). ''Perception auditive et reactivite foetale aux stimulations sonores'' (Auditory perception and fetal reaction to react to sound stimulation''), J. Gynecol. Obstet. Biol. Reprod. (Paris) **10**(4), 307–314.

Saeian, K., Weintraub, R., Hagen-Ansert, S., Sahn, C., Shiota, T., Kenny, A., and Sahn, D. J. (**1995**). ''Increased fetal activity and heart rate during and immediately after pulsed Doppler echocardiography,'' Echocardiography **12**, 71–77.

Torr, G. R. (**1984**). ''The acoustic radiation force,'' Am. J. Phys. **52**(5), 402–408.

Tsirulnikov, E. M., Vartanyan, I. A., Gersuni, G. V., Rosenblyum, A. S., Pudov, V. I., and Gavirolov, L. R. (**1988**). ''Use of amplitude-modulated focused ultrasound for diagnosis of hearing disorders,'' Ultrasound Med. Biol. **14**, 277–285.

Westervelt, P. J. (**1951**). ''The theory of steady force caused by sound waves,'' J. Acoust. Soc. Am. **23**(4), 312–315.

Whistles of tucuxi dolphins (*Sotalia fluviatilis*) in Brazil: Comparisons among populations

Alexandre F. Azevedo[a)]
PPGB/IBRAG, Departamento de Ecologia, Universidade do Estado do Rio de Janeiro. Rua São Francisco Xavier 524, Maracanã, Rio de Janeiro, 20550-013, RJ, and Projeto Mamíferos Aquáticos, Departamento de Oceanografia, UERJ

Monique Van Sluys
Departamento de Ecologia, Universidade do Estado do Rio de Janeiro, Rua São Francisco Xavier 524, Maracanã, Rio de Janeiro, 20550-013, RJ

(Received 29 June 2004; revised 27 December 2004; accepted 27 December 2004)

A quantitative and comparative analysis of the acoustic whistles parameters of tucuxi dolphins (*Sotalia fluviatilis*) from eight areas in Brazilian waters were conducted. Tucuxi produced mainly upsweep whistles, and despite their similar characteristics to other delphinids, whistles were shorter and less complex in shape. Some signals had fundamental frequency components above 24 kHz, and a broader band system may be useful in further studies concerning tucuxi whistles. The frequency parameters of the whistles had the lowest variation coefficients, whereas the duration and number of inflections had the highest variation coefficients. About 65% of all whistles had common characteristics between study localities, but SF, MinF, MeF, F1/4, F1/2, and F3/4 increased in frequency from southern to northern populations. The magnitude of the whistle characteristics variation was comparatively smaller between adjacent sites areas than between nonadjacent ones. However, for some pairwise comparisons the whistles were more similar between nonadjacent areas than between adjacent ones, indicating that other factors besides genetic and mixing affected tucuxi whistle characteristics. Only a few numbers of whistle variables were significantly different between adjacent areas and the parameter ranges of whistles overlapped, which made it difficult to discriminate whistles between adjacent tucuxi populations.
[DOI: 10.1121/1.1859232]

PACS numbers: 43.80.Ka Pages: 1456–1464

I. INTRODUCTION

Sounds produced by odontocetes have functions related to navigation and communication, and they are divided in two broad categories: pulsed and tonal sounds (Popper, 1980; Richardson *et al.*, 1998). Whistles are tonal sounds, which play an important underwater communication function in delphinid cetaceans (Tyack, 1998).

Although knowledge on whistles characteristics is an auxiliary tool for studies concerning cetaceans' behavior and ecology, only a few odontocete species have been studied in this respect (see Matthews *et al.*, 1999). Bioacoustical aspects are not well known in several species and a qualiquantitative description of whistles is rather scarce (Rendell *et al.*, 1999). Despite dolphin whistles having species-specific characteristics, intraspecific comparisons have shown that whistle variation seems to reflect the social relationship, the population structure, the environmental heterogeneity, and the behavioral state of the animals at time of recording (Steiner, 1981; Wang *et al.*, 1995a,b; Rendell *et al.*, 1999).

The bioacoustical aspects of tucuxi dolphin (*Sotalia fluviatilis*), including whistles, have received little attention. Two ecotypes are considered for this species, riverine (or freshwater) and marine, which differ in morphological aspects (Da Silva and Best, 1996). The riverine tucuxi is endemic to the Amazon River drainage, occurring in the main channel of rivers, tributaries, and lakes (Flores, 2002). The marine tucuxi has continuous distribution in western Atlantic coastal waters of South and Central Americas, from southern Brazil (27°35′S, 48°34′W) to Nicaragua (14°35′N, 83°14′W), with possible records in Honduras (Da Silva and Best, 1996; Flores, 2002). The marine tucuxi inhabits shallow waters and is often found in bays and estuaries (Da Silva and Best, 1996).

The first studies regarding tucuxi sound emissions focused on freshwater *Sotalia* were performed with wild and captive animals. Bioacoustic studies in captivity focused on echolocation (Caldwell and Caldwell, 1970; Wiersma, 1982; Kamminga *et al.*, 1993), sound production (Alcuri and Busnel, 1989), and auditory capacity of this species (Sauerland and Dehnhardt, 1998). The first free-ranging tucuxi studies on whistles were carried out in the Amazon River basin. Norris *et al.* (1972) described whistles about 0.2 s long, with most frequencies between 10 and 15 kHz. Nakasai and Takemura (1975) reported whistle frequencies ranging from 5 to 16 kHz and duration between 0.1 and 0.5 s. Wang *et al.* (2001) noted a frequency range between 3.6 and 23.8 kHz and whistles that lasted 1.04 s. Podos *et al.* (2002) reported an average whistle duration of 0.63 s and an average frequency ranging from 9.7 to 15.6 kHz.

The sound emissions of the marine ecotype were first studied in the 1990s with data obtained from a few areas

a)Electronic mail: azevedo.alex@uol.com.br

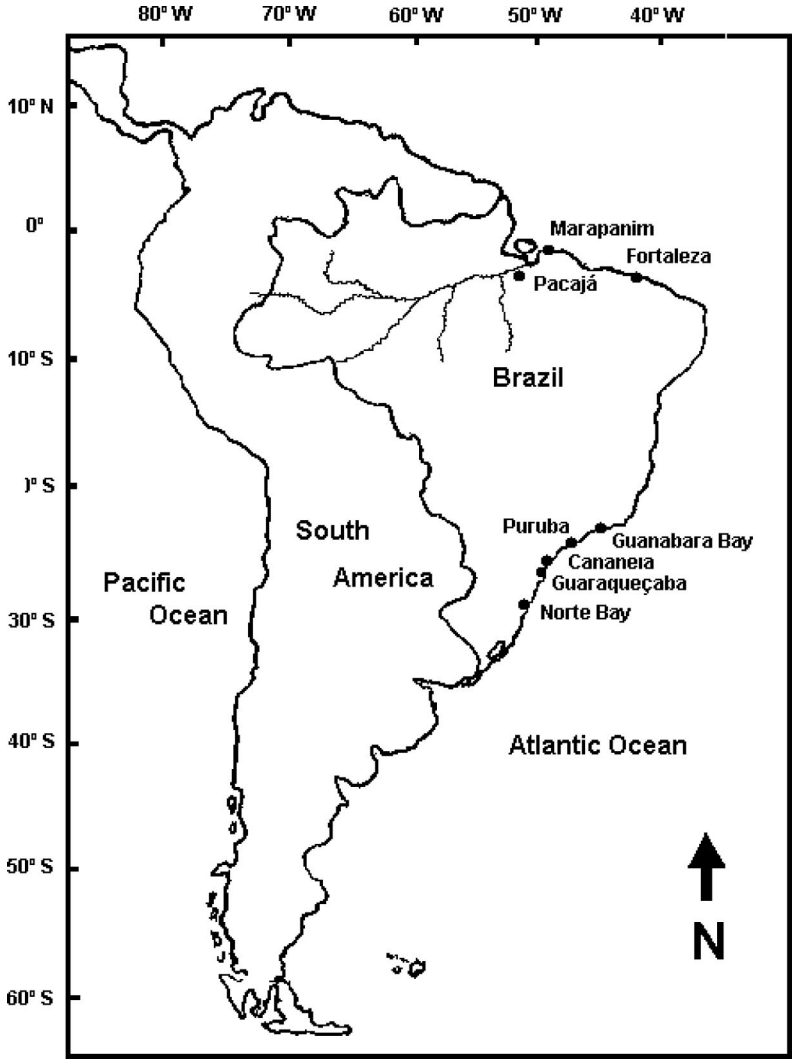


FIG. 1. Map of the Brazil denoting the eight sites where recordings of tucuxi dolphins whistles were made.

along the species' distribution, mainly in south and southeastern Brazil. Monteiro-Filho and Monteiro (2001) reported four different types of whistles with frequency components of up to 6 kHz for dolphins from Cananéia estuary, São Paulo State. Azevedo and Simão (2002) described a frequency range between 0.9 and 18.0 kHz and duration reaching 852 ms (102.5±81 ms) from marine tucuxis from Guanabara Bay, Rio de Janeiro State. Erber and Simão (2004) reported frequency range from 1.0 to 21.7 kHz and whistles that lasted 2.2 s for marine tucuxi from Sepetiba Bay, Rio de Janeiro State.

Despite the fact that a few studies had described acoustic parameter characteristics of tucuxi whistles, the intraspecific variability of the whistles are still unknown for *S. fluviatilis*. We conducted a study on tucuxi whistles in Brazil, in which we recorded the species along part of its distribution, in different environments and conditions. Here we report the whistle characteristics of wild tucuxis from eight areas in Brazilian waters and present a quantitative analysis of whistles parameters in order to examine for intraspecific variation.

II. METHODOLOGY

Between May 2002 and March 2004, acoustic recordings of underwater sound produced by marine tucuxi dolphins were made at eight different sites in Brazilian waters. The study areas are located in estuarine bays, mangroves sites, open sea, and rivers, from southern, southeastern, northern, and northeastern Brazil (Fig. 1, Table I).

All surveys were carried out under similar weather conditions (Beaufort sea states $\leqslant 2$), using small outboard-powered boats that ranged from 4 to 10 m in length. Acoustic recordings were made with the engine off and were monitored by headphones. When the dolphins' sounds became weak, we stopped recording and the boat was repositioned. The recording system consisted of a High Tech Inc. hydrophone (model HTI-96-MIN, frequency response: 5 Hz to 30 kHz ± 1.0 dB, -165 dB re: 1 V/μPa) and a digital audiotape recorder SONY TCD-D8 with on upper frequency limit of 24 kHz (sampling rate of 48 kHz).

Bioacoustical studies of cetaceans have sampling problems for representativeness and independence of data, leading to difficulties on statistical analyses (Matthews *et al.*, 1999; Rendell *et al.*, 1999). Trying to minimize these problems, in each study area the acoustic recordings were carried out across several individuals, groups, activities, and locations.

The DAT recordings of each group sampled were redigitized using Cool Edit Pro (Syntrillium Software) at a sampling rate of 48 kHz, 16-bit resolution. Whistles were defined

TABLE I. Location where tucuxi dolphins where recorded, noting the number of individuals sampled, total time of recording, days recorded, and activity of the dolphins.

Study area	Location	Individuals recorded[c]	Time recorded	Days	Activities
Norte Bay[a]	27°30′S, 48°30′W	70	4.5 h	4	Feeding, socializing and travel
Guaraqueçaba[a]	25°25′S, 48°15′W	80	5 h	3	Feeding, socializing and travel
Cananeia[a]	25°03′S, 47°55′W	180	10 h	6	Feeding, socializing and travel
Puruba[a]	23°23′S, 45°56′W	30	1.5 h	1	Feeding and travel
Guanabara Bay[a]	22°57′S, 43°10′W	70	6 h	10	Feeding, socializing and travel
Fortaleza[b]	03°42′S, 38°31′W	50	4 h	3	Feeding and travel
Marapanim[b]	00°43′S, 47°42′W	70	6 h	3	Feeding, socializing and travel
Pacaja River[b]	01°57′S, 50°44′W	50	3.5 h	3	Feeding, socializing and travel

[a]Southern Brazil.
[b]Northern Brazil.
[c]Estimated number of different individuals recorded.

as continuous, narrow-band sound emissions with or without harmonics (Popper, 1980). To avoid oversampling individuals and groups, due to repeated, individual-specific, or mimicked whistles (Rendell *et al.*, 1999), we analyzed whistles randomly from each group. We decided that the maximum number of whistles analyzed from each group would be of twice the group size, and the minimum interval between two selected whistles would be of one minute. If individual groups were oversampled, it could lead to the nonindependence of data (Matthews *et al.*, 1999).

The selected whistles were analyzed using the program Raven 1.1 (Cornell Laboratory of Ornithology, New York) with a FFT size of 1024 points, an overlap of 50%, and using a Hamming window. We only used whistles for which all parameters of a spectral contour were distinctly measurable. Whistles cut off by the upper frequency limit of the recording system were quantified but were not included in the acoustic analysis.

For each whistle its contour was determined by visual analyses of the frequency modulation. Despite the fact that whistles characterization is helpful in comparing studies and getting an overall view of the whistles' repertories, classifying dolphin whistles into categories is limited by the graded nature of dolphin whistles and by measurement bias by the observer (Bazúa-Duran and Au, 2002). Trying to minimize these problems, we used broad categories for classify whistle contours: ascending (whistles rising in frequency and no one inflection point), descending (whistles falling in frequency and no one inflection point), ascending–descending (initial rising in frequency, one inflection point, then falling in frequency), descending–ascending (initial falling in frequency, one inflection point, then rising in frequency), constant (whistles which the frequency changes 1000 Hz or less during more than 90% of duration), and others (Fig. 2).

The frequency of whistles in each category was compared between the study areas using a chi-square test

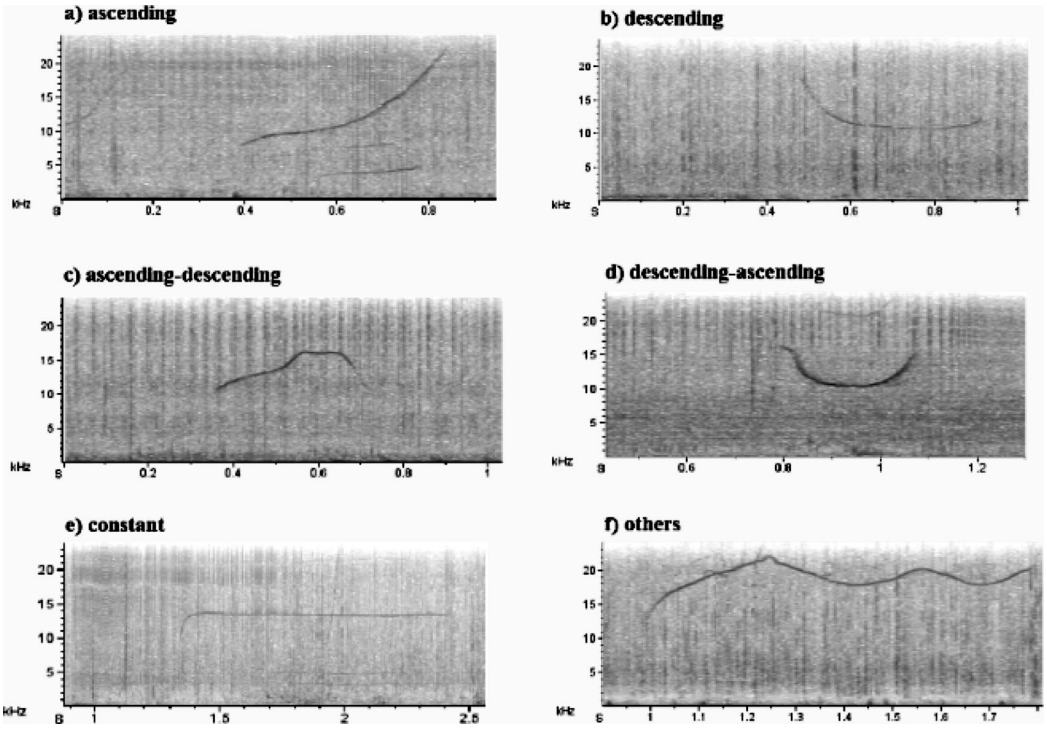


FIG. 2. The six classes to which tucuxi whistles contours were categorized. *X* axis=time (s), *Y* axis=frequency (kHz).

TABLE II. Number of tucuxi whistles for each of six categories in each study area.

Study area	Whistle contour category						
	Ascendant	Descendant	Ascendant–descendant	Descendant–ascendant	Constant	Others	All
Norte Bay	98	1	3	3	0	12	117
Guaraqueçaba	63	0	0	0	0	2	65
Cananeia	123	1	9	2	10	5	150
Puruba	35	0	5	3	4	5	52
Guanabara Bay	50	0	2	8	0	5	65
Fortaleza	68	2	6	7	5	3	91
Marapanim	54	4	10	8	7	12	95
Pacaja River	36	0	8	5	1	11	61
Percentage	75.7	1.1	6.2	5.2	3.9	7.9	
Total	527	8	43	36	27	55	696

(X^2; $P<0.01$). The larger the X^2 value represents the greater difference between sites (Bazúa-Duran, 2004).

Ten acoustic parameters from the fundamental component of each whistle were also measured: starting frequency (SF), ending frequency (EF), minimum frequency (MinF), maximum frequency (MaxF), frequency range (FR) [MaxF—MinF], duration (DUR), frequency at 1/4 of duration (F1/4), frequency at 1/2 of duration (F1/2), frequency at 3/4 of duration (F3/4), and number of inflection points (defined as points where the whistle contour changed from ascending to descending, or *vice versa*). We calculated a mean frequency (MeF) as the average of SF, EF, MinF, and MaxF. The frequency variables were measured in kHz and the duration in milliseconds. These whistle parameters were chosen to be consistent with previous studies of other odontocete species (e.g., Rendell *et al.*, 1999) and of *Sotalia* (e.g., Azevedo, 2000; Podos *et al.*, 2002; Erber and Simão, 2004). The descriptive statistics for all measured whistle variables includes the minimum values, maximum values, means, standard deviation, and coefficient of variation.

Acoustic parameters of whistles, such as frequency and duration components and number of inflections points, have been used for characterization of whistles of delphinid species, which also allows comparisons of whistles among groups and populations (Wang *et al.* 1995a,b; Rendell *et al.* 1999; Bazúa-Duran and Au, 2004). The paired-sample *t* test (Zar, 1999) was applied to evaluate if the end frequency of all whistles analyzed was significantly different from the start frequency. To investigate variations in whistle variables across sites using discriminant function analysis (Zar, 1999). To investigate differences among sites, a Tukey multiple comparisons test with unequal sample sizes (Zar, 1999) was performed to identify significant differences in each pair of whistle variables between the study areas.

III. RESULTS

A. General whistle characteristics

A total of 829 whistles emitted by tucuxi in Brazilian waters were selected. In all areas studied whistles cut off by the upper frequency limit of the recording system occurred, totaling 16% ($N=133$) of all selected whistles. These whistles were not included in the acoustic analyses, which left 696 whistles.

Ascending whistles corresponded to 75.7% of all whistles and they were the most frequent in all areas (Table II). Whistles with up to eight inflection points were found, but those with 0 and 1 inflection point corresponded to 92.5% of all whistles.

The whistle duration was short (308±137 ms) and 78.0% of the whistles lasted <400 ms. The average minimum frequency was 9.22±3.44 kHz and 78.4% of the whistles had MiF between 4.1 and 12.0 kHz. The maximum frequencies averaged 19.05±2.97 kHz and values between 16.01 and 22.00 kHz corresponded to 70.1% of the total. The frequency range of whistles averaged 9.83±4.03 kHz. The average mean frequency was 14.16±2.51 kHz and 74.3% of the whistles had MeFs ranging from 12.01 to 18.00 kHz. Descriptive statistics of all measured parameters of whistles are shown in Table III. The parameters of whistle frequency had the lowest variation coefficients. At all areas, the duration and number of inflection points had the highest variation coefficients, showing a high variability in these whistles' acoustic parameters.

The end frequency (mean 18.82±3.10 kHz) of the tucuxi whistles was, on average, higher than the start frequency (mean 9.57±3.76 kHz) (paired-sample *t* test, $t=52.72$, df=695, $P<0.001$).

B. Comparison among areas

In order to simplify, Norte Bay, Guaraqueçaba, Cananéia, Puruba, and Guanabara Bay were named as Southern Brazil and Fortaleza, Marapanim and Pacaja River as Northern Brazil. In all areas studied tucuxis produced preferably ascending whistles (Table II) and the end frequency was significantly higher than start frequency (paired-sample *t* test, $P<0.001$).

The whistles recorded in Fortaleza, Marapanim and Pacaja River had greater mean values of frequency parameters than areas located in southern Brazil (Table III). The frequency range was highest in South Brazil than in the northern areas (Table III). Pooled means of SF, MinF, MeF, F1/4, F1/2, and F3/4 from southern areas were about 25%, 29%, 11%, 21%, 21%, and 15%, respectively, lower than northern sites (*t* test; df=297; $P<0.001$). FR was about 24% lower in north than in south areas (*t* test; df=297; $P<0.001$). EF and MaxF were lower in southern localities (about 2% and 3%, respectively) than in northern ones, but

TABLE III. Descriptive statistics for acoustic parameters of tucuxi dolphin's whistles in Brazil, for each study area. Values are minimum–maximum, mean (s.d.) and variation coefficient. The frequency variables were measured in kHz and the duration in ms.

	Southern Brazil					Northern Brazil			
Acoustic parameters	Norte Bay ($N=117$)	Guaraqueçaba ($N=65$)	Cananéia ($N=150$)	Puruba ($N=52$)	Guanabara Bay ($N=65$)	Fortaleza ($N=91$)	Marapanim ($N=95$)	Pacaja River ($N=61$)	All whistles ($N=696$)
Start frequency	2.94–16.96 7.58 (2.80) 36.9%	4.61–13.79 8.25 (1.87) 26.7%	2.60–21.93 8.68 (3.52) 40.6%	4.48–16.51 9.58 (3.15) 32.9%	1.34–17.19 7.56 (2.84) 37.6%	2.44–21.82 12.60 (4.09) 32.5%	5.20–20.59 11.61 (3.39) 29.2%	5.19–20.62 11.36 (3.81) 33.5%	1.34–21.93 9.57 (3.76) 39.3%
End frequency	10.09–23.53 18.87 (3.17) 16.8%	13.28–23.01 18.65 (2.61) 14.0%	9.23–23.02 18.12 (3.39) 18.7%	10.40–22.98 18.75 (3.27) 17.4%	10.38–23.75 19.42 (3.13) 16.1%	12.62–23.64 18.99 (2.70) 14.2%	10.52–23.48 18.90 (3.36) 17.8%	12.30–23.07 19.63 (2.41) 20.7%	9.23–23.75 18.82 (3.10) 16.5%
Minimum frequency	2.94–16.84 7.39 (2.48) 33.6%	4.63–13.78 8.22 (1.91) 23.2%	2.60–18.34 8.43 (3.22) 38.2%	4.48–16.51 8.87 (2.52) 28.4%	1.34–17.19 7.21 (2.62) 36.3%	2.44–20.30 12.37 (3.76) 30.4%	5.20–20.01 10.85 (3.03) 27.9%	5.19–18.88 10.93 (3.39) 31.0%	1.34–20.3 9.22 (3.44) 37.3%
Maximum frequency	13.09–23.53 19.04 (2.85) 15.0%	13.28–23.01 18.68 (2.62) 14.0%	9.23–23.02 18.40 (3.28) 17.8%	10.40–23.00 18.88 (3.17) 16.8%	11.40–23.75 19.50 (2.96) 15.2%	12.62–23.84 19.24 (2.74) 14.2%	10.52–23.89 19.38 (3.05) 15.7%	12.30–23.87 19.95 (2.43) 12.2	9.23–23.89 19.05 (2.97) 15.6%
Mean frequency	8.55–18.70 13.22 (1.90) 14.4%	9.40–17.45 13.45 (1.65) 12.3%	6.70–20.10 13.41 (2.60) 19.4%	8.82–18.90 14.02 (1.88) 13.4%	8.75–18.90 13.42 (2.02) 15.1%	9.15–21.48 15.80 (2.75) 17.4%	7.95–20.68 15.19 (2.53) 16.6%	10.75–21.08 15.47 (2.41) 15.6%	6.70–21.48 14.16 (2.51) 17.7%
Frequency range	2.95–19.52 11.65 (3.80) 32.6%	2.53–17.67 10.44 (3.16) 30.3%	1.0–17.8 9.97 (3.91) 39.2%	1.5–16.1 10.01 (4.18) 41.8%	3.2–22.20 12.29 (3.90) 31.7%	0.21–18.0 6.87 (3.62) 52.7%	1.0–19.5 8.52 (3.39) 39.8%	1.9–15.2 9.02 (3.59) 39.8%	0.21–22.20 9.83 (4.03) 41.0%
Frequency at 1/4 of duration	4.28–18.58 10.01 (2.50) 25.0%	6.3–14.5 10.24 (1.70) 16.6%	5.4–18.8 11.13 (3.33) 29.9%	5.7–20.4 11.51 (3.09) 26.8%	3.9–17.9 9.40 (3.16) 33.6%	6.4–21.4 15.01 (3.26) 21.7%	7.0–21.5 13.10 (3.16) 24.1%	6.9–21.7 13.74 (3.59) 26.1%	3.9–21.7 11.73 (3.53) 30.1%
Frequency at 1/2 of duration	8.14–19.42 12.79 (2.67) 20.9%	7.29–18.49 12.94 (2.30) 17.8%	6.2–21.3 13.01 (3.59) 27.6%	6.9–20.6 13.03 (3.08) 23.6%	5.7–21.4 11.75 (3.61) 30.7%	9.2–23.2 16.51 (3.12) 18.9%	9.1–23.2 15.05 (3.66) 24.3%	8.0–23.4 15.99 (3.39) 21.2%	5.7–23.4 13.85 (3.58) 25.8%
Frequency at 1/2 of duration	8.63–21.37 15.25 (2.78) 18.2%	10.04–21.53 15.92 (2.70) 17.0%	8.10–22.73 15.14 (3.53) 23.3%	7.6–20.3 14.99 (3.28) 21.9%	7.4–23.6 15.22 (3.99) 26.2	12.1–23.4 17.62 (3.06) 17.4%	9.0–23.2 16.86 (3.70) 21.9%	10.3–23.4 17.47 (3.06) 17.5%	7.4–23.6 15.99 (3.43) 21.5%
Duration	90–620 291.1 (108.6) 37.3%	86–578 300.3 (119.8) 39.9%	80–810 318.4 (120.6) 37.9%	101–1064 325.7 (177.2) 54.4%	54–1015 300.3 (170.0) 56.6%	38–520 250.2 (107.1) 42.8%	64–774 321.4 (133.1) 41.4%	67–814 381.0 (172.1) 45.2%	38–1064 308.0 (137.3) 44.6%
Number of inflection points	0–4 0.46 (1.22) 265.2%	0–3 0.12 (0.55) 458.3%	0–5 0.17 (0.61) 358.8%	0–8 0.77 (2.00) 259.7%	0–4 0.38 (0.95) 250.0%	0–4 0.22 (0.49) 222.7%	0–5 0.61 (1.21) 198.4%	0–3 0.44 (0.74) 168.2%	0–8 0.37 (1.02) 275.7%

no significant differences were found (t test; df=297; $P>0.05$). The duration of whistles from southern areas was about 5% longer than northern sites, but not significantly different (t test; df=297; $P>0.05$). The distribution of the number of whistles cut off by the upper frequency limit of the recording system was similar between southern and northern localities (*Mann–Whitney* test, $U_{3,5}=6$; $P=0.65$).

The discriminant function analysis performed on whistle variables of tucuxis indicated a significant difference among the whistles from the eight areas [$Wilks\ Lambda=0.5598$; F (63,3835)=6.07; $P<0.0001$]. Almost all 11 whistle variables contributed for the discriminant function, only MinF and MaxF did not enter in the model, due to the strong correlation with SF and EF. The first two canonical roots accounted for 82% of the discrimination (Fig. 3). FR dominated the first canonical variable positively. The second canonical variable was dominated positively by SF and negatively by MeF. The *Mahalanobis* distances (D^2) values are shown in Table IV. In general, adjacent areas have the smallest distances, indicating that the whistles are more similar among sites geographically closer. The average D^2 was 0.823 between south sites and 0.910 between north areas. Comparing D^2 between south and north sites, the mean distance was 2.387. These D^2 values were significantly different for a comparison between adjacent and nonadjacent areas (*Mann–Whitney* test, $U_{13,15}=16$; $P<0.001$). However, for some comparisons the D^2 values were lower between nonadjacent areas than between adjacent areas (Table IV). The

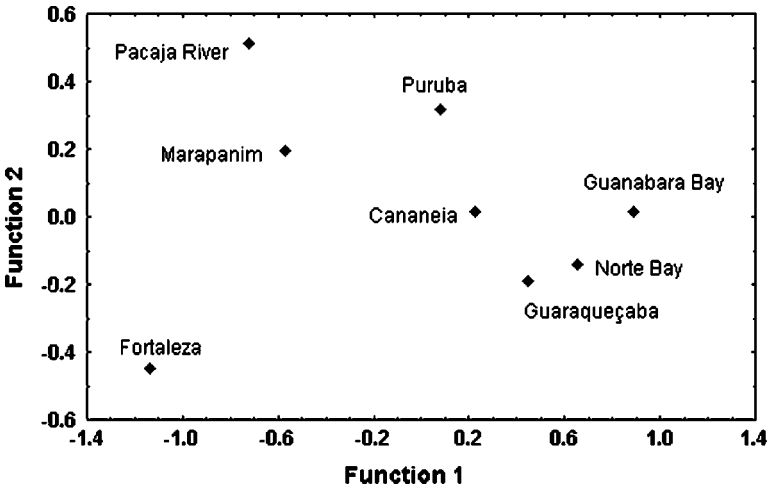


FIG. 3. Scatterplot of the canonical scores centroids for whistles emitted by tucuxi dolphins in each of the study areas. These two first functions combined account for 82% of the observed variance.

percentual correct of classification was 0.3491, indicating that approximately 65% of all whistles had similar characteristics between study areas.

Tukey multiple comparisons test between each pair of areas studied showed significant differences ($P<0.01$) in the whistles for almost all variables between southern and northern Brazil, whereas only a few variables were significantly different for comparisons between adjacent regions (Table V).

The comparisons of the whistle contour distribution of the six categories among each pair of site (Table VI) showed different results from the comparison of whistle acoustic parameters. For some sites X^2 values did not follow *Mahalanobis* distances and differences pointed out by the Tukey test.

IV. DISCUSSION

A. General whistle characteristics

We found higher averages of EF and MaxF than those previously reported (Table VII). These means had significant differences (two-sided *t* test, $P<0.01$) between this study and the values previously reported. For the riverine ecotype, Wang *et al.* (1995a) and Podos *et al.*, (2002) reported the mean of EF and MaxF 21%–31% lower than in this study. For marine tucuxi in southeastern Brazil, these differences were greater and the averages of EF and MaxF were about 45% higher than the values reported by Azevedo and Simão (2002) and Erber and Simão (2004). These differences may be related to the narrower frequency limit of the recording system used in some of those studies, but also may reflect differences in whistle characteristics. In addition, our results indicate that marine tucuxi produces whistles, on average, 54%–200% longer than previously reported in the Cananéia estuary (Monteiro-Filho and Monteiro, 2001) and Guanabara Bay (Azevedo and Simão, 2002) (Table VII; two-sided *t* test, $P<0.01$). These differences may be due to the upper limit of the recording systems used in previous studies, which probably cut off several whistles analyzed. Yet, when compared with the means of duration known for the riverine ecotype (Wang *et al.*, 2001; Podos *et al.*, 2002) we found lower values. We recorded riverine dolphins only in one area, not far from the marine environment (Pacajá River), for which the average duration of whistles was the longest of all study areas. However, the mean duration of whistles at Pacajá River was still significantly lower than those reported for tucuxis from isolated populations in the middle of the Amazon Basin (Wang *et al.*, 2001; Podos *et al.*, 2002).

Tucuxi dolphins had a varied repertoire of whistles, but these signals were simple in form and whistles with 0 and 1 inflection point were more abundant. Tucuxi dolphins produced a large number of whistles with rising frequency, which has already been noticed in previous studies (Norris *et al.*, 1972; Wang *et al.* 2001; Azevedo and Simão, 2002). At all sites studied here, the end frequency was, on average, higher than the start frequency. Upsweep whistles production is a species characteristic that has also been noticed in other delphinids (Matthews *et al.*, 1999). For instance, the mean end frequency is always higher than the start frequency for spinner dolphin whistles recorded in several places of the Pacific Ocean (Bazúa-Duran, 2004). However, the relation between the start and end frequencies of whistles may vary among populations of bottlenose dolphins (Wang *et al.*

TABLE IV. *Mahalanobis* distances (D^2) for comparisons of tucuxi whistles between each study area.

Study area	Norte Bay	Guaraqueçaba	Cananeia	Puruba	Guanabara Bay	Fortaleza	Marapanim
Norte Bay	···						
Guaraqueçaba	0.554	···					
Cananeia	0.658	0.506	···				
Puruba	0.907	1.266	0.632	···			
Guanabara Bay	0.952	0.664	0.998	1.097	···		
Fortaleza	3.695	2.933	2.221	2.568	4.705	···	
Marapanim	1.852	1.421	1.039	0.916	3.358	0.990	···
Pacaja River	2.658	2.094	1.443	1.519	3.382	1.252	0.489

TABLE V. Results of the Tukey multiple comparisons test with unequal sample sizes ($p<0.01$) for whistle variables among the study areas. Each study site has a code (numbers and letters), and for each site is shown whose other sites were significantly different in each of the acoustic parameters. No significant differences are represented by (···).

	Southern Brazil					Northern Brazil		
Area code	Norte Bay	Guaraqueçaba	Cananéia	Puruba	Guanabara Bay	Fortaleza	Marapanim	Pacajá River
Acoustic parameters	(1)	(2)	(3)	(4)	(5)	(A)	(B)	(C)
Start frequency	A, B, C	A, B, C	A, B, C	A, B, C	A, B, C	1, 2, 3, 4, 5	1, 2, 3, 4, 5	1, 2, 3, 4, 5
End frequency	···	···	···	···	···	···	···	···
Minimum frequency	A, B, C	A, B, C	A, B, C	A, B, C	A, B, C	1, 2, 3, 4, 5	1, 2, 3, 4, 5	1, 2, 3, 4, 5
Maximum frequency	···	···	···	···	···	···	···	···
Mean Frequency	A, B, C	A, B, C	A, B, C	A, C	A, B, C	1, 2, 3, 4, 5	1, 2, 3, 5	1, 2, 3, 4, 5
Frequency range	A, B, C	A	5, A	A	3, A, B, C	1, 2, 3, 4, 5	1, 5	1, 5
Frequency at 1/4 of duration	A, B, C	A, B, C	A, B, C	5, A, C	4, A, B, C	1, 2, 3, 4, 5, B	1, 2, 3, 5, A	1, 2, 3, 4, 5
Frequency at 1/2 of duration	A, B, C	A, B, C	A, B, C	A, B, C	A, B, C	1, 2, 3, 4, 5	1, 2, 3, 4, 5	1, 2, 3, 4, 5
Frequency at 3/4 of duration	A, C	···	A, B, C	A, C	A, C	1, 3, 4, 5	3	1, 3, 4, 5
Duration	C	···	···	···	···	B, C	A	1, A
Number of inflection points	···	···	···	···	···	···	···	···

1995b). Based on the relationship between start and end frequencies, Bazúa-Duran and Au (2002) suggest that *Sotalia fluviatilis, Delphinus delphis, Stenella attenuata, S. clymene, S. longirostris*, and *Lagenorhynchus albirostris* share some of the characteristics of their whistles and their repertoire and usage may be similar. Our findings are in agreement with the hypothesis of Bazúa-Duran and Au, once tucuxis produced a great number of upsweep whistles. But in most of those species the whistles contain a higher number of inflection points (see Matthews *et al.,* 1999).

Tucuxi is one of the smallest delphinids (Da Silva and Best, 1996) and the production of whistles with a higher frequency than other dolphins is consistent with the inverse relationship between the maximum fundamental frequency of whistles and body length, pointed out by Wang *et al.* (1995a,b) and Matthews *et al.* (1999). Of all selected whistles, 16% were cut off by the upper frequency limit of the recording system, indicating that tucuxi whistles can reach frequencies higher than 24 kHz, which had not been previously recorded. The maximum fundamental frequency recorded was 23.86 kHz for the riverine ecotype (Wang *et al.*, 2001) and 21.7 kHz for the marine ecotype (Erber and Simão, 2004). Odontocete whistles have most of their energy below 20 kHz, although energy may reach about 30 kHz. Fundamental frequencies above 24 kHz have been reported for only a few species (Richardson *et al.*, 1998; Matthews *et al.*, 1999). Wang *et al.* (1995a) reported whistle characteristics for seven odontocete species, and for only one (*Lagenorhynchus obscurus*) the authors reported frequencies higher than 24 kHz. For spinner dolphins in Hawaii, 2% of the whistles recorded with a broadband system (Au *et al.*, 1999) had fundamental whistle components above 24 kHz (Bazúa-Duran and Au, 2002). Our results suggest that tucuxis emit whistles with fundamental frequency components above 24 kHz more frequently than other odontocetes previously studied and a broader band system may be useful in further studies concerning *S. fluviatilis* whistles.

B. Comparison among areas

Some studies have shown geographic variation in intraspecific odontocete calls (e.g., Bazúa-Duran and Au, 2004). For tucuxis, the magnitude of the whistle characteristics variation was comparatively smaller between adjacent sites from southern and northern Brazil than between nonadjacent areas (average D^2 from adjacent areas<average D^2 from nonadjacent areas). Isolated populations have distinct adaptations to environmental conditions and have differences in taxonomy, which may have effects on whistle characteristics (Steiner, 1981; Wang *et al.*, 1995a; Rendell *et al.*, 1999). Wang *et al.* (1995b) found a similar pattern in whistle characteristics of bottlenose dolphin (*T. truncatus*) populations and suggested that each population might have been geographically isolated from the others, and thus developed their own acoustic characteristics. The closest tucuxi populations are not completely isolated and, consequently, genetic or behavioral processes of whistles divergence are less likely to operate. Northern populations are distant from southern ones

TABLE VI. Chi-square values (X^2) for comparisons of tucuxi whistle contours distribution between each study area. Italic numbers represent no significant differences ($P>0.01$).

Study area	Norte Bay	Guaraqueçaba	Cananeia	Puruba	Guanabara Bay	Fortaleza	Marapanim
Norte Bay	···						
Guaraqueçaba	*7.51*	···					
Cananeia	15.06	10.07	···				
Puruba	15.52	20.09	*8.35*	···			
Guanabara Bay	*7.69*	12.78	19.26	*8.87*	···		
Fortaleza	15.75	16.51	*7.74*	*4.58*	*8.21*	···	
Marapanim	25.57	32.35	24.66	*3.41*	14.25	*8.99*	···
Pacaja River	17.63	27.50	26.79	*4.57*	*9.70*	14.34	*5.93*

TABLE VII. Means and standard deviations of some whistle parameters of tucuxi from previous studies. The two-sided *t*-test (Zar, 1999) was performed to compare whistle parameters between the previous study and this study. Italic numbers represent no significant differences ($P > 0.01$).

	Studies				
Acoustic parameters	Wang *et al.* (1995a,b)	Monteiro-Filho and Monteiro (2001)	Azevedo and Simão (2002)	Podos *et al.* (2002)	Erber and Simão (2004)
Start frequency	10.76(3.53)	[a]	7.9(2.9)	*9.70(3.64)*	10.70(4.97)
End frequency	14.35(2.89)	[a]	12.8(4.5)	14.64(2.66)	13.31(5.86)
Minimum frequency	10.21(3.10)	[a]	7.6(2.9)	*9.18(3.05)*	10.52(4.52)
Maximum frequency	15.41(3.11)	[a]	13.0(4.1)	15.65(2.49)	13.31(4.85)
Frequency range	[a]	[a]	[a]	[a]	7.05(12.8)
Frequency at 1/4 of duration	[a]	[a]	[a]	[a]	11.11(4.72)
Frequency at 1/2 of duration	[a]	[a]	[a]	[a]	*13.66(6.19)*
Frequency at 3/4 of duration	[a]	[a]	[a]	[a]	15.37(6.44)
Duration	410 (210)	210(20)	102.5(81.0)	630 (320)	789 (3119)
N	155	94	5086	50	3350

[a]The frequency variables are given in kHz and the duration in ms.

and this may lead a to selective pressure for divergence in whistle characteristics. This pattern is somewhat in accordance with a preliminary study on the population genetic structure of this species, which presented evidence of divergence between south/southeastern and north/northeastern populations (Cunha, 2003). However, for some pairwise comparisons between nonadjacent localities, the D^2 values were smaller than between adjacent ones. For instance, the D^2 value was 0.916 between Puruba and Marapanim, and 1.266 between Puruba and Guaraqueçaba. This indicates that other factors besides genetic and geographic isolation are operating, since variations in odontocete calls also reflect their social relationship, environmental heterogeneity, and behavioral state (Wang *et al.*, 1995b; Rendell *et al.*, 1999).

Our results showed that differences in tucuxi whistles were small between adjacent regions. Only a few numbers of whistle variables were significantly different between adjacent areas and the parameter ranges of whistles overlapped, which made it difficult to discriminate whistles between adjacent tucuxi populations. For example, from the five areas from the south, only Guanabara Bay had a few whistle acoustic parameters significantly different from the other sites. The same pattern occurred in the northern sites, where also only a few whistle variables were significantly different among the three areas. Tucuxi movements among adjacent areas are poorly known. Photo-identification studies have been conducted only in Norte Bay, Guaraqueçaba, Cananeia and Guanabara Bay, but no comparisons among individuals catalogued were conducted. However, Cunha (2003) reports that populations in the southern and southeastern Brazilian coast share the same haplotype, which suggests gene flow between areas or recent isolation. For bottlenose dolphins, similarities between whistle characteristics may correspond to the degree of movement of individuals between areas (Wang *et al.*, 1995b). However, mixing is not the only factor to be considered in the geographic variation of dolphin whistles (Bazúa-Duran, 2004). Our findings agree with this author, since some pairwise comparisons showed that whistle similarities among adjacent areas did not follow the geographic distances. For instance, the D^2 value was 0.664 between Guanabara Bay and Guaraqueçaba, and 1.097 between Guanabara Bay and Puruba. Additionally, this same pattern was observed in comparisons of whistle contour distributions (Table VI) and in the first two canonical functions (Fig. 3).

The comparisons of whistle contour distribution of the six categories among each pair of site showed different results from comparison of whistle parameters. For instance, Fortaleza had the largest D^2 values from southern sites, but had small and not significant X^2 values for comparisons with Cananeia, Puruba, and Guanabara Bay. Additionally, for some sites where the Tukey test did not point out differences in whistle variables, the chi-square test found significant differences for whistle contour distribution. These results agree with Bazúa-Duran (2004), who also found differences when comparing whistle contour distributions and whistle acoustic parameters, for spinner dolphins. It reinforces the Bazúa-Duran supposition of that *Stenella longirostris* and *Sotalia fluviatilis* have a similar whistle system.

V. CONCLUSIONS

The frequency parameters of the whistles had the lowest variation coefficients, whereas the duration and number of inflections had the highest variation coefficients. This high intraspecific variability in both duration and number of inflection points may result from an individual modulation of these parameters so that information may be transmitted from different contexts or from different individuals (Steiner, 1981; Wang *et al.*, 1995b; Rendell *et al.*, 1999; Bazúa-Duran and Au, 2004).

Acoustic parameters of whistles, such as frequency and duration components and number of inflections points, have been used for the characterization of whistles of the delphinid species, which also allows comparisons of whistles among groups and populations (Wang *et al.*, 1995b; Rendell *et al.*, 1999; Bazúa-Duran and Au, 2004). When geographical whistle differences are well studied, they can help in population discrimination. Despite the fact that 65% of all whistles have common characteristics between study localities, the whistles from northern dolphins differed from those of the south, mainly due to their higher values of some frequency parameters. SF, MinF, MeF, F1/4, F1/2, and F3/4

increased in frequency from southern to northern populations. Biacoustical studies on tucuxi populations in the middle Brazilian coast, would confirm if this is a natural tendency in tucuxi whistles along the species distribution. Additionally, information regarding tucuxi sounds along the Amazon basin would also allow an evaluation of whistle differences between marine and freshwater tucuxis. The two ecotypes live in different habitats and ecology can exert a considerable influence on the evolution of sound communication (Van Parijs *et al.*, 2000). Freshwater and marine tucuxis are morphologically distinct, the latter being significantly larger-bodied (Monteiro-Filho *et al.*, 2002), and whistle characteristics also follow morphology (Matthews *et al.*, 1999). Recently, Monteiro-Filho *et al.* (2002) has suggested the separation of these ecotypes in two distinct species and whistle characteristic comparisons may be useful in this context.

ACKNOWLEDGMENTS

This study was funded by Idea Wild, The Society for Marine Mammalogy (Research Grants-In-Aid Program), Cetacean Society International, The Humane Society of the United States, PPGB (IBRAG, UERJ) and Instituto Biomas. We thank two anonymous reviewers who made useful suggestions for improving the manuscript. We also thank José Lailson Brito Jr. (Dept. Oceanografia, UERJ, Brazil), Haydée Cunha (IB, UFRJ, Brazil), Lilian S. Hoffman (IB, UFRGS, Brazil), and Davor Vrcibradic (Dept. Ecologia, UERJ, Brazil) provided useful comments on an earlier version of the manuscript. Paulo A. Flores and Mariel Bazzalo (Projeto Sotalia—Norte Bay, SC); Marcos César O. Santos (Projeto Atlantis—Guaraqueçaba e Cananeía, SP); Simone Castro and Alvaro Miranda (Projeto MAQUA—Guanabara Bay, RJ); Aline Cerqueira and Alexandra Costa (Fortaleza, CE); Renata Enim and Luana França (Projeto Botos—Marapanim e Pacajá River, PA) helped us in data collection. We particularly thank the Departamento de Ecologia (IBRAG, UERJ), Departamento de Oceanografia (UERJ), and Iate Club Jardim Guanabara, which provided logistical support. The *Conselho Nacional para o Desenvolvimento Científico e Tecnológico* (*CNPq*—Brazil) granted a graduate fellowship to A. F. Azevedo (Grant No. 140853/02-2). Monique Van Sluys had a research grant by *CNPq—Brazil* (Grant No. 302405/02-0).

Alcuri, G., and Busnel, R. G. (**1989**). ''Unicité functionelle des clics sonar et des sifflements dans les emissions acoustiques d'un Dauphin Amazonien *S. fluviatilis* Gervais et Deville,'' C. R. Acad. Sci. Paris. **308**, 379–384.

Au, W. W. L., Lammers, M. O., and Aubauer, R. (**1999**). ''A portable broadband data acquisition system for field studies in bioacoustics,'' Marine Mammal Sci. **15**, 526–531.

Azevedo, A. F. (**2000**). ''Emissões sonoras do boto-cinza (*Sotalia fluviatilis* Gervais, 1853) na Baía de Guanabara—RJ,'' Master thesis, Universidade Federal Rural do Rio de Janeiro, Seropédica, Brazil.

Azevedo, A. F., and Simão, S. M. (**2002**). ''Whistles produced by marine tucuxi dolphins (*Sotalia fluviatilis*) in Guanabara Bay, southeastern Brazil,'' Aquat. Mamm. **28**, 261–266.

Bazúa-Duran, M. C., and Au, W. W. L. (**2002**). ''Whistles of Hawaiian spinner dolphins,'' J. Acoust. Soc. Am. **112**, 3064–3072.

Bazúa-Duran, M. C. (**2004**). ''Differences in the characteristics and repertoire of bottlenose and spinner dolphins,'' Ann. Braz. Acad. Sci. **76**, 386–392.

Bazúa-Duran, M. C., and Au, W. W. L. (**2004**). ''Geographic variations in the whistles of spinner dolphins (*Stenella longirostris*) of the Main Hawaiian Islands,'' J. Acoust. Soc. Am. **116**, 3757–3769.

Caldwell, M. C., and Caldwell, K. D. (**1970**). ''Echolocation-type signals by two dolphins, genus *Sotalia*,'' Q. J. Fla. Acad. Sci. **33**, 124–131.

Cunha, H. A. (**2003**). ''Distribuição e diversidade das linhagens de DNA mitocondrial do ecótipo marinho de *Sotalia fluviatilis* (Cetacea: Delphinidae),'' Master's thesis, Universidade do Estado do Rio de Janeiro, Brazil.

Da Silva, V. M. F., and Best, R. C. (**1996**). ''Sotalia fluviatilis,'' Mammalian Species **527**, 1–7.

Erber, C., and Simão, S. M. (**2004**). ''Analysis of whistles produced by the tucuxi dolphin *Sotalia fluviatilis* from Sepetiba Bay, Brazil,'' Ann. Braz. Acad. Sci. **76**, 381–385.

Flores, P. A. C. (**2002**). ''Tucuxi *Sotalia fluviatilis*,'' in *Encyclopedia of Marine Mammals*, edited by W. F. Perrin, B. Wursig, and J. G. M. Thewissen (Academic Press, San Diego), pp. 1267–1269.

Kamminga, C., Van Hove, M. T., Engelsma, F. J., and Terry, R. P. (**1993**). ''Investigations on cetacean sonar X: A comparative analysis of underwater echolocation clicks of *Inia* spp. and *Sotalia* spp.,'' Aquat. Mamm. **19**, 31–43.

Matthews, J. N., Rendell, L. E., Gordon, J. C. D., and MacDonald, D. W. (**1999**). ''A review of frequency and time parameters of cetacean tonal calls,'' Bioacoustics **10**, 47–71. 21

Monteiro-Filho, E. L. A., and Monteiro, K. D. K. A. (**2001**). ''Low-frequency sounds emitted by *Sotalia fluviatilis guianensis* (Cetacea: Delphinidae) in an estuarine region in southeastern Brazil,'' Can. J. Zool. **79**, 59–66.

Monteiro-Filho, E. L. A., Monteiro, L. R., and Reis, S. F. (**2002**). ''Skull shape and size divergence in dolphins of the genus *Sotalia*: a tridimensional morphometric analysis,'' J. Mammal. **83**, 125–134.

Nakasai, K., and Takemura, A. (**1975**). ''Studies on the underwater sound VI: On the underwater calls of fresh dolphins in South America,'' Bull. Fac. Fish. Nagasaki. Univ. **40**, 7–13.

Norris, K. S., Harvey, G. W., Burzell, L. A., and Kartha, T. D. K. (**1972**). ''Sound production in the freshwater porpoises *Sotalia fluviatilis* and *Inia geoffrensis*, in the Rio Negro, Brazil,'' Investig. Cetacea **4**, 251–260.

Podos, J., da Silva, V. M. F., and Rossi-Santos, M. R. (**2002**). ''Vocalizations of Amazon river dolphins, *Inia geoffrensis*: insights into the Evolutionary origins of delphinid whistles,'' Ethology **108**, 601–612.

Popper, A. N. (**1980**). ''Sound emission and detection by delphinids,'' in *Cetacean Behaviour: Mechanisms and Functions*, edited by L. M. Herman (Wiley, New York), pp. 1–52.

Rendell, L. E., Matthews, J. N., Gill, A., Gordon, J. C. D., and MacDonald, D. W. (**1999**). ''Quantitative analysis of tonal calls from five odontocete species, examining interspecific and intraspecific variation,'' J. Zool. **249**, 403–410.

Richardson, W. J., Greene, C. R. J., Malme, C. I., and Thomsom, D. H. (**1998**). *Marine Mammals and Noise* (Academic, New York).

Sauerland, M., and Dehnhardt, G. (**1998**). ''Underwater audiogram of a tucuxi (*Sotalia fluviatilis guianensis*),'' J. Acoust. Soc. Am. **103**, 1199–1204.

Steiner, W. W. (**1981**). ''Species-specific differences in pure tonal whistle vocalizations of five western North Atlantic dolphin species,'' Behav. Ecol. Sociobiol. **9**, 241–246.

Tyack, P. L. (**1998**). ''Acoustic Communication under sea,'' in *Animal Acoustics Communication—Sound Analysis and Research Methods*, edited by S. L. Hopp, M. J. Owen, and C. S. Evans (Springer-Verlag, New York), pp. 163–219.

Van Parijs, S. M., Parra, G. J., and Corkeron, P. J. (**2000**). ''Sounds produced by Australian Irrawaddy dolphins, *Orcaella brevirostris*,'' J. Acoust. Soc. Am. **108**, 1938–1940.

Wang, D., Wursig, B., and Evans, W. E. (**1995a**). ''Comparisons of whistles among seven odontocete species,'' in *Sensory Systems of Aquatic Mammals*, edited by R. A. Kastelien, J. A. Thomas, and P. E. Nachtigal (De Spil, Woerden, The Netherlands), pp. 299–323.

Wang, D., Wursig, B., and Evans, W. E. (**1995b**). ''Whistles of bottlenose dolphins: comparisons among populations,'' Aquat. Mamm. **21**, 65–77.

Wang, D., Wursig, B., and Leatherwwod, S. (**2001**). ''Whistles of boto, *Inia geoffrensis*, and tucuxi, *Sotalia fluviatilis*,'' J. Acoust. Soc. Am. **109**, 407–414.

Wiersma, H. (**1982**). ''Investigations on Cetacea sonar IV. A comparation of wave shapes of odontocete sonar signals,'' Aquat. Mamm. **9**, 57–66.

Zar, J. H. (**1999**). *Biostatistical Analysis* (Prentice–Hall, Englewood Cliffs, NJ).

Underwater, low-frequency noise in a coastal sea turtle habitat

Y. Samuel[a)]
Department of Earth and Atmospheric Sciences, Ocean Resources and Ecosystems Program, Cornell University, Snee Hall, Ithaca, New York 14853

S. J. Morreale
Department of Natural Resources, Cornell University, Fernow Hall, Ithaca, New York 14853

C. W. Clark
Bioacoustics Research Program, Laboratory of Ornithology, Cornell University, Ithaca, New York 14853

C. H. Greene
Department of Earth and Atmospheric Sciences, Ocean Resources and Ecosystems Program, Cornell University, Snee Hall, Ithaca, New York 14853

M. E. Richmond
U.S.G.S., New York Cooperative Fish and Wildlife Research Unit, Department of Natural Resources, Cornell University, Fernow Hall, Ithaca, New York 14853

(Received 16 June 2004; revised 12 November 2004; accepted 22 November 2004)

Underwater sound was recorded in one of the major coastal foraging areas for juvenile sea turtles in the Peconic Bay Estuary system in Long Island, New York. The recording season of the underwater environment coincided with the sea turtle activity season in an inshore area where there is considerable boating and recreational activity, especially during the summer between Independence Day and Labor Day. Within the range of sea turtle hearing, average noise pressure reached 110 dB during periods of high human activity and diminished proportionally, down to 80 dB, with decreasing human presence. Therefore, during much of the season when sea turtles are actively foraging in New York waters, their coastal habitats are flooded with underwater noise. During the period of highest human activity, average noise pressures within the range of frequencies heard by sea turtles were greater by over two orders of magnitude (26 dB) than during the lowest period of human activity. Sea turtles undoubtedly are exposed to high levels of noise, most of which is anthropogenic. Results suggest that continued exposure to existing high levels of pervasive anthropogenic noise in vital sea turtle habitats and any increase in noise could affect sea turtle behavior and ecology. [DOI: 10.1121/1.1847993]

PACS numbers: 43.80.Nd, 43.30.Nb, 43.50.Rq [WWL] Pages: 1465–1472

I. INTRODUCTION

Underwater noise levels have increased dramatically in recent decades due to anthropogenic sources, such as commercial, industrial, and recreational maritime activities (Richardson *et al.*, 1995; Curtis *et al.*, 1999; Andrew *et al.*, 2002). Notably, a predominant component of sounds from these sources is in the low-frequency range (less than 1000 Hz). Currently, low-frequency noise is of great concern for sea turtles because their hearing range of highest sensitivity is confined to low frequencies (Ridgway *et al.*, 1969; Bartol *et al.* 1999), and they are endangered or threatened worldwide (Plotkin, 1995). Furthermore, sea turtles often aggregate in coastal areas where human activity, and therefore anthropogenic disturbance and underwater noise, is greatly heightened (Morreale and Standora, 1998). Our study focused on quantifying noise in a sea turtle foraging habitat, and identifying the anthropogenic component of underwater noise.

Many previous studies have linked anthropogenic noise to effects on the natural ecology of marine organisms. While some of these marine organisms have a different hearing sensitivity than sea turtles, and while some of the noise exposures are to different kinds of anthropogenic noise that are relevant to sea turtles, we present these studies here to demonstrate how different sources of anthropogenic noise have been seen to affect marine animals. Among the higher vertebrates, it has been shown that whales exhibit strong avoidance reactions to the sounds from oil and gas exploration activities (Malme *et al.*, 1983) and to seismic exploration noises (Richardson *et al.*, 1986; Richardson, 1997). In addition, beluga whales (*Delphinapterus leucas*) have been observed to decrease their call rates in response to boats moving closer, presumably responding to the associated noise (Lesage *et al.*, 1999). Killer whales (*Orcinus orca*) off British Columbia, Canada, when approached by motorboats, adopted a less predictable pathway; female killer whales in particular responded by swimming faster and increasing the angle between successive dives (Williams *et al.*, 2002). Further, impact models of boat noises on killer whales in the same area were shown to interfere with killer whale communication, cause behavioral avoidance, and cause a temporary threshold shift (TTS) in hearing (Erbe, 2002). When exposed to low-frequency sonar broadcasts, humpback whales

[a)]Electronic mail: ys88@cornell.edu

(*Megaptera novaeangliae*) often responded by singing longer songs, possibly to compensate for acoustic interference (Miller *et al.*, 2000; Fristrup *et al.*, 2003).

Among fish, several species, including Pacific herring (*Clupea harengus*), cod (*Gadus morhua*), and capelin (*Mallosus villosus*), have been shown to react to noise stimuli by increasing their swimming speed (Olsen *et al.*, 1983), by swimming downward (Suzuki *et al.*, 1980), or by moving away from the sound sources (Blaxter *et al.*, 1981; Schwarz and Greer, 1984; Engås *et al.*, 1998). Sound can have physical effects, too, causing measurable damage to sensory cells in the ears of fishes (Hastings *et al.*, 1996; McCauley *et al.*, 2003).

Adverse affects of underwater noise have been demonstrated for invertebrates as well. When exposed to higher pressure levels of noise in an experimental setting, brown shrimp (*Crangon crangon*) exhibited increased aggression and higher mortality rates, as well as significant reductions in their food uptake, growth, and reproductive rates (Lagardère, 1982).

For sea turtles, several studies have identified their ability to perceive low-frequency sounds (Ridgway *et al.*, 1969; Lenhardt *et al.*, 1983, 1996; O'Hara 1990; Bartol *et al.*, 1999). Hearing range coincides with the predominant frequencies of anthropogenic noise, increasing the likelihood that sea turtles might experience negative effects from noise exposure. At present, sea turtles are known to sense low-frequency sound, however, little is known about the extent of noise exposure from anthropogenic sources in their natural habitats, or the potential impacts of increased anthropogenic noise exposure on sea turtle biology.

Much of the acoustic research on sea turtles has focused on studying sea turtle ear anatomy and auditory sensory capabilities. These studies clearly demonstrated that sea turtles are able to detect and respond to sounds, and that their hearing is limited to low frequencies, with the range of highest sensitivity between 200 and 700 Hz, and with a peak near 400 Hz (Ridgway *et al.*, 1969; Bartol *et al.*, 1999). A related study has shown that when presented with acoustic stimuli at 430 Hz and 1.5 dB re 1 μ, sea turtles placed in 50-gallon tanks respond with abrupt body movements, such as blinking, head retraction, and flipper movement, all of which were interpreted as startle responses (Lenhardt *et al.*, 1996). Similar responses were observed when sinusoidal stimuli at 250 and 500 Hz within the range of 55–59 dB (acceleration re 10^{-3} m/s^2) were directly transmitted to the head of sea turtles, which were submerged in 0.5 m of sea water in 1.44-m-diam tanks (Lenhardt *et al.*, 1983). In a separate study, higher level responses, such as changes in swimming patterns and orientation, were noted when sea turtles in a confined canal (300 m long, 45 m wide, and up to 10 m deep) were subjected to high-pressure air gun pulses (120 dB re 1 μbar at 1 m), suspended at 2-m depth and positioned 33 m inward from one side of the tank, with frequencies ranging from 25 to 750 Hz (O'Hara, 1990).

In the current study, we focused on a near-shore habitat in New York state where there is great overlap between sea turtle activity and human activities that generate high levels of low-frequency noise. Each year, juvenile sea turtles, mostly loggerhead sea turtles (*Caretta caretta*), kemp's ridley sea turtles (*Lepidochelys kempii*), and green sea turtles (*Chelonia mydas*), are found in New York waters and the Peconic Bay Estuary system between July and October (Morreale and Standora, 1998). These very productive waters provide important developmental habitat for sea turtles during their early life stages. However, because this region also hosts some of the highest human densities along the coast of the western North Atlantic, as well as shipping lanes and major ports, foraging young sea turtles can be exposed to high levels of human activities that pose a risk of potential disturbance. Furthermore, a major portion of the sea turtle coastal foraging season occurs during the summer months, when recreational boating activities are highest.

We report here the results of a seasonal study designed to record underwater sound and to determine the noise levels to which sea turtles are exposed in one of their major coastal foraging areas in the western North Atlantic. The primary objective was to characterize and quantify the noise levels from both ambient and introduced sounds in the underwater environment, especially sounds within the sea turtle hearing range. The secondary objective was to distinguish and quantify the anthropogenic component of the noise to which sea turtles are exposed in their foraging habitat. This study was complementary to a separate behavioral study (Samuel, 2004) conducted in experimental tanks, which was designed to detect behavioral responses from sea turtles exposed to sounds similar to the ones recorded in the natural environment.

II. METHODS

A. Overview

Research was conducted in the Peconic Bay Estuary system in Long Island, New York from 20 July to 27 September 2001, a period that coincides with the sea turtle activity season in inshore waters. Over the span of the sea turtle foraging season, sounds were recorded in the underwater environment in Southold Bay (41°03′N, 72°24′W), one of the estuary's major foraging areas for juvenile sea turtles. Simultaneous with sea turtle activity, the waters within the estuary system are the site of much boating and recreational activity during the summer period, starting around Independence Day (4 July) and ending on Labor Day (first Monday in September). However, by early September, after a final burst of human activity on the Labor Day holiday, recreational boating drops off abruptly. These factors combine for a natural experiment in which the sea turtle season can be separated into five different categories, each of which is characterized by a distinctly different level of human motor-boating activity. The presumption is that the five different levels of human boating activity generate different levels of underwater noise that can be measured and compared. The five levels, ranging from periods of highest to lowest human activity were (1) daytime summer weekends before Labor Day; (2) daytime summer weekdays before Labor Day; (3) daytime weekends after Labor Day; (4) daytime weekdays after Labor Day; and (5) nighttime after Labor Day.

TABLE I. Sampling design of the 14 different recording sessions conducted between 20 July 2001 and 27 September 2001. Recording sessions were conducted both before and after Labor Day (3 September 2001). Sessions were further divided into weekday and weekend recordings. In addition, there was one recording conducted during nighttime after Labor Day. All recordings were 1 h long. Time indicates the beginning of each 1-h recording session.

	Before Labor Day			After Labor Day		
	Date	Day	Time	Date	Day	Time
Weekday	20 July	Friday	13:05	3 September	Monday	15:05
	7 August	Tuesday	14:40	6 September	Thursday	16:15
	15 August	Wednesday	14:45	18 September	Tuesday	14:50
				26 September	Wednesday	12:45
Weekend	22 July	Sunday	16:30	9 September	Sunday	15:10
	18 August	Saturday	14:40	22 September	Saturday	15:25
	25 August	Saturday	15:10			
	1 September	Saturday	14:40			
Night				27 September	Thursday	19:55

B. Sampling design

Monitoring of underwater sound and collection of acoustical data occurred on 14 separate days, spanning all five levels of human activity. Seven recording sessions were conducted during the two summer months of high human activity before Labor Day, and seven sessions were conducted during periods of lower human activity, after the Labor Day holiday. Recording sessions for high and low human activity were further divided into daytime weekdays (three recordings before Labor Day and four after Labor Day), daytime weekends (four recordings before Labor Day and two after Labor Day), and one nighttime recording during the period of very low human activity after Labor Day (Table I). All recordings were conducted from a kayak, anchored in the same location in Southold Bay, 120 m from shore, in water with a mean depth of 5 m. From visual observations, watercrafts in the Bay never exceeded 9 m in length. During data collection, the wind force was mostly zero or one on the Beaufort scale, sometimes between one and three, and once approached five. No precipitation fell during any of the data collection periods. Using this sampling scheme, a total of 14 one-hour samples were collected, for all five different human activity levels, in the same location under similar environmental conditions that differed principally only in the amount of anthropogenic activities.

C. Acoustic recording

The underwater acoustic environment was monitored and measured using a calibrated Sippican Inc. hydrophone system with a sensitivity of -162 dBV at 1000 Hz re 1 μPa. Sounds were recorded with a Sony TCD-D8 Digital Audio Tape (DAT) Walkman, which sampled at a rate of 44.1 kHz. The lowest significant bit of the DAT recorder's ADC was 70 dB re 1 μPa. Acoustic data were recorded on BASF-DAT Master® 124-min digital tapes. The recording system's frequency response was flat (± 2 dB) in the 100–14 000-Hz frequency range, with a -10 dB per octave roll-off in the 15–100-Hz frequency band. During the monitoring sessions, the hydrophone was suspended in mid-water column from a kayak (approximately 2.5 m below the surface of the water); the cable was attached by surgical tubing to reduce flutter.

D. Acoustic data processing

Digital acoustic data were transferred from digital audio tapes (DATs) to a computer and low-pass filtered at 1200 Hz to include only sounds within the range of sea turtle hearing, using Cornell Raven software (Charif *et al.*, 2003). The filtered data were graphically and statistically analyzed, and spectrograms were created (FFT size: 512 samples, overlap: 0%, Window type: Hann) using Cornell Canary 1.2.4 software. Spectrograms of noises, typical for each of the five different activity levels, along with their power spectrum statistics of the 5th, 50th, and 95th percentiles are shown in Fig. 1. Graphical representation of spectra was performed using interactive data language (IDL) software.

Band levels of received sound pressure in the band of peak hearing range of sea turtles (200–700 Hz), for each of the five different human activity levels, were calculated in a direct computation from the received sound pressure spectral density levels (Table II). Statistical comparisons of noise spectra from the four different daytime levels of human activity were performed using the Mann-Whitney rank sum test in which noise levels in the 50–1000-Hz frequency band were considered to be significantly different (at $p \leqslant 0.05$) when they were separated by 5 dB or more.

III. RESULTS

There was a distinct increase in noise pressure within the low-frequency range of sea turtle hearing that corresponded to an increase of human boating activity in the bay ecosystem (Fig. 2). On weekends during the boating season, noise pressure levels were the highest, and, as boating activity decreased, noise levels systematically diminished. High human activity also was accompanied by an increase in the complexity of noises, as measured by the increase of signals across a broad range of frequencies [Figs. 2(b)–(e)].

Below 50 Hz, noise levels were dominated by cable fluttering for all five levels of human activity. No comparisons were made between the spectral patterns for frequencies be-

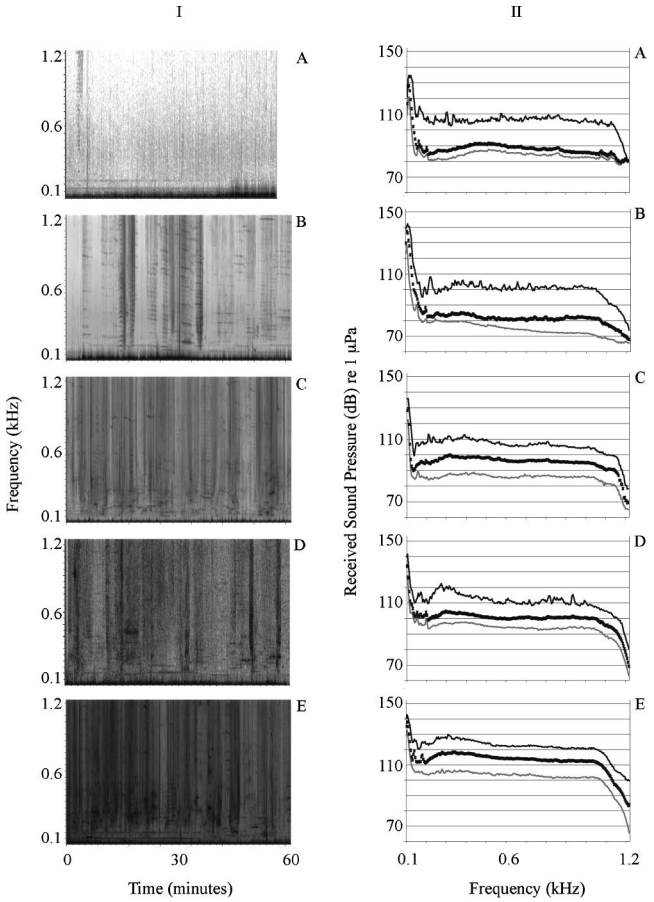


FIG. 1. (I) One-hour noise spectrograms from recordings during each of the five different human activity levels in Southold Bay. These spectrograms are considered to be typical for each activity level. (a) Lowest human activity—nighttime; (b) Low human activity—weekday after Labor Day; (c) Intermediate human activity—weekend after Labor Day; (d) High human activity—weekday before Labor Day; and (e) Highest human activity—weekend before Labor Day. Frequencies are between 0.1 and 1.2 kHz. All spectrograms were set to the same brightness and contrast. The darkness in the graphs represents pressure of sound, and shows not only how much noise is in the environment, but also how loud it is. (II) Power spectrum statistics for each of the one-hour noise spectrograms from recordings during each of the five different human activity levels in Southold Bay. The statistics show the 5th (solid light gray line—bottom line), 50th (thick black line—middle line), and 95th (solid black line—top line) percentiles of sound. Below each percentile line is where the corresponding percent of sound is found at those pressures and along the entire frequency range. (a) Lowest human activity—nighttime; (b) Low human activity—weekday after Labor Day; (c) Intermediate human activity—weekend after Labor Day; (d) High human activity—weekday before Labor Day; and (e) Highest human activity—weekend before Labor Day.

low 50 Hz. Above 50 Hz, the spectral patterns were dramatically different for each of the different levels of human activity (Fig. 2). At the lowest level of human activity, which occurred at night after Labor Day, the majority of the ambient noise occurred in the 50–200-Hz frequency band, with pressure levels below 75 dB *re* 1 μPa. At frequencies above 200 Hz, which include the peak hearing range for sea turtles, there was very little noise, with detectable signals (above 70 dB) in these frequencies comprising less than 5% of the total noise present at night [Fig. 2(a)]. Even the maximum noise pressures in these higher frequencies were moderate, rarely exceeding 90 dB [Fig. 2(a)].

At slightly higher levels of human activity, during both

TABLE II. Band levels of received sound pressure within the 200–700-Hz band during each of the five different human activity levels in Southold Bay. The received sound pressure spectral density levels have been integrated over the peak hearing range of sea turtles (200–700 Hz) in order to calculate the received sound pressure band levels.

Human activity level		Band levels of received sound pressure in the 200–700-Hz band (dB) *re* 1 μPa
Before Labor Day	Weekend	113
	Weekday	102
After Labor Day	Weekend	91
	Weekday	87
	Night	83

weekdays and weekends after Labor Day, there was an increase over nighttime levels in both the total amount of noise and the pressure levels at higher frequencies [Figs. 2(b) and (c)]. During these daytime recordings, pressure levels for sounds at all frequencies were greater than nighttime levels, with levels rising to as high as 100 dB. On weekdays after the Labor Day holiday, the noise was spread widely across the entire hearing range of sea turtles. The pressure levels were roughly equal at middle and higher frequencies, with an approximately uniform amount of noise with levels ranging from 70 to 80 dB [Fig. 2(b)]. On weekends after Labor Day, there was a further increase in the amount of noise present at pressures below 80 dB across all frequencies, and a much greater amount of noise that was louder than 80 dB [Fig. 2(c)]. Overall, in the period after Labor Day, there was nearly a 3-dB increase in the amount of noise levels from the nighttime to the weekday samples, and a further 4-dB increase in noise levels from weekdays to the weekend samples.

During the mid-summer levels of higher human activity, the bay ecosystem was flooded with noise across all frequencies. On weekdays, more than 50% of the sound at all frequencies was at pressures between 80 and 100 dB, with noise levels peaking up to 110 dB [Fig. 2(d)]. On weekends, when boating activity was at its peak, the highest levels of noise also occurred. Along with a shroud of noise across the entire range of sea turtle hearing, 50% of the noise levels were at pressures between 90 and 110 dB, with approximately 5% at levels greater than 120 dB [Fig. 2(e)].

Using the profile from the weekdays after Labor Day as a baseline reference for daytime noise levels, the mean difference in sound pressure spectral densities was calculated for the three other daytime activity levels (Fig. 3). This was used as a general index of anthropogenic noise. When sound pressures varied by more than 5 dB from the quiet baseline levels, the difference was considered to be significant. Using this measure, during two of the three daytime human activity levels, there was a significant difference ($p<0.05$) from the baseline [Figs. 3(b)–(e)]. At frequencies greater than 200 Hz, during levels of higher human activity, the mean sound pressures on weekdays were approximately 15 dB greater than the baseline [Fig. 3(d)], and, on weekends, pressures were 26 dB greater than the baseline [Fig. 3(e)]. Even during weekends of the lower human activity levels [Fig. 3(c)], when the

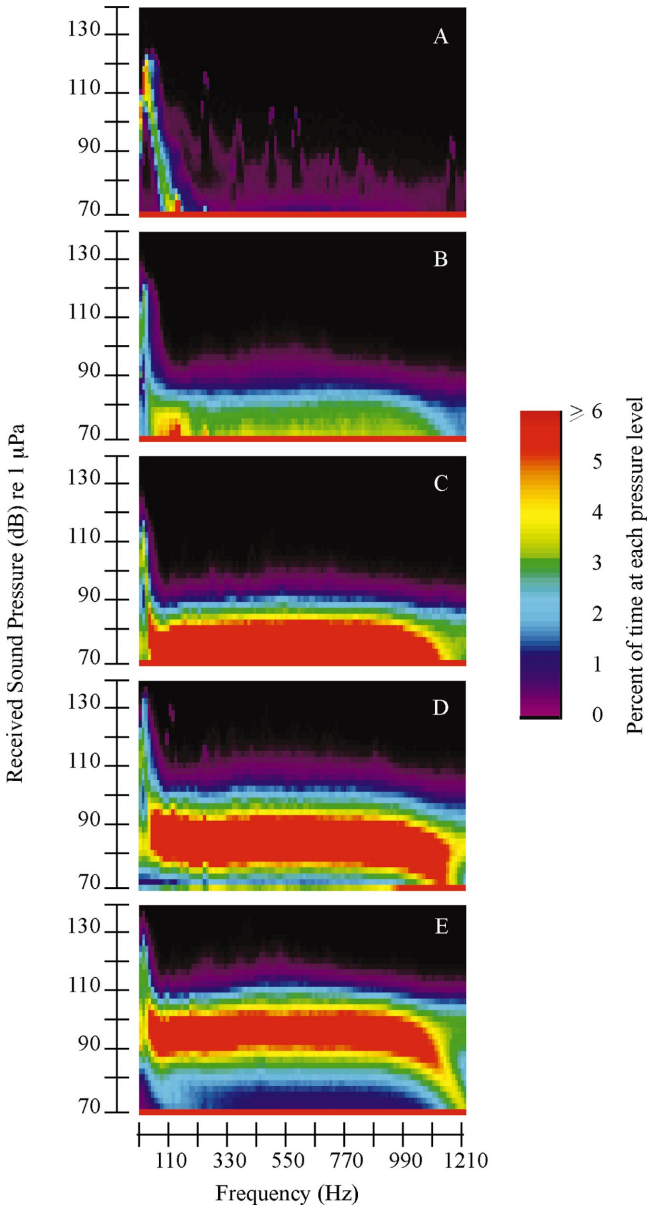


FIG. 2. A graphic depiction of both sound pressure and frequency, within the range of sea turtle hearing, over the entire range of human activity in Southold Bay. Distribution of sound is calculated by the percent of time a sound at a particular frequency throughout the sea turtle hearing range is found within a 2-dB range of received sound pressure at the hydrophone over the total hydrophone sampling period. (a) Lowest human activity—nighttime (number of spectral frames, $n=77\,504$); (b) Low human activity—weekday after Labor Day ($n=300\,988$); (c) Intermediate human activity—weekend after Labor Day ($n=159\,028$); (d) High human activity—weekday before Labor Day ($n=232\,512$); and (e) Highest human activity—weekend before Labor Day ($n=290\,640$).

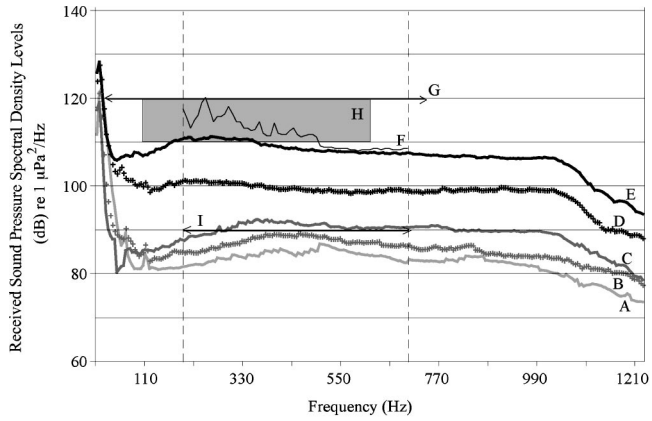


FIG. 3. Sound pressure spectral density level profiles in Southold Bay filtered for frequencies within the sea turtle hearing range (below 1200 Hz) during five periods of different human activity levels, compared to pressures of known levels of noise and to U.S. Navy LFA signals that elicit a response in sea turtles. (a) Lowest human activity—nighttime (number of spectral frames, $n=77\,504$); (b) Low human activity—weekday after Labor Day ($n=300\,988$); (c) Intermediate human activity—weekend after Labor Day ($n=159\,028$); (d) High human activity—weekday before Labor Day ($n=232\,512$); (e) Highest human activity—weekend before Labor Day ($n=290\,640$); and (f) maximum sound pressure levels recorded within the sea turtle peak hearing range, during the highest human activity level. Vertical dashed lines bracket the peak hearing range of sea turtles (200–700 Hz). The maximum noises recorded in the bay were comparable in pressure to the documented levels, which evoked disturbance responses in sea turtles: (g) O'Hara (1990) and (h) Samuel *et al.*, in manuscript. Pressure levels measured in the bay ecosystem during high human activity levels were much higher than levels at which sea turtles showed a behavioral response: (i) Samuel *et al.* (in manuscript).

difference was not significant, sound pressures were higher in many frequencies than baseline values by as much as 4 dB.

A direct measure of anthropogenic noise was seen best in a comparison of noise levels between weekdays and weekends within the same human activity season. Any differences observed in mean sound pressures between weekdays and weekends could not readily be attributed to biological or seasonal differences and, therefore, were ascribed to the higher boating activity on weekends. After the main boating season, even though sound levels were not significantly different between weekdays and weekends, there was still a 4-dB increase in noise levels on weekends. During the height of the boating season differences were even more obvious, with weekend mean sound pressure levels significantly higher ($p<0.05$) than weekdays, often by more than a factor of 8 dB (Fig. 3). This was particularly true at frequencies within the peak hearing range of sea turtles during this season. The difference in the band levels of sound pressure within the 200–700-Hz band between weekend and weekday was approximately 11 dB (Table II).

During these extremely noisy conditions, when human activity and mean sound pressures were highest, maximum levels of noise observed [Fig. 3(f)] rose far above the mean noise levels. For a sense of perspective, these observed means and maxima were compared to values from other studies in which sea turtles were subjected to a range of pressure levels to determine their behavioral response. Indeed, the maximum pressures recorded in the bay were comparable to sound pressures produced by high-pressure air gun pulses [Fig. 3(g)], which evoked changes in swimming patterns and orientation of sea turtles (O'Hara, 1990). In addition, the maximum levels in the bay also were within the range of pressures [Fig. 3(h)] at which sea turtles in tanks exhibited agitated behavior when subjected to simulated boat signals and recordings of the U.S. Navy's Low Frequency Active sonar signals (LFA). Details of these measurements are in a separate manuscript. Similar tank studies demonstrated that noise levels in the bay during the high human activity are at least 10 to 20 dB higher than pressures at which sea turtles showed a behavioral response [Fig. 3(i)].

IV. DISCUSSION

During much of the season when sea turtles are actively foraging in New York waters, their coastal habitats are flooded with underwater noise, a substantial component of which is anthropogenic in origin. Much of this noise occurs at frequencies below 1000 Hz, squarely within the range of highest sensitivity for sea turtle hearing. In mid-summer, during periods of high human activity, levels of noise were correspondingly high, and, as human activity decreased, both the pressure and complexity of underwater noise diminished. In general, daytime noise levels remained high for several weeks, until after the Labor Day holiday, when both recreational activity and noise pressure decreased abruptly.

A more detailed analysis of the noise profiles within the seasons highlighted the extent of anthropogenic noise. In addition, the observed patterns indicated that much of the low-frequency noise stemmed from recreational boating activity. On weekends, there were distinctly higher noise levels than on weekdays, and the same pattern was evident during both summer and fall sampling periods. Furthermore, the measured increases in sound pressures were substantial. At the peak of the recreational boating season, average noise pressures at many frequencies within the range of sea turtle hearing were more than 8 dB greater than those recorded during the corresponding weekdays. These levels were more than 26 dB greater than noise levels recorded at night after Labor Day (Fig. 3). Notwithstanding seasonal and diurnal differences, the noise signature from boats was strikingly evident in the sea turtle foraging habitat.

There was a substantial change also in the character of the noise at times of increased boating activity. At the lowest level of boating activity, namely at nighttime after Labor Day, most of the detectable sounds were at frequencies below 200 Hz. At frequencies in the range of peak sea turtle sensitivity (200–700 Hz) and higher, there were very few detectable signals throughout the entire monitoring session [Fig. 2(a)]. On weekdays after Labor Day, although there was an increase in signals at higher frequency ranges, most of the underwater sounds similarly were in the lower (50–200 Hz) frequencies [Fig. 2(b)]. In contrast, on weekends and during the height of the recreational season, underwater noise was louder and much more complex, extending relatively evenly across the entire measured frequency spectrum ranging up to 1200 Hz [Figs. 2(c)–(e)].

A comparison of the sound pressure levels in the 200–700-Hz band measured under the five different human activity levels (Table II), with the corresponding sound pressure band levels of ambient sounds from environmental sources as reported by Wenz (1962), reveals that the noises recorded in Southold Bay, even during the daytime period of lowest human activity, are much higher than the reported levels. In shallow waters, such as the Southold Bay area, the primary source of noise from environmental sources is wind and waves (Richardson *et al.*, 1995). According to Wenz (1962), at a wind force of 1, on the Beaufort scale, sound pressure in the 200–700-Hz band is approximately 46 dB re 1 μPa. As the wind force and consequently wave height increase, so does the sound pressure. At 2 Beaufort, noise levels rise to 55 dB, and at 3 Beaufort reach 59 dB. Even during the highest wind force we collected data, at 5 Beaufort, noise levels in the 200–700-Hz band attributed to wind by Wenz (1962) are 66 dB. Although wind contributes to underwater noise levels, our high band levels, even during the period with the lowest daytime recreational human activity, indicate that biological noise, distant shipping, or even noise from land are increasing ambient noise levels in Southold Bay throughout our recording period. The contribution of these sources to the noise levels in the bay can be seen in the sound pressure band levels recorded during weekdays after Labor Day, a period which we consider to be the baseline of daytime noise levels (Table II).

Juvenile sea turtles are found in New York coastal waters and estuarine ecosystems for a period of about 21 weeks each year, starting at the beginning of July (Morreale and Standora, 1998). The data from this study indicate that these foraging sea turtles spend nearly half of their activity season surrounded by higher than the baseline levels of underwater noise. Rather than being a unique situation, the Long Island study site is likely to be very representative of numerous other coastal habitats where foraging, mating, and nesting sea turtles congregate during key stages in their lives. Similarly, many of these crucial habitats exist in very populous areas where sea turtle activity coincides with exposure to high human activity and underwater noise.

From previous research it is evident that sea turtles can detect sound, and that their hearing is confined to lower frequencies, mainly below 1000 Hz (Ridgway *et al.*, 1969; Bartol *et al.*, 1999). Therefore, sea turtles are able to hear the low-frequency sound emitted underwater by anthropogenic sources, and such exposure could be directly influencing their health and ecology. However, little is known about the potential effects of noise exposure on the short-term or longer-term behavior and health of sea turtles. Nevertheless, several studies have shown that sea turtles in experimental tanks respond to signals at pressures well below those recorded in the Southold Bay foraging habitat, and some of the responses indicate at least short-term disturbance. Sea turtles displayed agitated behavior, abrupt body movements, startle responses, and even prolonged inactivity at the bottom of the tank in response to low-frequency signals (Lenhardt *et al.*, 1983, 1996). Such responses are similar to the ones we observed during exposure of sea turtles to LFA signals and simulated boat sounds. In a similar study in outdoor enclosures, sea turtles exposed to louder pulses from high-pressure air guns exhibited changes in swimming patterns and orientation (O'Hara, 1990).

Long-term effects on sea turtles after exposure to noise, although more insidious, would be much harder to detect. However, it is possible that prolonged exposure could be highly disruptive to the health and ecology of the animals as it has been recorded to affect other marine animals, by encouraging avoidance behavior (Malme *et al.*, 1983; Richardson *et al.*, 1986; Richardson, 1997), increasing stress and aggression levels (Lagardère, 1982), causing physiological damage to the ears through either temporary or even permanent threshold shifts (Hastings *et al.*, 1996; Scholik and Yan, 2001; Erbe, 2002; McCauley *et al.*, 2003), altering surfacing and diving rates (Suzuki *et al.*, 1980; Blaxter *et al.*, 1981;

Olsen *et al.*, 1983; Schwarz and Greer, 1984; Engås *et al.*, 1998; Williams *et al.*, 2002), or even confounding orientation cues. Sea turtles have been shown to exhibit strong fidelity to fixed migratory corridors, habitual foraging grounds, and nesting areas (Morreale *et al.*, 1996; Morreale and Standora, 1998; Avens *et al.*, 2003), and such apparent inflexibility could prevent sea turtles from selecting alternate, quieter habitats.

It is likely that sea turtles and other marine animals are not adapted to contemporary noise levels. Over the past 33 years there has been a documented 9- to 10-dB increase in ambient sound levels in some coastal marine environments (Andrew *et al.*, 2002). Furthermore, there has been a 6.3% increase in the usage of recreational boats in the last decade (NMMA, 2003). Thus, even with the current noise abatement regulations, which prohibit vessel operation in inland waters that exceed noise levels above 82 dB re 1 μPa, measured at a distance of 25 m from the vessel (36CFR3 Title 36, Chapter I, Part 3, Sec. 3.7), cumulative noise levels are likely to keep increasing as recreational motor boat use rises. Additional new sources, such as seismic exploration activities in shallow water, which are increasing worldwide (Frisk *et al.*, 1998), also will undoubtedly contribute to the pervasive noise levels.

Anthropogenic noise in the underwater environment appears to be more than an abstract issue inevitably related to increasing human populations. Our study in a representative coastal habitat highlights the extent of the problem. Sea turtles tend to cluster, as is the habit of many other marine animals, in continental shelf and pelagic waters of the western North Atlantic where there is extensive human activity. This spatial overlap undoubtedly exposes many of them to high levels of ambient noise, especially during warmer months. Currently, the effects of such exposure are largely unknown. At the least, underwater noise must be accounted for when devising appropriate management strategies for the protection and recovery of rare and endangered organisms such as sea turtles. At best, future marine conservation plans should include careful acoustic monitoring to prevent future unabated noise pollution.

ACKNOWLEDGMENTS

We would like to thank a number of people for their contribution. From the Cornell Laboratory of Ornithology, we thank Melissa Craven Fowler, Christopher Tessaglia-Hymes, Connie Gordon, Kurt Fristrup, Bob Grotke, and Marguerite McCartney for providing all the equipment and helping with the data retrieval. From the Cornell Ocean Resources and Ecosystems Program, we thank all members of the laboratory group, and especially Bruce Monger, for his invaluable help and contributions to the data analyses. From Suffolk County Cornell Cooperative Extension Marine Program, we thank Chris Smith, Rory MacNish, Stacy Myers, Emerson Hasbrouck, Eileen Brennan, Sonia Tulipano, and Ronnie Matovcik for providing funding, support, and access to the study site. Special thanks to Kim Durham, Robert DiGiovanni, Chris Buckman, Kelly Cantara from New York State Marine Mammal and Sea Turtle Stranding Program for support and access to animals. Bennett S. Orlowski from the Long Island Horticultural Research and Extension Center kindly provided living accommodations. We further thank Lt. Commander USCG, Jeannot Smith, for field and technical support. Our thanks also to Riverhead Foundation for Marine Research and Preservation, and National Marine Fisheries Service for support and funding.

36CFR3 (**2001**). Code of Federal Regulations. Title 36, Volume 1, Chapter I, Part 3, Section 3.7. Revised July 2001.

Andrew, R. K., Howe, B. M., Mercer, J. A., and Dzieciuch, M. A. (**2002**). ''Ocean ambient sound: Comparing the 1960s with the 1990s for a receiver off the California coast,'' ARLO **3**(2), 65–70 (2002).

Avens, L., Braun-McNeill, J., Epperly, S., and Lohmann, K. J. (**2003**). ''Site fidelity and homing behavior in juvenile sea turtles (*Caretta caretta*),'' Mar. Biol. (Berlin) **143**(2), 211–220.

Bartol, S. M., Musick, J. A., and Lenhardt, M. L. (**1999**). ''Auditory evoked potentials of the loggerhead sea turtle,'' Copeia **3**, 836–840.

Blaxter, J. H. S., Gray, J. A. B., and Denton, E. J. (**1981**). ''Sound and startle responses in herring shoals,'' J. Mar. Biol. Assoc. U.K. **61**, 851–869.

Charif, R. A., Clark, C. W., and Fristrup, K. M. (**2003**). ''Raven 1.1 User's Manual,'' Cornell Laboratory of Ornithology, Bioacoustics Research Program, Ithaca, NY.

Curtis, K. R., Howe, B. M., and Mercer, J. A. (**1999**). ''Low-frequency ambient sound in the North Pacific: Long time series observations,'' J. Acoust. Soc. Am. **106**, 3189–3200.

Engås, A., Haugland, E. K., and Øvredal, J. T. (**1998**). ''Reactions of cod (*Gadus morhua L.*) in the pre-vessel zone to an approaching trawler under different light conditions. *Preliminary Results*,'' Hydrobiologia **371/372**, 199–206.

Erbe, C. (**2002**). ''Underwater noise of whale-watching boats and potential effects on killer whales (*Orcinus orca*), based on an acoustic impact model,'' Marine Mammal Sci. **18**(2), 394–418.

Frisk, G. V., Jackson, D., DeFerrari, H., Simmen, J., Spindel, R. C., Buck, J. R., D'Spain, G., Hilderbrand, J. A., and Greene, C. R. (**1998**). ''Ocean Acoustics,'' in Proceedings of the workshop on the effects of anthropogenic noise in the marine environment, edited by R. C. Gisiner, Marine Mammal Science Program, Office of Naval Research, pp. 10–18.

Fristrup, K. M., Hatch, L. T., and Clark, C. W. (**2003**). ''Variation in humpback whale (*Megaptera novaeangliae*) song length in relation to low-frequency sound broadcasts,'' J. Acoust. Soc. Am. **113**, 3411–3424.

Hastings, M. C., Popper, A. N., Finneran, J. J., and Lanford, P. J. (**1996**). ''Effects of low-frequency underwater sound on hair cells of the inner ear and lateral line of the teleost fish *Astronotus ocellatus*,'' J. Acoust. Soc. Am. **99**, 1759–1766.

Lagardère, J. P. (**1982**). ''Effects of noise on growth and reproduction of *Crangon crangon* in rearing tanks,'' Mar. Biol. (Berlin) **71**, 177–185.

Lenhardt M., Moein, S., and Musick, J. (**1996**). ''A method for determining hearing thresholds in marine turtles,'' in Proceedings of the fifteenth annual workshop on sea turtle biology and conservation, NOAA technical Memorandum NMFS-SEFSC-387.

Lenhardt, M. L., Bellmund, S., Byles, R. A., Harkins, S. W., and Musick, J. A. (**1983**). ''Marine turtle reception of bone-conducted sound,'' J. Aud Res. **23**, 119–125.

Lesage, V., Barrette, C., Kingsley, M. C. S., and Sjare, B. (**1999**). ''The effect of vessel noise on the vocal behavior of belugas in the St. Lawrence river estuary, Canada,'' Marine Mammal Sci. **15**(1), 65–84.

Malme, C. I., Miles, P. R., Clark, C. W., Tyack, P., and Bird, J. E. (**1983**). ''Investigations of the potentials effects of underwater noise from petroleum industry activities on migrating gray whale behaviour: final report for the period of 7 June 1982–31 July 1983,'' Report No. 5366, prepared by Bolt, Beranek and Newman Inc., Cambridge, MA, USA, for U.S. Minerals Management Service, Alaska OCS Office, Anchorage, USA.

McCauley, R. D., Fewtrell, J., and Popper, A. N. (**2003**). ''High pressure anthropogenic sound damages fish ears,'' J. Acoust. Soc. Am. **113**, 638–642.

Miller, P. J. O., Biassoni, N., Samuels, A., and Tyack, P. L. (**2000**). ''Whale songs lengthen in response to sonar,'' Nature (London) **405**(6789), 903–903.

Morreale, S. J., and Standora, E. A. (**1998**). ''Early life stage ecology of sea turtles in Northeastern U.S. waters,'' NOAA Tech. Mem. NMFS-SEFSC-413.

Morreale, S. J., Standora, E. A., Spotila, J. R., and Paladino, F. V. (**1996**).

''Migration corridor for sea turtles,'' Nature (London) **384**, 319–320.
National Marine Manufacturers Association (**2003**). ''Boating 2002: At a glance—Facts & Figures,'' Marketing Statistics Department, Chicago, IL.
O'Hara, J. (**1990**). ''Avoidance responses of loggerhead turtles, *Caretta caretta*, to low frequency sounds,'' Copeia **2**, 546–567.
Olsen, K., Angell, J., Pettersen, F., and Løvik, A. (**1983**). ''Observed fish reactions to a surveying vessel with special reference to herring, cod, capelin and polar cod,'' FAO Fish. Rep. **300**, 131–138.
Plotkin, P. T. (ed.). (**1995**). National Marine Fisheries Service and U.S. Fish and Wildlife Service Status Reviews for Sea Turtles Listed under the Endangered Species Act of 1973, National Marine Fisheries Service, Silver Spring, MD.
Richardson, W. J. (**1997**). ''Marine mammal and acoustical monitoring of BPXA's seismic program in the Alaskan Beaufort Sea,'' LGL Report TA2150-3, prepared by LGL Ltd. and Greeneridge Sciences, Inc. for BP Explorations (Alaska), Inc. and National Marine Fisheries Service. King City, Can., LGL Ltd.
Richardson, W. J., Würsig, B., and Greene, Jr., C. R. (**1986**). ''Reactions of bowhead whales, *Balaena mysticetus*, to seismic exploration in the Canadian Beaufort Sea,'' J. Acoust. Soc. Am. **79**, 1117–1128.
Richardson, W. J., Greene, Jr., C. R., Malme, C. I., and Thomson, D. H. (**1995**). *Marine Mammals and Noise* (Academic, San Diego, CA).
Ridgway, S. M., Wever, E. G., McCormick, J. G., Palin, J., and Anderson, J. H. (**1969**). ''Hearing in the giant sea turtle, *Chelonia mydas*,'' Psychology **64**, 884–890.
Samuel, Y. (**2004**). ''Underwater low-frequency noise and anthropogenic disturbance in a critical sea turtle habitat,'' M.S. thesis, Cornell University.
Scholik, A. R., and Yan, H. Y. (**2001**). ''Effects of underwater noise on auditory sensitivity of a cyprinid fish,'' Hear. Res. **152**, 17–24.
Schwarz, A. L., and Greer, G. L. (**1984**). ''Responses of pacific herring, *Clupea harengus pallasi*, to some underwater sounds,'' Can. J. Fish. Aquat. Sci. **41**, 1183–1192.
Suzuki, H., Hamada, E., Saito, K., Maniwa, Y., and Shirai, Y. (**1980**). ''The influence of underwater sound on marine organisms,'' J. Nav. **33**, 291–295.
Wenz, G. M. (**1962**). ''Acoustic ambient noise in the ocean: Spectra and sources,'' J. Acoust. Soc. Am. **34**, 1936–1956.
Williams, R., Trites, A. W., and Bain, D. E. (**2002**). ''Behavioural responses of killer whales (*Orcinus orca*) to whale-watching boats: opportunistic observations and experimental approaches,'' J. Zool. **256**, 255–270.

Three-dimensional beam pattern of regular sperm whale clicks confirms bent-horn hypothesis

Walter M. X. Zimmer
NATO Undersea Research Centre, V.le San Bartolomeo 400, 19138 La Spezia, Italy

Peter L. Tyack, Mark P. Johnson, and Peter T. Madsen
Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

(Received June 30, 2004; revised September 30, 2004; accepted October 11, 2004)

The three-dimensional beam pattern of a sperm whale (*Physeter macrocephalus*) tagged in the Ligurian Sea was derived using data on regular clicks from the tag and from hydrophones towed behind a ship circling the tagged whale. The tag defined the orientation of the whale, while sightings and beamformer data were used to locate the whale with respect to the ship. The existence of a narrow, forward-directed P1 beam with source levels exceeding 210 dB_{peak} *re*: 1 μPa at 1 m is confirmed. A modeled forward-beam pattern, that matches clicks $>20°$ off-axis, predicts a directivity index of 26.7 dB and source levels of up to 229 dB_{peak} *re*: 1 μPa at 1 m. A broader backward-directed beam is produced by the P0 pulse with source levels near 200 dB_{peak} *re*: 1 μPa at 1 m and a directivity index of 7.4 dB. A low-frequency component with source levels near 190 dB_{peak} *re*: 1 μPa at 1 m is generated at the onset of the P0 pulse by air resonance. The results support the bent-horn model of sound production in sperm whales. While the sperm whale nose appears primarily adapted to produce an intense forward-directed sonar signal, less-directional click components convey information to conspecifics, and give rise to echoes from the seafloor and the surface, which may be useful for orientation during dives. © *2005 Acoustical Society of America.*
[DOI: 10.1121/1.1828501]

PACS numbers: 43.80.Ka, 43.80.Ev, 43.30.Sf [WWA] Pages: 1473–1485

I. INTRODUCTION

The vocal repertoire of sperm whales is generally reported to be limited to click sounds. Individual sperm whales can produce a variety of kinds of clicks, which are thought to function for several different echolocation and communicative functions. When diving, sperm whales produce long series of clicks with regular interclick intervals (ICI) of 0.5–2 s (called ''usual clicks'' by Whitehead and Weilgart, 1991). The purpose of these regular clicks was long disputed. Most biologists assumed by analogy with the better-studied biosonar of bats and dolphins that regular clicks are used for echolocation (Backus and Schevill, 1966; Gordon, 1987). However, several bio-acousticians have argued that the observed mean source level (160–180 dB *re*: 1 μPa at 1 m) and the lack of significant directionality were not compatible with successful echolocation of prey (Watkins, 1980; Fristrup and Harbison, 2002). Recent measurements suggest that regular clicks are in fact highly directional, with source levels of up to 235 dB_{rms} *re*: 1 μPa at 1 m on the axis of the sound beam (Møhl *et al.*, 2003). Tyack (1997) showed echoes of regular clicks from the seafloor, and Jaquet *et al.* (2001) and Gordon and Tyack (2002) demonstrated that the first clicks at the start of the descent of a dive correlated with the round-trip travel time to the bottom in some locations. These data suggest that sperm whales echolocate at least on the seafloor.

Diving sperm whales also make bursts of clicks with higher repetition rates, called ''creaks'' (Gordon, 1987; Mullins *et al.*, 1988; Madsen *et al.*, 2002). By analogy with the terminal buzz produced when some bat species close on aerial prey (Griffin, 1958), most authors have suggested that usual clicks produced at the bottom of foraging dives may represent a search phase of echolocation, and that sperm whales produce creaks, or terminal buzzes, as they close on prey.

Sperm whales also make clicks assumed to be used for social communication. Weilgart and Whitehead (1988) describe distinctive intense reverberant clicks with long ICI typically of 5–7 s. These were only recorded in the presence of mature males and are thought to advertise the maturity and competitive ability of the male. Stereotyped repetitive series of clicks called ''codas'' (Watkins and Schevill, 1977) are recorded from many different groups of sperm whales, especially when whales are socializing near the surface (Gordon, 1987; Whitehead and Weilgart, 1991). These coda vocalizations are thought to function for social communication within these groups (Weilgart and Whitehead, 1993; Moore *et al.*, 1993; Rendell and Whitehead, 2003).

Backus and Schevill (1966) observed that sperm whale clicks last in excess of 10 ms and are composed of a series of pulses of short duration (0.1–2 ms). They showed further that the relative amplitude and timing of the pulses within the clicks obey no consistent rule among whales. However, the relative amplitudes and timing show an apparent reproducibility from click to click from individual whales at least on a short time scale; Backus and Schevill (1966) considered this as signature for the individuals. Norris and Harvey (1972) proposed that sperm whale clicks are generated by the so-called phonic lips (also known as the *museau de singe*, or monkey lips) and that the dominating first pulse is directly transmitted into the water ahead of the whale, while the re-

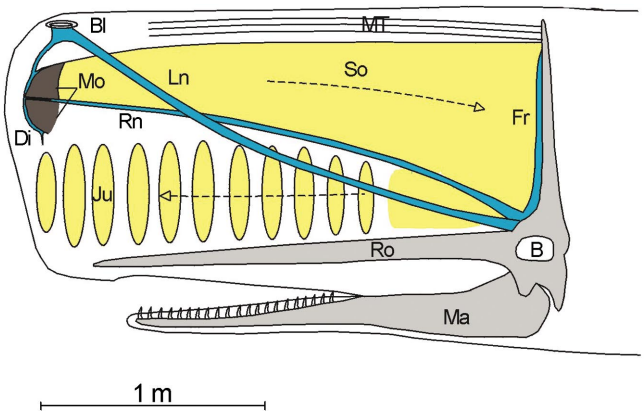


FIG. 1. Diagram of bent-horn model of sound production in sperm whales (modified from Fig. 1 of Madsen *et al.* 2002). B, brain; Bl, blow hole; Di, distal air sac; Fr, frontal air sac; Ju, junk; Ln, left naris; Ma, mandible; Mo, monkey lips/*museau de singe*; MT muscle/tendon layer; Rn, right naris; Ro rostrum; So, spermaceti organ.

maining pulses represent reverberations of the backwardly directed portion of the original pulse that are reflected between two reflecting air sacs that bound the ends of the spermaceti organ. Concerning the way the reverberated pulses leave the spermaceti organ, Norris and Harvey (1972) were inconclusive and proposed two possibilities for the exact exit site, the upper phonic lip, at the anterior termination of the spermaceti organ, or alternatively the well-developed mesorostral cartilage of the rostrum (Norris and Harvey, 1972).

The Norris and Harvey (1972) sound generation model for sperm whales has been modified by Møhl (2001), who proposed in his ''bent-horn'' model that some of the acoustic energy generated by the phonic lips (labeled ''Mo'' in Fig. 1) escapes directly into the water, generating a P0 pulse. The majority of acoustic energy propagates back through the spermaceti organ (''So'' in Fig. 1) to the frontal air sac (Fr) in front of the skull, where it is reflected downward and forward into the junk (Ju), from which it propagates into the seawater as a forward-directed P1 pulse. The remaining sound energy is reflected from the frontal sac back into the spermaceti organ where it returns to the distal sac. Most of this energy is again reflected backwards to repeat the path of the original path, and so on for P2, P3, etc pulses (Møhl, 2001). Acoustic data from a sperm whale neonate in rehabilitation confirmed that sperm whale clicks are produced at phonic lips in the anterior end of the spermaceti organ, and that sound produced there reverberates in the spermaceti organ and is transmitted to the seawater via the junk complex (Madsen *et al.*, 2003; Møhl *et al.*, 2003).

Møhl *et al.* (2000, 2003) used a dispersed array of hydrophones to test the hypothesis of an intense, directional P1 pulse predicted by the bent horn model. By selecting a few clicks from entire seasons of recording, they showed that some sperm whale clicks have centroid frequencies of about 15 kHz, high directionality of over 27 dB, and source levels of up to 235 dB_{rms} re: 1 μPa at 1 m. While they could not measure the orientation of the clicking whale directly, Møhl *et al.* (2000, 2003) assumed that if they detected a strong click with a dominant single pulse, this would represent an on-axis recording in the beam of the P1 pulse. They argue that this high source level and the narrow sound beam represent adaptations for long-range echolocation on mesopelagic prey with low target strength (Møhl *et al.*, 2003; Madsen *et al.*, 2002).

By contrast, Watkins (1980) analyzed data from dozens of cruises and reported that sperm whale clicks have variable spectra and no apparent propagation directionality. Watkins (1980) observed that the spectral component of sperm whale clicks extended over 30 kHz when the whales were within 20 m. At about 2 km most of the audible energy was below 5 to 6 kHz with apparent emphases at 2 to 4 kHz. Watkins and Daher (2004) presented the underwater recordings of fast click sequences (2.3/s to 8.8/s) of a small whale near the surface during head-out episodes when the whale exposed the lower jaw as far as the jaw hinge. They state that they did not observe directional clicks as reported by Møhl *et al.* (2000).

These two views on the acoustic properties of clicks led to strikingly different interpretations as to their function. Watkins (1980) concluded that sperm whale regular clicks appeared to be used mostly for communication to coordinate movements of whales as they dive and disperse. While recognizing that most other researchers assumed an echolocation functionality for sperm whale clicks, Watkins (1980) pointed out that his observations of sperm whale sounds did not match the characteristics expected for an echolocation signal, especially when compared with the click characteristics from echolocating dolphins: sperm whale clicks do not appear to be highly directional; the click repetition rate is generally very regular and in particular does not vary with the changing distance to approaching targets; sperm whales can be silent for long periods, especially when they are alone; sperm whale clicks can be heard over such long distances that sound signal distortion becomes a limiting factor; individual clicks are longer and more complex than the echolocation signals of other odontocetes, most of which are thought to approximate an impulse. In summary, Watkins concluded that sperm whale clicks do not have acoustic features expected for echolocation, but rather seem to fit a context of communication (Watkins, 1980). Other researchers such as Fristrup and Harbison (2002) have also been skeptical of the idea that regular clicks can be successfully used for echolocation due to the low target strength of squid.

Here, we present data to suggest how such two very different views on sperm whale clicks may have been reached. We demonstrate that the P1 pulse does form a high-powered, forward-directed beam with properties consistent with the conclusions of Møhl *et al.* (2000, 2003). The initial pulse generates a P0 pulse and associated low-frequency components that are relatively omnidirectional, which may explain the conclusions reached by Watkins, who apparently never recorded the narrow, forward-directed P1 pulse on axis. Our data indicate that the P0 pulse contains $<10\%$ of the energy of the P1 pulse, but this is still intense enough to be detectable over long distances and even to generate echoes from the seafloor and the surface. The P0 pulse has a backwards directionality that is a necessary and heretofore unpredicted consequence of the bent-horn model. We show that regular clicks of sperm whales have temporally and

spectrally separated acoustic components by which the animal simultaneously produces a narrow, high-frequency sonar beam to search for prey, and less directional components that may be used for communication and perhaps orientation. A similar dual function of biosonar signals has been proposed for the high- and low-frequency components of harbor porpoise clicks (Møhl and Andersen, 1973).

II. MATERIALS AND METHODS

Our paper is based on data recorded in the Ligurian Sea in 2001 during *Sirena-01*, a field trial organized by the NATO Undersea Research Center (NURC) as part of its Marine Mammal Acoustic Risk Mitigation program. The data collection during *Sirena* trials was based on multiple components: visual observation of animals at the surface, passive sonar detection and tracking while animals were diving, and tagging of animals with a compact acoustic data recorder (Johnson and Tyack, 2003) developed by the authors at the Woods Hole Oceanographic Institution (WHOI).

The procedure for tagging sperm whales was based on the following scheme: During the night and early morning, visual observers and operators of a passive sonar tried to locate and to approach sperm whales for tagging. Once a sperm whale was tracked acoustically and visually, a small workboat was deployed from NRV ALLIANCE to attach the tag to the animal. Any potential responses to tagging were monitored visually and acoustically from NRV ALLIANCE as well as from the small vessel. After tagging, the whale could be followed visually when close to the ship, acoustically when it was clicking, and using a radio direction finder to track a VHF radio transmitter on the tag when the whale surfaced.

A. Visual observation

A visual watch was established during daylight hours on the flying bridge of NRV ALLIANCE, which provided a stable platform for visual observation at a height of 16 m over the water. Big-eye binoculars enabled observation up to 10 km. Once a whale was located either visually or acoustically, it was selected as the ''focal follow'' animal and its detailed behavior (blow rate, swim speed, etc.) was recorded to establish behavioral patterns.

B. Passive sonar

A major asset available during *Sirena* trials was the passive sonar system on the NRV ALLIANCE. This passive sonar was developed at NURC and consisted of a horizontal line array that was towed at about 80-m depth, just below any substantial thermocline, a real-time digital beamformer, and sonar display system. It was deployed almost continuously, enabling a 24-h listening operation during most weather conditions up to sea state 7. Passive sonar technology was applied to detect and to track vocalizing animals when they were submerged and therefore not visible. Sperm whales are known to emit intense clicks at regular intervals while they are diving, and consequently are ideal for tracking by passive sonar.

Underwater signals were received by a towed line array of 128 hydrophones with the system saturation set to 140 dB_{peak} re: 1 μPa. The hydrophone separation was 9 cm and the sampling frequency was 31.25 kHz, allowing a maximum bandwidth of about 15 kHz. The acquired acoustic data were archived on a 240-Mbit/s digital tape recorder, together with relevant nonacoustic data such as array depth and ship's position, heading, and speed. The received array data were transformed to angular space using a digital time-delay beamformer (Zimmer *et al.*, 2003). The beamformed data were used to find potential animals for tagging during the night and to track them during their deep dives. A close handshaking with the visual team made sure that focal animals were not lost. The continuous acoustic watch brought NRV ALLIANCE also in good positions to sight surfacing animals.

While the beamformer was useful for tracking the focal animal, a broadband analysis is more properly done with the data from individual hydrophones. For the subsequent analyses, data from two hydrophones (10.62 m apart) and relevant nonacoustic data (array depth, ship's position, GPS time stamp) were extracted from the hydrophone recordings.

C. Tagging

The passive sonar was able to track the gross movements of the whale, but not to record the detailed orientation or short-term movements of the animal between clicks. A digital tag (DTAG), developed by the authors at WHOI, was therefore used to record sound and high-resolution movement patterns directly from the whale (Johnson and Tyack, 2003). Key features of the DTAG were 16-bit analog-to-digital conversion at a hydrophone sampling rate of 32 kHz and clipping level set to 153 dB_{peak} re: 1 μPa, further pressure sensor, 3-axis accelerometer, and 3-axis magnetometer, all sampled at 50 Hz.

The tagging team approached the focal animal in a small boat at low speed. The tag was deployed by means of a 12-m carbon-fiber pole, mounted in a bow-mounted cantilever and attached to the whale with suction cups. After a preprogramed release time, the tag floated to the surface and was then tracked by taking bearings to a built-in VHF radio transmitter. The data acquired by the sensors of the tag were recorded on 3-GB flash memory, downloaded after recovery, and stored on CD-R for archiving and processing.

The recordings of the DTAG on the whale are in general not in synchrony with the passive sonar recordings on board NRV ALLIANCE. Synchronization is therefore a critical step in the data processing and will be addressed in Sec. III C 3.

III. DATA PROCESSING

The main goals of the data analysis were to obtain temporal, spectral, and level characteristics of sperm whale clicks as a function of spatial orientation of the whale with respect to a far-field sensor. Two data processing tasks had to be undertaken to achieve these analysis goals: to determine the motion and orientation of the sperm whale and to extract the temporal, spectral, and level characteristics of each click

from the acoustic recordings. The data available were visual observations, tag and passive sonar recordings.

A. Visual

When the focal whale was at the surface, the visual team observed it continuously from the flying bridge of NRV ALLIANCE. Range and bearing were logged into a computerized logging system for each behavioral event, such as blow, fluke-up, breach, etc. To estimate the target range, the reticule scale of the Big-eyes was used to measure the vertical angle of the focal animal below the horizon (Kinzey and Gerrodette, 2001). To estimate the bearing of the animal, the Big-eyes was fitted with a bearing encoder, which gave the bearing relative to the bow of the ship. The readout of the ship's gyroscope was then used to convert the relative bearing to absolute bearing of the focal animal. The availability of the ship's gyro eliminated the requirement for a magnetic compass that was considered problematic in proximity to metal structures on the flying bridge.

B. Tag data

The tag was the primary data source for the time at which the whale emitted each click, and for animal depth and orientation.

1. Click detection and classification

The times at which the clicks were emitted by the tagged animal were extracted from the acoustic recordings using an automatic click detector. The click detector was based on the Page test, a sequential probability ratio test that takes as input the time series of the received sound and determines beginning and end of transients (detections) by means of the following algorithm (Page, 1954; Wald, 1947; Abraham, 2000):

Given the instantaneous signal magnitude x_n, calculate a test variable V_n and make the decision for detection of a transient or signal according to

$$V_n = V_{n-1} + \left(\frac{x_n^2}{N_n} - b \right) \times \begin{cases} > V_1 & \text{decide detection and set } V_n = V_1, \\ < V_0 & \text{decide noise and set } V_n = V_0, \end{cases} \quad (1)$$

with N_n = noise estimate; b = bias for test variable; V_0 = threshold for decision of noise; and V_1 = threshold for decision of detection.

To obtain the signal magnitude, the real-valued data were first Hilbert transformed to a complex-valued (analytic) representation.

For each sample the detector algorithm could output one of three states, decision for signal, decision for noise, or decision deferred. In the last case the test variable V_n was augmented by a new measurement. We defined the duration of the signal by the time between the last noise decision and the time of the last signal decision. It can be shown that in theory, bias and thresholds may be related to the detection probability and probability of false alarm (Wald, 1947). However, for this analysis the values were determined empirically to give good detection performance: $b=4$, $V_0=1$, and $V_1=10^3$. The noise was estimated by exponential weighting $N_n=(1-\alpha)N_{n-1}+\alpha x_n^2 (\alpha=1/32)$ while noise was detected, or kept constant during signal detection. To avoid unrealistic long detections the maximum duration of any single detection was limited to 0.1 s, a value that is about 10 times a complete sperm whale click.

After signal (transient) detection, classification was based on the temporal characteristics of regular clicks, such as the slowly varying time interval between clicks and slowly varying peak levels of clicks. Initially, only two detection categories were defined: sperm whale clicks and echoes reflected from the surface. The signal detection was classified as a sperm whale click when the interval to the next sperm whale click was similar to the interval to the previous sperm whale click with compatible signal levels and click durations. A surface-reflected echo was expected to occur shortly after a sperm whale click with a delay that corresponded to twice the whale depth.

2. Animal depth and orientation

The animal depth is measured by a calibrated pressure sensor in the DTAG. A key innovation of the DTAG is its ability to measure the orientation of the tagged animal as a function of time. Orientation is deduced from the 3-axis accelerometer and magnetometer signals and is expressed in terms of the Euler angles, pitch, roll, and heading, with reference to the fixed (earth) frame. As the tag may be placed anywhere on the back of a whale, the tag axes do not generally coincide with the whale axes. There are thus three frames involved: the tag frame, the whale frame, and the earth frame. To determine the orientation of the whale, the angles of rotation (pitch, roll, and heading) relating the whale frame to the earth frame were determined from the DTAG's accelerometer and magnetometer. (Johnson and Tyack, 2003; Zimmer *et al.*, 2003).

3. Animal tracking

In general, for known animal speed $v(t)$, pitch and heading the animal track may be estimated by

$$\begin{pmatrix} p_x(t) \\ p_y(t) \\ p_z(t) \end{pmatrix} = \begin{pmatrix} p_x(t_0) \\ p_y(t_0) \\ p_z(t_0) \end{pmatrix} + \int_{t_0}^{t} v(\tau) \begin{pmatrix} \cos\beta(\tau)\sin\gamma(\tau) \\ \cos\beta(\tau)\cos\gamma(\tau) \\ \sin\beta(\tau) \end{pmatrix} d\tau. \quad (2)$$

The earth frame coordinate system is here assumed to be (east, north, up), the pitch $\beta(t)$ is positive up, and the heading $\gamma(t)$ is magnetic (relative) north.

Here, we start by considering only the data from the DTAG, that is, we have a detailed description of the orientation of the animal but no reliable speed information, and assume therefore as first approximation some (mean) speed $v(t)=v=$ const. The track generated by this assumption of constant speed is called a pseudotrack. The mean speed can be independently estimated using time and GPS position where the tagging occurred and a sighting toward the end of the tag attachment.

To estimate the final animal track, the DTAG data were combined with the passive sonar data. For this the range and bearing components of the constant-speed pseudotrack (r_p and b_p) were adjusted to approximate the measured acoustic range and bearing (r_a and b_a) to obtain final range and bearing estimates (r_e and b_e)

$$r_e(t) = r_p(t) + F_{\mathrm{LP}}(r_a(t) - r_p(t)),$$
$$b_e(t) = b_p(t) + F_{\mathrm{LP}}(b_a(t) - b_p(t)). \quad (3)$$

The purpose of the low-pass filter F_{LP} was to avoid impulsive (high-frequency) correction to the constant-speed (zero acceleration) pseudotrack, and therefore to obtain smooth variation in the resulting acceleration of the animal.

C. Passive sonar

The passive sonar data were used to estimate the location of the whale while it was clicking during the dive, to correct the speed estimation of the tagged animal, and to characterize the sperm whale clicks. The array was deployed from NRV ALLIANCE and all available nonacoustic data, e.g., GPS position and array depth, were recorded together with the acoustic data.

1. Array position, depth, and heading

While the ship position was measured using GPS, the array position had to be estimated by the following method. The array is towed behind the ship with a constant tow cable length. The depth of the array varies as a function of the tow-ship speed. Maneuvers were generally made smoothly to avoid excessive bending in the array. One may therefore assume that the array follows the ship track delayed by a certain time offset, that is, the array position is estimated as the ship's position at a later time

$$P_{\mathrm{array}}(t + \tau(t)) = P_{\mathrm{ship}}(t). \quad (4)$$

To obtain the time offset when the array was found at a given position, the ship's mean speed V_m is incorporated

$$\tau(t) = \frac{C_0 + \sqrt{C^2 - h(t)^2}}{V_m(t)}, \quad (5)$$

where C_0 is the horizontal distance from GPS receiver to stern of NRV ALLIANCE; C is the length of the tow cable; and $h(t)$ is the array depth.

The array depth was measured with a pressure sensor in the array and, as a heading sensor was not available, the array heading was derived directly from the array motion.

2. Acoustic bearing

To obtain the acoustic bearing of the sperm whale the time delay between two distant hydrophones of the same click was measured using cross correlation and transformed to angle. The passive sonar (beamformer)-based bearing estimate was not used for this task because, due to the limited number of beams, the existing angular accuracy was not sufficient. However, the clicks were tracked minute by minute on the passive sonar display to cross check the bearing derived from the time delay estimation.

3. Acoustic ranging

The range estimation is based on the travel time required for the sperm whale clicks to reach the passive sonar system. Independent range estimation was further carried out by exploiting the multipath structure of the received clicks and used to synchronize the acquisition systems of the DTAG and the passive sonar system.

a. *Multipath ranging*. The multipath ranging uses the path difference between the sperm whale click and the surface-reflected echo of the same click to estimate the range from the passive sonar to the clicking animal. As we have measured the animal depth by the DTAG, and as we know the array depth of the towed array, multipath ranging only requires the path difference between direct and surface-reflected arrival (Zimmer *et al.*, 2003). Depending upon the geometry, the expected delay between these two arrivals is less than a second. The travel time difference (surface reflected − direct click) was obtained from the automatic click detector applied to the hydrophone data. The surface reflection from some clicks interfered with the direct arrival; these were eliminated from the analysis. Before running the automatic click detector, the hydrophone data were bandlimited to 0.3 to 15 kHz (using a 128-taps FIR filter) to reduce the influence of ambient noise at low frequencies.

b. *Click travel time ranging*. To estimate the range using the click travel time, the actual time difference of the clicks between passive sonar (t_S) and DTAG (t_T) is measured and multiplied with sound speed c, which for small time differences and similar ranges may be assumed to be constant

$$r = c(t_S - t_T). \quad (6)$$

However, in order for this method to be successful the data must be sampled synchronously on both systems. As the sampling rate on the tag varies with the temperature of the oscillator crystal, the tag sampling rate is expected to vary as the animal dives deep into regions of colder water. To compensate for this temperature-dependent drift, the click travel time ranging was compared with the data from the multipath ranging and the difference in range estimation was fitted to a temperature-dependent model

$$\Delta R = c_0 + c_1 t + c_2 t(T - 20)^2. \quad (7)$$

Solving this equation for c_0, c_1, and c_2 for all range differences ΔR and temperatures T gave an optimal correction to the click travel time range estimation.

c. *Hydrophone data processing*. We did not use the beamformed data to estimate the spectra and levels of clicks, but used the data directly recorded from individual hydrophones. The time-delay beamforming process uses low-pass filters (8-taps FIR) for interpolation, and a detailed spectral analysis may experience some difficulties due to varying filter characteristics. Also, due to spatial aliasing frequencies above 8.3 kHz may be contaminated by sounds from other directions. The individual hydrophones, on the other hand, are calibrated and have a flat frequency response over the band of interest. The lack of array gain limits the detection capabilities, but this was not considered a major problem for

this paper. Hydrophones also lack the angular resolution of the beamformer and are susceptible to more interference from shipping noise and conspecifics.

The method for relating orientation of the tagged whale at the time a click was produced to the level received at the hydrophone required matching clicks from the tag and array. The range between the whale and the array was often great enough that once the whale made one click, it might have made several more clicks before the first click was received at the array. However, the varying sequence of ICI makes it possible to uniquely associate each click recorded on the tag with the same click recorded on the array. The two data sets were matched by generating a waterfall sonar plot of the hydrophone data, in which the start of each trace was triggered using the click times as measured on the tag but shifted by a constant time increment to reduce the effect of travel time. This was analyzed on a dive-by-dive basis, with the time increment selected to correspond to the average range during the dive. Due to varying ranges during each dive, the resulting (hydrophone) clicks were not exactly aligned at the same offset time throughout the dive. In a second step, an automatic click detector similar to the one that extracted the clicks from the tag data was used to estimate the correct click travel time difference. The performance of the click detector could be checked graphically to test whether the clicks on the hydrophone waterfall sonar plot lined up at the same time.

To describe the signal strength, the received peak pressure was estimated for each click by taking the maximum value of the signal envelope. For this, the real-valued signal was first Hilbert transformed to obtain an analytic representation after which the magnitude was estimated as the absolute value of the analytic signal (Randall, 1987). An apparent source level (ASL) was obtained by correcting the received pressure level with the transmission loss for the estimated range between whale and hydrophone, for which we assumed spherical spreading. The maximum ASL of a source defines the axis of the sound beam (acoustic axis) and the ASL in that case gives the source level (SL). Off the acoustic axis the ASL is reduced and its beam pattern characterizes sound emission from the source.

IV. RESULTS

The DTAG was attached for nearly 7 h to a whale (SW01_275b) estimated by visual observation (Miller *et al.*, 2004) to be about 12.2 m long. While tagged, the whale performed eight complete deep dives to foraging depths of 550–900 m (Fig. 2). As the water depth for all dives in this Ligurian Sea site was over 2600 m, the dive pattern indicates that the tagged whale was feeding on a midwater prey layer between 550 and 900 m. While the whale transited from surface to foraging depth and from foraging depth to surface without major points of inflection in the vertical axis, during the bottom portion of its dive, it regularly moved up and down through 200 m of vertical excursion, suggesting that it was exploiting a prey layer about 200 m thick. The whale started to emit regular clicks during descent soon after fluke-up and continued while at foraging depth. While the tagged whale dove, the surface echo of each click was regularly recorded on the tag independent of pitch angle. Figure 3 shows a waterfall sonar plot of the first 350 clicks of a dive. The surface echoes are not clipped and are very sharp at onset. On ascent no acoustic activity was observed, apart from a total of 18 codas at end of dives 1 and 8. The whale only emitted creaks at depths greater than 550 m.

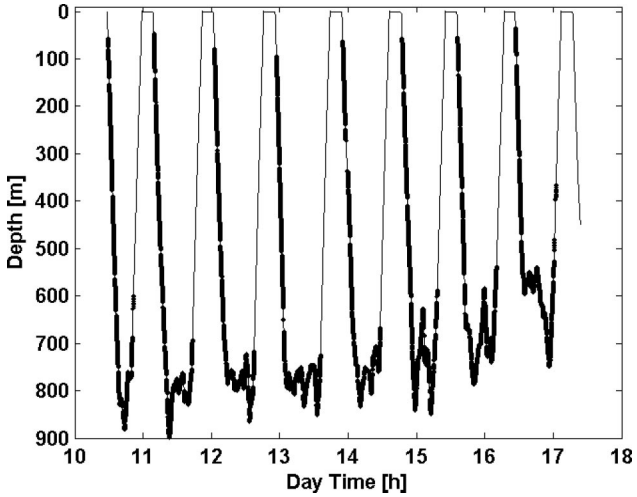


FIG. 2. Depth profile of tagged sperm whale. The whale made eight full dives to depths ranging from 550–900 m. The dots overlaid on the profile indicate the times of click emissions by the whale.

While the whale was tagged, it moved with a mean horizontal speed of 1.22 m/s in north–north–westerly direction. Figure 4 shows the reconstructed track of the whale with color-coded depth profile. The track of NRV ALLIANCE shows that the passive sonar system circled the tagged whale four times at ranges from about 0.7 to 6 km. Due to this circling and the varying pitch of the whale, it was possible to record the whale clicks from nearly all aspects.

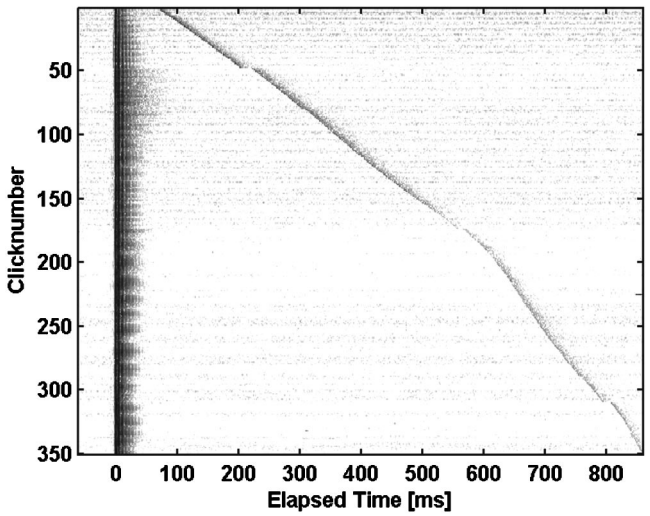


FIG. 3. Waterfall sonar plot of sperm whale clicks recorded from the tag on the whale during descent. Successive clicks are aligned on the left at time $=0$. Surface echoes of the emitted clicks are clearly visible (line with elapsed time increasing to 860 ms at the bottom of the figure). The corresponding depth values are 52 m for the first click and 645 m the 350th click. The depth of the ocean bottom during this dive was over 2600 m and echoes from the ocean bottom are not visible. The small discontinuities visible along the surface echo are due to short pauses within the sequence of regular clicks.

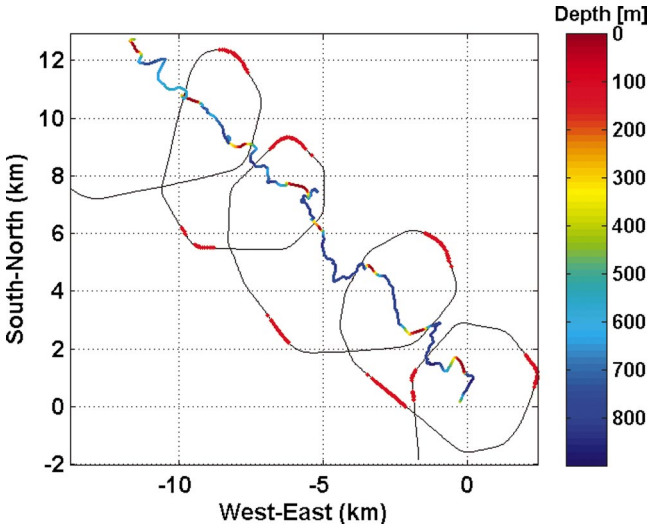


FIG. 4. Reconstruction of the track of the sperm whale tagged in the Ligurian Sea. The track of NRV ALLIANCE is shown in black. The red dots on the ship's track indicate where the visual team sighted the tagged sperm whale. The color-coded line is the track of the animal, where the color map ranges from red for surfacing to blue for a depth of 900 m.

The tag recorded a total of over 36 000 regular and creak clicks. Figure 5 shows the histogram of the ICI of all clicks. The data are plotted on a log–log scale and show a three-mode structure. One mode peaks at ICI of around 1 s and can be attributed to regular clicks. A weaker mode shows a peak around an ICI of 0.25 s and the third mode peaks at an ICI of 0.03 s. For the analysis of this paper we selected only regular clicks with ICI above 0.45 s (vertical dashed line in Fig. 5) and ignored all clicks with ICI below 0.45 that we associated with creaks (Gordon, 1987). According to this threshold, the whale emitted over 14 100 regular clicks during the tag attachment. For this paper, we only used regular clicks where the surface reflection could be distinguished from the direct arrival, reducing the number of clicks to just below 13 000.

The knowledge of the three-dimensional orientation of the whale at each click allowed the transformation of the

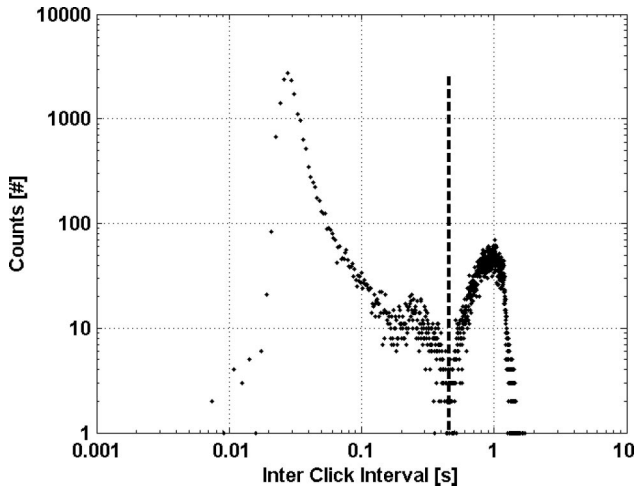


FIG. 5. Log–log plot of distribution of interclick-intervals (ICI) from the tagged whale using all detected clicks ($n>36\,000$). The peak centered on 0.025 s marks the dominant ICI for creak clicks. The peak centered on 1 s marks the dominant ICI for regular clicks. The vertical dashed line at 0.45 s marks the lower limit of ICI we used in this paper to define regular clicks.

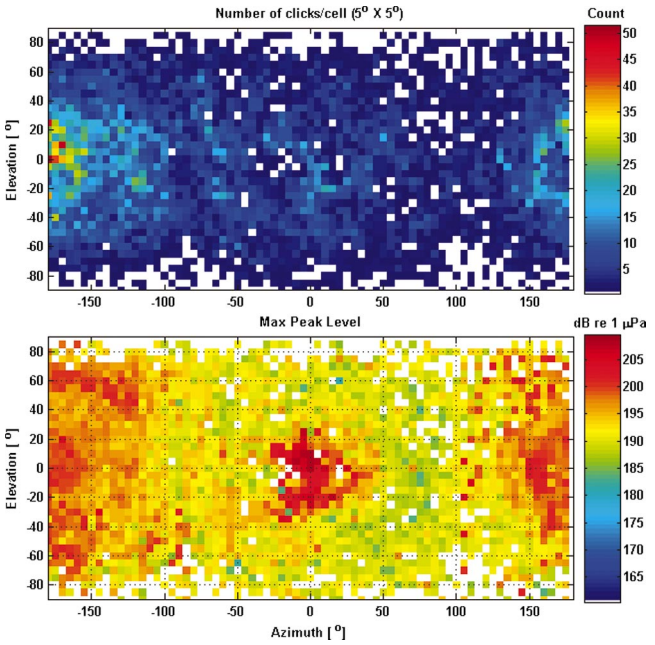


FIG. 6. Aspect coverage and estimated transmit beam pattern of the tagged sperm whale. The pattern is shown as seen from the animal's perspective. The top panel shows the number of clicks falling into each $5°\times5°$ cell. The lower panel shows the maximum apparent source level within each cell. The 3D beam pattern shows two distinct features: a confined concentration of intense clicks in forward directions (around 0° azimuth and 0° elevation) and a more diffuse concentration in backward direction ($\pm180°$ azimuth and 0° elevation). The observed levels reach about 210 dB_{peak} *re*: 1 μPa at 1 m in forward direction and 200 dB_{peak} *re*: 1 μPa at 1 m in backward direction.

coordinates relative to the hydrophone receiver into a whale-frame coordinate system, which made it possible to estimate the azimuth and elevation angles of the whale sound source as seen from the passive sonar. While at depth, the whale rolled and changed pitch from $-90°$ to $+90°$, so a circling passive sonar could record sperm whale clicks from nearly all vertical and horizontal angles. Using the apparent source levels of nearly 13 000 detected regular clicks, we constructed a map of the emitted three-dimensional sound field, or beam pattern of regular sperm whale clicks (Fig. 6).

To obtain this three-dimensional beam pattern, the azimuth and elevation values of all regular clicks were allocated into 5° by 5° cells. The cell size resulted as compromise between expected resolution and the number of obtained histogram samples. The top panel of Fig. 6 shows the coverage of the beam pattern in azimuth and elevation with color-coded counts for the 5° by 5° cells. The center of this plot indicates an azimuth and elevation of 0°. This corresponds to the direction directly in front of the whale. This plot uses a coordinate system as seen by the whale. The point with 0° azimuth and 0° elevation points directly ahead of the whale. Negative azimuth values indicate angles to the left of the whale, positive azimuth to the right. Similarly positive elevation indicates angles above the horizontal, while negative angles are downwards. A zero count (white) means that no click was measured in the cell with corresponding azimuth and elevation. Azimuth and elevation angles of the beam pattern were estimated assuming constant sound speed between whale and array. The impact of ray refraction due to

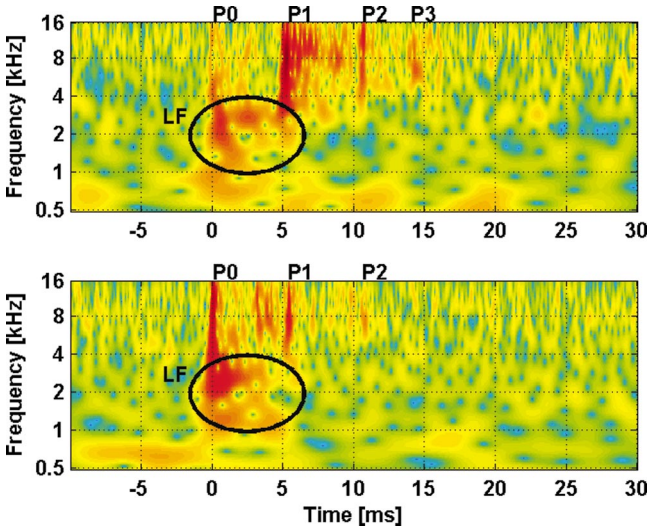


FIG. 7. Wavelet-type spectra (based on Gabor kernel) of forward click (top panel) and backward click (bottom panel) as received by a towed hydrophone. The forward click corresponds to 9° azimuth and 7° elevation in the beam pattern, and the backward click to −137° azimuth and −60° elevation. The individual pulses within each click are labeled P0, P1, P2, and P3, consecutively. The ellipse indicates the low-frequency (LF) component of the click.

depth-dependent variation of the sound-speed velocity (Urick, 1983) varies as a function of range (~−3° at 4 km) and must be considered only at larger ranges (>6 km) and for increased demands on angular accuracy.

The lower panel of Fig. 6 shows the broadband beam pattern of the regular sperm whale clicks, including frequencies from 300–15 000 Hz. To obtain this pattern, the maximum apparent source level (ASL) within each cell was plotted. This approach was necessary because not all of the variations in ASL are due to the click beam pattern, but may also result from variations in click source level (Madsen *et al.*, 2002). By taking the maximum ASL within each cell, a beam pattern is obtained that approximates the maximum level in each direction. This should correspond to the true beam pattern when a sufficient number of clicks are sampled within each cell.

The three-dimensional beam pattern shows two distinct features: a confined concentration of intense clicks in forward direction (around 0° azimuth and 0° elevation) and a more diffuse concentration in backward direction (±180° azimuth and 0° elevation). The observed levels reach about 210 dB_{peak} *re*: 1 μPa at 1 m in forward direction and 200 dB_{peak} *re*: 1 μPa at 1 m in backward direction.

To further understand this bimodal beam pattern, two representative clicks were selected for detailed analysis, one produced when the ship was in front of the whale and the other from when the ship was behind. These two clicks are from the same dive and separated by 3 min. They are consequently from about the same range. Figure 7 shows wavelet spectra (using a Gabor kernel) (Zimmer *et al.*, 2003) of this forward and backward click. The forward click (top panel) was emitted at 10° azimuth, 7° elevation and is characterized by a weak, 4-ms-long component (0 to 4 ms) at frequencies around 2.5 kHz, a very weak short pulse, called P0, with frequencies above 3 kHz at 0 ms, a strong pulse, called P1, with frequencies above 3 kHz at 5.3 ms, followed by a weaker pulse, P2 at 10.6 ms, and an again weaker pulse, P3 at 14 ms. The backward click (lower panel) was emitted at −137° azimuth, −60° elevation, and shows the same low frequency (LF) component (0 to 4 ms around 2.5 kHz) as seen in the forward click. But, the high-frequency part is different. The short P0 pulse at 0 ms is now the strongest pulse, followed by a weaker P1 pulse around 5.3 ms and a very weak P2 pulse at about 10.6 ms. Both clicks are characterized by an IPI of 5.3 ms, which corresponds to a whale of just over 12.5 m: 12.53 m using the formula of Gordon (1991): $L=4.833+1.453(\mathrm{IPI})-(\mathrm{IPI}/1000)^2$, and 12.57 m using the formula of Rhinelander and Dawson (2004): $L=17.12-2.189(\mathrm{IPI})+0.251(\mathrm{IPI})^2$. This is a close match to the 12.2 m estimated visually. In both clicks, the pulse structure occurs only at higher frequencies and there is significant reverberation following the strongest pulse. The LF component of the clicks starts with the first pulse P0 and its duration remains shorter than the IPI. No pulse-like repetition of the low-frequency component is observed.

Figure 7 shows that the individual clicks are too complex to support a single beam pattern. There are at least three different characteristic features: a low-frequency (LF) component common to both the forward- and backward-directed parts of the click, the forward-directed portion dominated by P1, and the backward-directed part dominated by P0. We therefore reprocessed the data with narrower time and spectral windows to separate the beam patterns of the LF component, the P0 and the P1 pulses. For the following analysis the window definitions were as follows: The LF component was estimated with a spectral window from 300 Hz to 3 kHz and a time window from −2 to 10 ms. The P0 component was defined to fall within 3 and 15 kHz and between −2 and 3 ms. The P1 component was defined to fall again within 3 and 15 kHz but between 3 and 8 ms.

Figure 8 shows peak-level beam pattern of the LF component, the P0 and P1 pulses. The figures confirm a nearly omnidirectional LF component with maximum levels ranging from ~170–190 dB_{peak} *re*: 1 μPa at 1 m, the P0 beam pattern pointing backwards with low directionality at maximum levels of ~200 dB_{peak} *re*: 1 μPa at 1 m and the P1 component pointing forward with high directionality and a maximum measured level of 210 dB_{peak} *re*: 1 μPa at 1 m.

To simplify the graphic representation of the beam pattern and to get a better feeling for the directionality, we observe that all three beam patterns exhibit rotational symmetry. It therefore makes sense to plot the peak level estimated as function of the off-axis angle, which is the angle between the forward direction of the acoustic axis and any other combination of azimuth and elevation.

Figure 9 shows the peak level as function of off-axis angle for the three components of a sperm whale click (LF component: top panel; P0 pulse middle panel, P1 pulse bottom panel). All panels show in gray all of the peak-level measurements and in black the 90th percentile of the measurements for a given off-axis angle. Again, we observe that the LF component is nearly independent of the off-axis angle; the P0 pulse increases slowly with the off-axis angle and reaches a maximum close to 180°. The P1 pulse de-

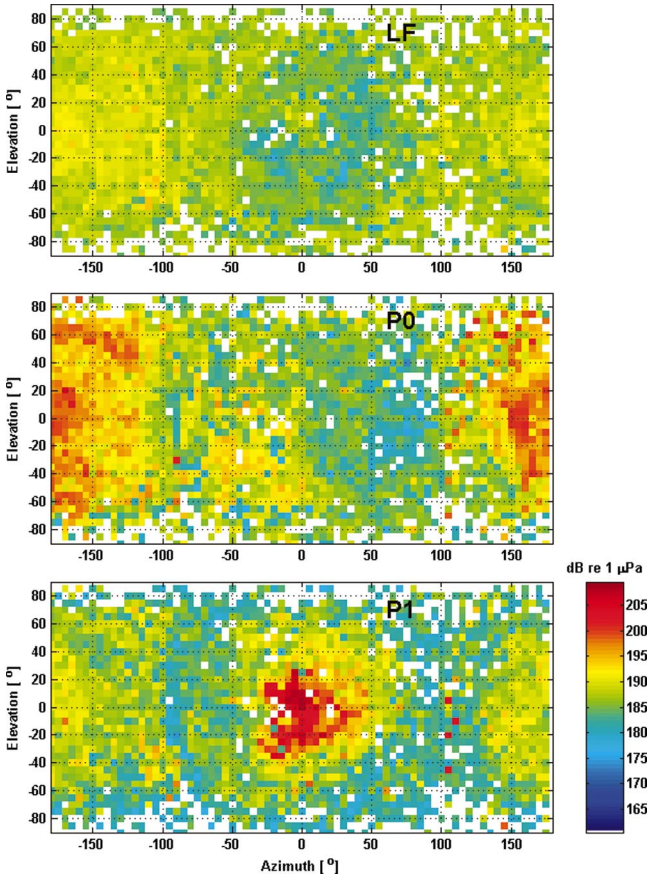


FIG. 8. Map of transmit beam pattern of three components of clicks recorded from the tagged sperm whale. The pattern is shown as seen from the animal's perspective. Each pixel indicates the maximum apparent source level (ASL) recorded within each cell. The LF component is nearly omnidirectional, with click ASLs ranging from 170–190 dB_{peak} *re*: 1 μPa at 1 m. The P0 component has a broad beam directed backward (near $\pm 180°$, with ASLs in the beam of about 200 dB_{peak} *re*: 1 μPa at 1 m. The P1 beam has maximum ASLs near 210 dB_{peak} *re*: 1 μPa at 1 m.

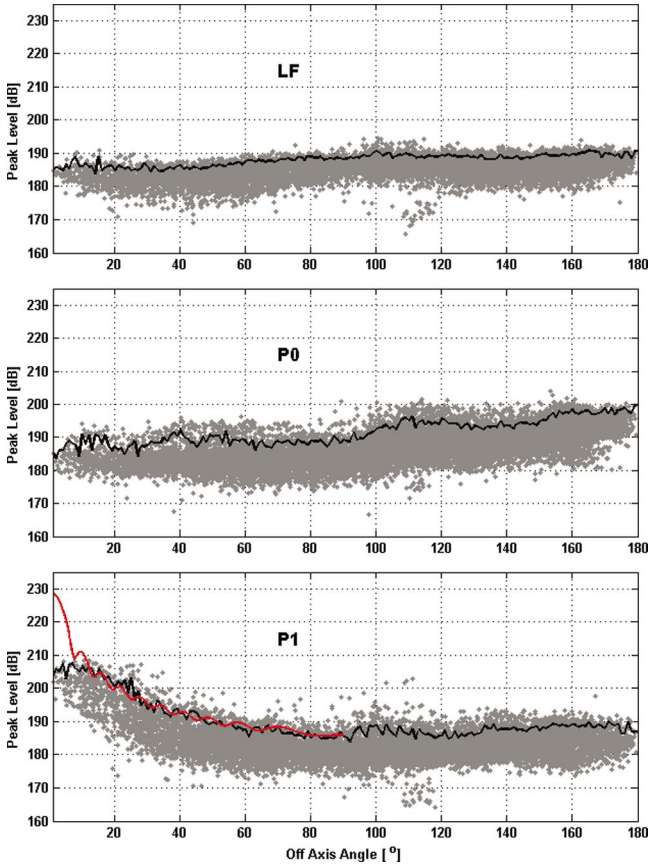


FIG. 9. Beam patterns of click components as a function of off-axis angle. The gray dots are a scatter plot of each measured value. The black line represents the 90th percentile of level for each off-axis angle. For P1, the red line indicates the beam pattern predicted for a circular piston source with the parameters indicated in the text, and is fitted to off-axis angles $>20°$ with a maximum on-axis click source level of 229 dB_{peak} *re*: 1 μPa at 1 m. Clicks with off-axis angles $<20°$ were not used in this fit and the resulting differences are discussed in the text.

creases with the off-axis angle and remains constant beyond 90°. The standard deviation of peak level was found to be nearly constant at 3.5 dB as a function of off-axis angle for all three components.

The red curve in the lower panel was modeled as broadband beam pattern for the P1 pulse and is based on a short, Gaussian-shaped pulse emitted from circular piston (Au, 1993) and fitted to the 90th percentile for angles $>20°$. The obtained modeled source parameters for the P1 pulse were center frequency 13 kHz, signal duration 0.21 ms, piston radius 0.55 m, and maximum source level 229 dB_{peak} *re*: 1 μPa at 1 m. Replacing the 90th percentile values for angles $<90°$ with the modeled beam pattern, the broadband directivity index for the P1 pulse became 26.7 dB, which is close to the 27 dB given by Møhl *et al.* (2003). For comparison, the measured broadband directivity index (based on the 90th percentile) of the backward beam (P0) is 7.4dB.

To investigate the origin of the LF component, the peak frequency of the LF spectrum of each pulse was extracted. If the LF component were based on air resonance, then one would expect that the peak frequency would vary as a function of dive depth due to the compression of available air volume. Under certain assumptions (wavelength larger than the linear dimension of the air volume, speed of sound nearly constant) theory predicts that the resonance frequency should vary with the square root of the static pressure (Kinsler *et al.*, 1982; Medwin and Clay, 1998). Figure 10 shows a scatter plot where the dominant peak frequency in the LF component is plotted against the square root of the static pressure. For pressures up to 53 atm (vertical dashed line; equivalent to 520 m) one can observe two modes for which the peak frequency varies linearly with square root of pressure. For pressures greater 53 atm, or higher depth values, this relationship breaks down. The two black lines correspond to the modeled depth dependence of an air resonator ($f_{1,0} = 150$ Hz and $df_1/d\sqrt{p} = 160$ Hz/$\sqrt{\text{atm}}$ for the lower line and $f_{2,0} = 50$ Hz and $df_2/d\sqrt{p} = 260$ Hz/$\sqrt{\text{atm}}$ for the steeper line).

Figure 10 was generated by plotting a single peak LF frequency for each click. Consequently, the scatter plot does not reflect completely the presence of the two modes in each click, but only the stronger of the two modes within each click. However, on a statistical basis both modes got a chance to dominate.

The two lines in Fig. 10 cross at surface pressure, indicating a single (or similar) volume at surface with a reso-

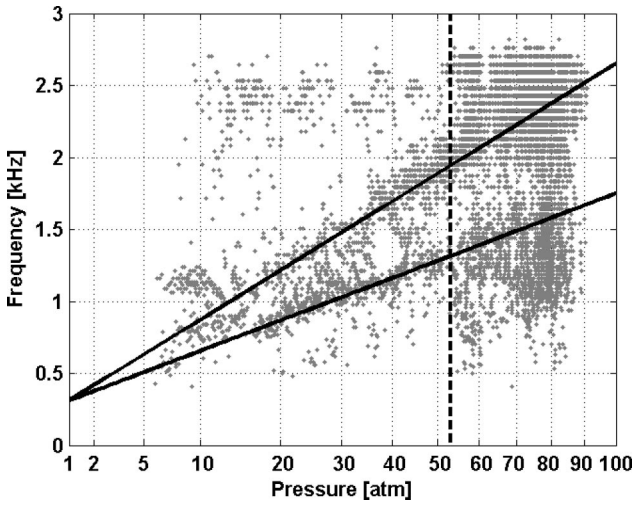


FIG. 10. Variation of the peak frequency of the LF component as a function of pressure. The two solid lines indicate two modes of resonance frequency and indicate that the resonance frequencies vary with the square root of pressure for pressures up to 53 atm (or 520 m; vertical dashed line).

nance frequency of 310 Hz. We do not know the exact mechanism behind this possible air resonance so we only can speculate on the air volume required for this resonance frequency. However, to get a feeling for the dimension of air volume, two models were investigated: free-oscillating air bubbles (Kinsler *et al.*, 1982; Medwin and Clay, 1998) and a Helmholtz resonator (Kinsler *et al.*, 1982). For a freely oscillating bubble, a volume of 4.2 ml is required for a resonance frequency of 310 Hz at 1 atm. To obtain the volume of a Helmholtz resonator, the dimensions of the neck must be specified. As we only are interested in the order of magnitude, we assume a neck with diameter of 8.6 cm and an effective length of 2.6 cm. A neck diameter of 8.6 cm corresponds to the aperture that is consistent with the beam pattern of the P0 pulse, and the effective neck length of 2.6 cm is equivalent to the end corrections of a zero-length neck in air. With these assumptions we obtain a resonance air volume of 6400 ml.

Even if the estimated volumes are only indicative, they show clearly that the resonance frequency depends on the assumed physical model. An open resonator, like the Helmholtz model, requires a larger air volume than a closed resonator, or air bubble. A more accurate estimate of air volume must use more detailed knowledge of the anatomy of sound production in the sperm whale, especially the location of air during the dive.

V. DISCUSSION

Our paper is based on a unique data set from a sperm whale tagged in the Ligurian Sea in 2001. While tagged, the whale emitted over 36 600 clicks, of which over 14 100 are regular clicks. Nearly 13 000 regular clicks were used for estimating the three-dimensional beam pattern. The orientation of the whale was obtained from the accelerometer and magnetometer values from the DTAG. After synchronizing the clocks of the DTAG and the passive sonar towed from the NATO RV ALLIANCE, the click travel time from the whale to the hydrophone array was used to estimate the acoustic range of the whale during its dives. The acoustic bearing was obtained by correlating two hydrophones separated by 10.62 m.

This method, which is totally independent of the technique applied by Møhl and co-workers to discover the high directionality of sperm whale clicks (Møhl *et al.* 2000, 2003), confirms the basic conclusion that most of the energy of regular clicks is directed forward in a high-powered narrow beam. Møhl *et al.* (2000, 2003) measured the beam pattern using multiple dispersed receivers to record the same clicks simultaneously. We measured the peak levels of nearly 13 000 clicks at varying aspects to a single receiver. Our analysis assumed that the source spectra, and therefore the beam pattern, are similar enough for all of these clicks to allow integration of the measurements into a single beam pattern. The technique used by Møhl *et al.* (2000, 2003) was unable to define the orientation of the clicking whale with respect to the hydrophones. Our analysis was able to do this, and confirms their conclusion that the main acoustic axis is aligned forward near an azimuth of 0° and an elevation of 0° with respect to the whale's body.

Møhl *et al.* (2003) reported a maximum source level on the axis of the P1 pulse of up to 235 dB_{rms} *re*: 1 μPa at 1 m. Our data for the P1 pulse match a modeled beam pattern for off-axis angles $>20°$ [Fig. 9(lower panel)] with source levels of 229 dB_{peak} *re*: 1 μPa at 1 m. When comparing our results with the values given by Møhl *et al.* (2003), one should note that they present the source level as rms value, while we have chosen to use peak values. This allowed us to use data with low signal-to-noise ratio (SNR) for the beam pattern, whereas the rms estimation would have been difficult at low SNR. As rule of thumb, we estimate that, for on-axis sperm whale clicks, rms values would be 5 dB lower than peak levels of the analytic signal representation of this paper. For off-axis angles $<20°$ our measurements only extend up to 210 $dB_{0\text{-}p}$ *re*: 1 μPa at 1 m, well below the maximum value of the modeled beam pattern and Møhl's results. Part of this discrepancy may stem from Møhl *et al.*'s (2000, 2003) selection of the few clicks with the highest apparent source levels from weeks of sampling large males in polar waters compared to our analysis of all clicks from a 12-m whale in the Mediterranean. This discrepancy may also stem from three limitations in our data: our limited sampling of on-axis clicks, limitations in the bandwidth of our sampling, and clipping of received levels >140 dB_{peak} *re*: 1 μPa. The number of points we sampled near the axis of the beam is relatively small: $352<20°$, $116<10°$, $37<5°$, and $2<2°$. The expected number of clicks/degree is 13 000/180, or ~ 70; the low numbers of clicks sampled on axis probably reflects the tendency for the whale not to pitch upwards while clicking. Given the small sample size, it is very likely that we missed the strongest clicks with off-axis angles less than $2°-5°$. We also underestimate by some dB the level of each click due to the limited bandwidth of our data that cover only $\frac{1}{2}$ of the bandwidth of on-axis version of the P1 pulse in sperm whale clicks (Møhl *et al.*, 2000, 2003). Finally the gain settings of the passive sonar system limited the received level to 140 dB_{peak} *re*: 1 μPa. To verify how many P1 clicks are clipped, the received level of each click with off-axis

angle less than 20° was inspected: 131of these clicks were recorded within 3 dB of the clipping level, indicating that for about 1/3 of the near-axis measurements clipping may have occurred.

This paper is the first to use a method capable of separating the P0 and P1 pulses based upon orientation of the whale with respect to the receiver. Our data show a difference of 39 dB between the modeled source level of P1 and the measured ASL of P0 in forward direction. This difference describes the processing gain of the sperm whale sound production system. Comparing the directivity indices for both the P1 and P0 beam, pattern we obtain a gain of the ''bent horn'' of about 19 dB and an effective source level of an initial omnidirectional pulse source of about 210_{peak} dB re: 1 μPa at 1 m. The modeled values for the source levels are only indicative and are partially based on extrapolation and rough model assumptions. Nevertheless, these values are consistent with recent results of Møhl *et al.* (2003).

The dimension of the frontal air sac is about 1 m, corresponding to a good sound reflector for frequencies over 3 kHz (at 3 kHz, $2\lambda \approx 1$ m). As the spermaceti organ is most likely neither a perfect waveguide nor the frontal air sac a perfect reflector for frequencies below 3 kHz, significant lower frequency (LF) sound energy should leave the whale in all directions. Physics can explain why we see an omnidirectional LF frequency component for all sperm whale clicks, but this does not explain how and why this LF frequency component is generated in the first place.

The LF component always immediately follows the P0 pulse and has a long duration, with peak frequencies that are depth dependent down to over 500 m. The lack of repeated pulse structures in the LF component after the P1 pulse indicates that the P1 pulse does not reinforce the resonance as it travels from the frontal sac through the junk. We propose that the initial pulse (P0) generated by the phonic lips activates air volumes connected to the phonic lips, which generates the LF component. The two dominant frequencies in the LF component indicate either one resonator with aspect-dependent radiation patterns or that two resonators exist with similar volumes at the surface but different rates at which the volumes are reduced by increasing static pressure. The prediction that resonance frequency should vary with the square root of the static pressure fits well for depths of <520 m, but not for deeper depths. The reasons for the breakdown of the air resonance model at depth larger than 520 m are not known.

Our data show that sperm whale clicks are composed of three components with different characteristics: P0, P1, and LF. We suggest that they are all generated by the same acoustic event, the generation of a short pulse at the phonic lips. This pulse excites a low-frequency resonance in adjacent gas cavities that radiates nearly omnidirectionally. The initial pulse itself is mostly directed into the spermaceti organ, but some energy ($<10\%$) leaks as broad backward P0 beam. If the source level of the P1 pulse has a source level of 230 +dB dB_{peak} re: 1 μPa at 1 m as indicated by our model results and as reported by Møhl *et al.* (2003), and the corresponding P0 pulse has a source level of 200 dB_{peak} re: 1 μPa at 1 m, then the P0 pulse may have 0.1% of the energy of the P1 pulse. Most of the energy of the initial backward-directed pulse reflects forward off the frontal sac into the junk and leaves the junk as a narrow, forward-directed P1 pulse. A small fraction of that energy is reflected by the frontal sac back into the spermaceti organ to generate higher-order pulses (Norris and Harvey, 1972; Møhl *et al.*, 2003).

Due to its high directionality, the forward-directed P1 pulse is well suited for echolocation as demonstrated by Møhl and co-workers (Møhl *et al.*, 2003; Møhl, 2001; Madsen, 2002; Madsen *et al.*, 2002). The high source level of the P1 pulse and the long ICI of regular clicks suggest a potential for long detection ranges. Working in areas where the water depth was <1000 m, Thode *et al.* (2002) correlated the ICI of sperm whales on the descent phase and the round-trip travel time to the seafloor, confirming earlier reports that sperm whales may time their clicks to include echoes from the seafloor. On descent, the ICI of sperm whale regular clicks usually have a curious oscillation. Zimmer *et al.* (2003) demonstrated that this oscillation can correlate with the pitch of a tagged whale. These data stem from a whale tagged in water 2600 m deep, where the ICI never was long enough to correspond to the round-trip travel time to the seafloor, but does correlate with the depth at which the whale will forage. As the whale changes pitch from straight down to more horizontal, the ICI increases proportional to the slant range of the narrow forward beam to the maximum depth of his foraging dive (Zimmer *et al.*, 2003). This implies that the whale clicks only after the echo from the forward beam returns from the prey layer at which the whale will feed. The correlation of ICI with pitch of the whale would not occur if the whale were listening for echoes from an omnidirectional click. While at depth, sperm whale ICI typically range from 0.5–1.5 s, corresponding to round-trip travel times that are equivalent to maximum sonar ranges of 375–1125 m, assuming that whales do not continue listening for echoes of a previous click once they produce the next one (Au, 1993). It is possible that sperm whales are echolocating for prey at these long detection ranges, but the long ICI may also be used to maintain an overview of the entire auditory scene, similar to bats (Moss and Surlykke, 2001).

While it has not been addressed before, the backward-oriented beam pattern of the P0 is a necessary consequence of the bent-horn model of sound production for sperm whale clicks, and therefore the beam pattern we have described for the P0 supports the bent-horn model. Considering that the source level of the P0 is 1–3 orders of magnitude weaker than the P1 pulse, one could argue that the existence of the P0 is only a by-product of the generation of the P1 pulse, and that a backward-oriented beam has no special functionality. On the other hand, the absolute source level of the backward beam is high enough that significant echoes are received on the tagged whale from directions outside of the forward beam. For example, while the whale was descending, the hydrophone on the DTAG usually picked up distinct reflections from the surface (Fig. 3). Echoes from the surface are likely to provide a useful orientation cue, acting like a constant ''acoustic horizon,'' as well as informing the whale of depth, which is important in timing its ascent. Reception of sound from behind requires hearing capabilities in backward

direction. As a complete three-dimensional receiving beam pattern is not known for sperm whales, one can only speculate that in analogy to dolphins (e.g., *Tursiops truncatus*; Au, 1993), sperm whales have a frequency-dependent receiving beam pattern with equal or less directionality than the transmit beam pattern. Given the signal-to-noise ratio of surface echoes on the tag, it would be surprising if the whale could not hear the surface echoes of its own click and use that information. The ICI of the regular clicks is frequently long enough to allow reception of surface echoes before the next click is emitted. The whale may time the ICI in order to receive information from all relevant elements of its auditory scene, including surface and bottom.

If our interpretation that the LF component is produced by a resonating gas volume stimulated by the initial pulse is correct, then this supports the hypothesis that the initial pulse energy is generated by a pneumatic mechanism. Thode *et al.* (2002) reported a shift in frequencies in the 10–15-kHz band of sperm whale clicks as a function of depth, and proposed that this was caused by changing resonant frequencies of reducing air volumes. They suggested that the high-frequency sounds they were analyzing might represent ''generator noise,'' or incidental sounds that are by-products of the sound generation system. Before ruling out any functionality of the LF component, one should note that this low-frequency component with its low directionality often travels furthest and allows the detection over long distances. The long-range propagation would allow conspecifics to listen over long distances of more than 15 km to the acoustic activities of sperm whales (Madsen *et al.*, 2002). This LF component is the most likely signal that enables separated sperm whales to synchronize their diving behavior (Whitehead, 1996).

In summary, regular clicks of sperm whales have several components by which the whale simultaneously produces a narrow, high-frequency sonar beam to search for prey, a less-directional backward pulse, by-product or not, that provides orientation cues, and a low-frequency component of low directionality that conveys sound to a large part of the surrounding water column with a potential for reception by conspecifics at considerable ranges.

ACKNOWLEDGMENTS

This work was carried out within the Acoustic Risk Mitigation Program of NATO Undersea Research Center. Thanks to the science party and ship's crew on the research cruise during which these data were collected. The whale was tagged under a scientific research permit No. 981-1578-00 issued by the U.S. National Marine Fisheries Service to Peter L. Tyack. This is contribution No. 11019 from the Woods Hole Oceanographic Institution. The Woods Hole Oceanographic Institution Animal Care Use Committee approved this research. This work was funded by grants from the Office of Naval Research Grants N00014-99-1-0819 and N00014-01-1-0705, and the Packard Foundation.

Abraham, D. A. (**2000**). ''Passive Acoustic Detection of Marine Mammals,'' SM-351 (SACLANT Undersea Research Centre, La Spezia, Italy).

Au, W. W. L. (**1993**). *The Sonar of Dolphins* (Springer, New York).

Backus, R. H., and Schevill, W. E. (**1966**). ''*Physeter* Clicks,'' in *Whales, Dolphins and Porpoises*, edited by K. S. Norris, (University of California Press, Berkeley), pp. 510–527.

Fristrup, K. M., and Harbison, G. R. (**2002**). ''How do sperm whales catch squids?'' Marine Mammal Sci. **18**, 42–54.

Gordon, J. C. D. (**1987**). ''Sperm whale groups and social behaviour observed off Sri Lanka,'' Ph.D. thesis (University of Cambridge, UK).

Gordon, J. C. D. (**1991**). ''Evaluation of a method for determining the length of sperm whales (*Physeter catodon*) from their vocalizations,'' J. Zool. **224**, 310–314.

Gordon, J., and Tyack, P. L. (**2002**). ''Acoustic techniques for studying cetaceans,'' in *Marine Mammals: Biology and Conservation*, edited by P. G. H. Evans and T. Raga (Plenum, London).

Griffin, D. R. (**1958**). *Listening in the Dark* (Yale University Press, New Haven).

Jaquet, N., Dawson, S., and Douglas, L. (**2001**). ''Vocal behavior of male sperm whales: Why do they click?'' J. Acoust. Soc. Am. **109**, 2254–2259.

Johnson, M., and Tyack, P. L. (**2003**). ''A digital acoustic recording tag for measuring the response of wild marine mammals to sound,'' IEEE J. Ocean. Eng. **28**, 3–12.

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

Kinzey, D., and Gerrodette, T. (**2001**). ''Conversion factors for binocular reticules,'' Marine Mammal Sci. **17**, 353–361.

Madsen, P. T. (**2002**). ''Sperm whale sound production,'' PhD. dissertation (University of Aarhus).

Madsen, P. T., Wahlberg, M., and Møhl, B. (**2002**). ''Male sperm whale (*Physeter macrocephalus*) acoustics in a high latitude habitat: Implications for echolocation and communication,'' Behav. Ecol. Sociobiol. **53**, 31–41.

Madsen, P. T., Payne, R., Kristiansen, N. U., Wahlberg, M., Kerr, I., and Møhl, B. (**2002**). ''Sperm whale sound production studied with ultrasound time/depth-recording tags,'' J. Exp. Biol. **205**, 1899–1906.

Madsen, P. T., Carder, D. A., Au, W. W. L., Nachtigall, P. E., Møhl, B., and Ridgway, S. H. (**2003**). ''Sound production in neonate sperm whale,'' J. Acoust. Soc. Am. **113**, 2988–2991.

Medwin, H., and Clay, C. S. (**1998**). *Fundamentals of Acoustic Oceanography* (Academic, San Diego).

Miller, P. J. O., Johnson, M. P., Tyack, P. L., and Terray, E. A. (**2004**). ''Swimming gaits, passive drag and buoyancy of diving sperm whales *Physeter macrocephalus*,'' J. Exp. Biol. **207**, 1953–1967.

Møhl, B. (**2001**). ''Sound transmission in the nose of the sperm whale *Physeter catodon*. A postmortem study,'' J. Comp. Physiol., A **187**, 335–340.

Møhl, B., and Andersen, S. (**1973**). ''Echolocation: High-frequency component in the click of the harbor porpoise (*Phocoena ph. L.*),'' J. Acoust. Soc. Am. **54**, 1368–1372.

Møhl, B., Wahlberg, M., Madsen, P. T., Heerfordt, A., and Lund, A. (**2000**). ''Sperm whale clicks: Directionality and source level revisited,'' J. Acoust. Soc. Am. **107**, 638–648.

Møhl, B., Wahlberg, M, Madsen, P. T., Heerfordt, A., and Lund, A. (**2003**). ''The monopulsed nature of sperm whale clicks,'' J. Acoust. Soc. Am. **114**, 1143–1154.

Moore, K. E., Watkins, W. A., and Tyack, P. L. (**1993**). ''Pattern similarity in shared codas from sperm whales (*Physeter catodon*),'' Marine Mammal Sci. **9**, 1–9.

Moss, C., and Surlykke, A. (**2001**). ''Auditory scene analysis by echolocation in bats,'' J. Acoust. Soc. Am. **110**, 2207–2226.

Mullins, J., Whitehead, H., and Weilgart, L. (**1988**). ''Behaviour and vocalization of two sperm whales, *Physeter macrocephalus*, off Nova Scotia,'' Can. J. Fish. Aquat. Sci. **45**, 1736–1743.

Norris, K. S., and Harvey, G. W. (**1972**). ''A theory for the function of the spermaceti organ of the sperm whale (*Physter catodon L.*),'' in *Animal Orientation and Navigation*, edited by S. R. Galler, K. Schmidt-Koenig, G. J. Jacobs, and R. E. Belleville, SP-262 (NASA, Washington D. C.), pp. 397–417.

Page, S. E. (**1954**). ''Continuous inspection schemes,'' Biometrika **41**, 100–115.

Randall, R. B. (1987). *Frequency Analysis* (Brüel & Kjær, Nærum, Denmark).

Rendell, L. E., and Whitehead, H. (**2003**). ''Vocal clans in sperm whales (*Physeter macrocephalus*),'' Proc. R. Soc. London, Ser. B **270**(1512), 225–231.

Rhinelander, M. Q., and Dawson, S. M. (**2004**). ''Measuring sperm whales

from their clicks: Stability of interpulse intervals and validation that they indicate whale length,'' J. Acoust. Soc. Am. **115**, 1826–1831.
Thode, A., Mellinger, D. K., Stienessen, S., Martinez, A., and Mullin, K. (**2002**). ''Depth dependent acoustic features of diving sperm whales (*Physeter macrocephalus*) in the Gulf of Mexico,'' J. Acoust. Soc. Am. **112**, 308–321.
Tyack, P. L. (**1997**). ''Studying how cetaceans use sound to explore their environment,'' Perspec. Ethol. **12**, 251–297.
Urick, R. J. (**1983**). *Principles of Underwater Sound* (McGraw-Hill, New York).
Wald, A. (**1947**). *Sequential Analysis* (Wiley, New York).
Watkins W. (**1980**). ''Acoustics and the behavior of sperm whales,'' in *Animal Sonar Systems*, edited by R. Busnel and J. F. Fish (Plenum, New York), pp. 291–297.
Watkins, W. A., and Daher, M. A. (**2004**). ''Variable spectra and nondirectional characteristics of clicks from near-surface sperm whales (*Physeter catodon*),'' in *Echolocation in Bats and Dolphins*, edited by J. A. Thomas, C. F. Moss, and M. Vater (The Univeristy of Chicago Press, Chicago), pp. 410–413.
Watkins, W. A., and Schevill, W. E. (**1977**). ''Sperm whale codas,'' J. Acoust. Soc. Am. **62**, 1485–1490.
Weilgart, L., and Whitehead, H. (**1988**). ''Distinct vocalizations from mature male sperm whales (*Physeter macrocephalus*),'' Can. J. Zool. **66**, 1931–1939.
Weilgart, L., and Whitehead, H. (**1993**). ''Coda communications by sperm whales (*Physeter macrocephalus*) off the Galapagos Islands,'' Can. J. Zool. **71**, 744–752.
Whitehead, H., and Weilgart, L. (**1991**). ''Pattern of visually observable behavior and vocalizations in groups of female sperm whales,'' Behaviour **118**, 275–296.
Whitehead, H. (**1996**). ''Babysitting, dive synchrony, and indication of alloparental care in sperm whales,'' Behav. Ecol. Sociobiol. **38**, 237–244.
Zimmer, W. M. X., Johnson, M. P., D'Amico, A., and Tyack, P. L. (**2003**). ''Combining data from a multisensor tag and passive sonar to determine the diving behavior of a sperm whale (*Physeter macrocephalus*),'' IEEE J. Ocean. Eng. **28**, 13–28.

Indication of a Lombard vocal response in the St. Lawrence River beluga

P. M. Scheifele, S. Andrew, R. A. Cooper, and M. Darre
Animal Science Department, University of Connecticut, Storrs, Connecticut 06269

F. E. Musiek and L. Max
Communications Science Department, University of Connecticut, Storrs, Connecticut 06269

(Received 2 August 2004; revised 27 October 2004; accepted 1 November 2004)

Noise pollution is recognized as a potential danger to marine mammals in general, and to the St. Lawrence beluga in particular. One method of determining the impacts of noise on an animal's communication is to observe a natural and repeatable response of the vocal system to variations in noise level. This is accomplished by observing intensity changes in animal vocalizations in response to environmental noise. One such response observed in humans, songbirds, and some primates is the Lombard vocal response. This response represents a vocal system reaction manifested by changes in vocalization level in direct response to changes in the noise field. In this research, a population of belugas in the St. Lawrence River Estuary was tested to determine whether a Lombard response existed by using hidden Markhov-classified vocalizations as targets for acoustical analyses. Correlation and regression analyses of signals and noise indicated that the phenomenon does exist. Further, results of human subjects experiments [Egan, J. J. (1966), Ph.D. dissertation; Scheifele, P. M. (2003), Ph.D. dissertation], along with previously reported data from other animal species, are similar to those exhibited by the belugas. Overall, findings suggest that typical noise levels in the St. Lawrence River Estuary have a detectable effect on the communication of the beluga. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1835508]

PACS numbers: 43.80.Nd, 43.80.Ka, 43.80.Lb [WWA] Pages: 1486–1492

I. INTRODUCTION

The St. Lawrence River Estuary is habitat to a sub-Arctic population of beluga whales on a year-round basis. This region is also a mainstream route for commercial shipping and, in the last 10 years, has become the primary region of eco-tourism activities primarily consisting of whale watching. A great debate continues regarding whether or not these activities have any effect on the hearing and communication abilities of these animals. One mechanism that can be used to determine whether noise is having an effect on an animal's ability to communicate is to observe some natural and repeatable response of the vocal system in response to changes in noise level. The fact that an animal has to alter its vocalization level in the presence of anthropogenic noise is indicative that its vocalizations are being influenced by that noise, possibly with long-term adverse energetic consequences. Vocal changes in response to noise may also impede normal auditory feedback or ''sidetone'' levels (Lane and Tranel, 1971; Lombard, 1911).

One natural reaction such as this has been observed in humans and is known as the Lombard vocal response (Lombard, 1911). The Lombard vocal response (also known as the Lombard effect or reflex) represents a reaction of the vocalization system directly manifested by changes in vocalization level (Egan, 1966) and refers to a noise-induced phenomenon and the unconscious tendency of a person or animal to raise their voice when confronted with a noisy environment. The underlying principle is the maintenance of the normally expected loudness of the vocalizer's sidetone. Measuring the Lombard response allows the study of the communication system in an integrated manner. It may also be used as an indicator of noise effects on animal communication. While exhibiting a Lombard response provides a mechanism for animals to cope with varying levels of noise, the response is also indicative of the animal attempting to cope with noise levels that are potentially rising toward a point where masking will occur. This level is the ceiling of the Lombard response. During the process of responding to elevated levels of noise, the animal is also expending more energy than normal to achieve total communication.

Acoustic communication relies on the integrated and interdependent functioning of the auditory and vocal systems (Levelt, 1989; Bradbury and Vehrencamp, 1998). The auditory subsystem plays a pivotal role as an external feedback loop in the overall ability of the animal to communicate. With no external feedback loop, the ability to properly construct, deliver, and process sounds is severely reduced (Guenther, 2001).

The Lombard vocal response is a phenomenon not limited to humans; it is known to occur in monkeys, bats, cats, quail, nightingales, and budgerigars (Potash, 1972; Sinott *et al.*, 1975; Manabe *et al.*, 1998; Egnor *et al.*, 2003; Cynx *et al.*, 1998); however, studies have not been made to determine whether it occurs in marine mammals. Observation of this response is a critical step in the analysis of vocalization-in-noise studies and the study of the general dynamic relationship between auditory feedback and acoustic communication, especially under conditions of altered auditory feedback (Lane and Tranel, 1971). Table I shows gross comparative results of Lombard tests on various animals and humans. Although the Lombard response has not been tested in

TABLE I. Human/animal Lombard response comparisons showing the level of the vocalization during a Lombard response per dB increase in noise level.

Animal	Increase in Lombard response as function of increase noise level (dB)	Reference
Human	1.3	Egan, 1966; Lane and Tranel, 1971
Human	1.09	Scheifele, 2003
Tamarin	2.8	Egnor *et al.*, 2003
Macaque	2.0	Sinott *et al.*, 1975
Zebra finch	3.3	Cynx *et al.*, 1998
Quail	0.60	Potash, 1972
Budgerigar	0.35	Manabe *et al.*, 1998
Cat	1.8	Nonaka *et al.*, 1997
Beluga	1.0	Scheifele, 2003

many animals, Table I illustrates a range of response levels in mammals and songbirds. The current study was conducted with a group of sub-Arctic beluga whales in the St. Lawrence River Estuary to investigate whether a vocalization-as-a-function-of-noise response exists in the St. Lawrence beluga whale.

II. METHODS

A. General methodology

Vocalizations of subgroups of the population of 700 St. Lawrence River belugas were collected at different sites of the upper estuary where these whales congregate during the summer. A selection of these sites, all near the confluence of the Saguenay River and the St. Lawrence Estuary, was made in view of the

(1) Different vessel traffic and use causing individual background noise intensities;
(2) Regular use by different social groups of belugas during the same portion of the summer range (in an effort to reduce confounding factors due to differences among whales of different social groupings and/or of different areas);
(3) Intrinsic quality of the site's acoustical environments (topography, depth); and
(4) Proximity to one another and, hence, ability to sample them numerous times during a single day.

Site 1. Saguenay site (latitude: 48°07.34′, longitude: 69°41.40′) is located approximately 1 km outside of the harbor of Tadoussac at the mouth of the Saguenay fjord.

Site 2. The Channel Head site (latitude: 48°67.83′, longitude: 69°33.38′) is located approximately 8 km east of the Saguenay site on the north side of the St. Lawrence estuary.

Site 3. The Alouette site (latitude: 48°02.56′, longitude: 69°40.71′) is located 8 km west of the Saguenay site on the south side of the St. Lawrence River.

Site 4. Baie St. Marguerite site (latitude: N48° 15.00, longitude: W69° 55.30) is a cul-de-sac located north of Tadoussac and up the Saguenay tributary. Recordings of whales vocalizing in background noise were taken in July and August at each site during the hours of 0700, 1000, and 1400. In addition, some recordings were taken by following the pod from site to site during peak mid-day hours.

Lombard testing in humans is well known. Typically, the test involves having a person repeat a set of sentences or words while the tester varies the noise level presented to the subject. A Lombard effect would be indicated by the subject's voice level rising and/or falling in coincidence with the increase and decrease in the presented noise. In this study noise was generated by the presence or nonpresence of ships. The ''sentences'' were replaced by known beluga vocalizations that were chosen by a hidden Markhov classification system devised by Clemins and Johnson (2003) at Marquette University. Four classified vocalizations that the belugas routinely made at all of the sample sites were chosen. All four vocalizations were whistles.

Recordings were taken on groups of whales at selected sites when no vessels were present, followed by the purposeful presence of a vessel passing through that site and again, afterward (when the vessel was gone). The selected vocalizations were inspected during the before-vessel (no-noise) and during the vessel present (in-noise) situations to determine more specifically whether the noise had a direct effect on the vocalization level of that group of animals by sampling during these specifically created treatments. In all cases the group (subpod) of animals was first identified (group and identification number) and the standoff range from the boat to the pod was at most 400 m but not less than 100 m. The animals in the pod were kept in sight at all times during the test as best as could be done from the observation/recording boat. In two instances digital video recordings of the pod were made during the test. The recording hydrophone was deployed from the recording boat to a depth of 8 m and at no time were whales seen close to or approaching the hydrophone.

B. Data acquisition and analyses

Recordings were made from the R/V BLEUVET of the Center d' Interpretation des Mammifers Marins (CIMM) near the shoreline of the Saguenay River tributary and St. Lawrence Seaway. The BLEUVET is a 26 ft. Cabin cruiser with a Volvo 6 cylinder 3, 21 Turbo diesel 200-hp engine and Volvo Penta Dp stern drive with dual counter-rotating propellers. A total of 230 h of recording was used for this research.

All recordings were made with an International Transducer Corporation model ITC-1042 omnidirectional hydrophone with preamplifier (frequency response flat ±3 dB from 20 Hz to 40 kHz). Recordings were made on a Sony TCD-D8 digital audio (DAT) tape recorder with 48-kHz sampling frequency and 16-bit linear quantization using the LINE input. The TCD-D8 recorder had a flat frequency response from 20 Hz to 20 kHz. Recording instrumentation was calibrated relative to a 1000-Hz calibration tone. Edited portions of the recordings were analyzed with a PC using PRAAT 4.1 speech analysis program (Boersma and Weenink, 2003) for spectrographic wideband analysis with a sampling rate of 20–48 kHz and ATSPEC PRO spectrum analysis soft-

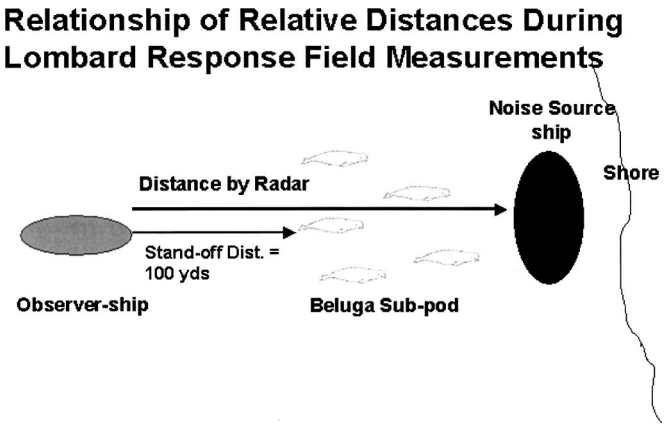


FIG. 1. Showing logistics and the negligibility of changes in vocalization level of any animal based on logistical standoff distance from the recording vessel.

ware (Taquis Corp.) for power spectra (sampling rate of 50–80 kHz was used, vocalization dependent, and an FFT size of 2048 points).

During recordings the recording vessel was shut down. Recordings of merchant- and whale-watching ship noise were made with the omnidirectional recording hydrophone placed at 8-m depth and situated such that the whale subpod was at a standoff distance of 100 m from the recording ship (required by law) and with the whale-watching vessel (noise source) on the far side of the pod. The position of the recording vessel and recording hydrophone in relation to the pod was with the pod directly forward of the recording hydrophone. A census of whales was taken at the beginning of each recording and again upon completion. In most cases the whales remained at or near the surface and the pods generally remained together. That is, the pod remained grouped within a roughly 50-m circle. The distance from the recording vessel to the whale-watching vessel was taken by radar and never exceeded 500 m. The logistics are shown in Fig. 1.

PRAAT 4.1 was used to segment individual whale vocalization ''sound cuts'' from the recordings. Using the software, each vocalization that was visible above the noise was extracted in its entirety (signal and noise). To get specific vocalizations for use in Lombard response analysis presorting of all beluga calls followed by vocal classification methods was used: batch spectral moments and hidden Markhov (HMM) classification. Spectral moments were run on each signal to make an initial similarity grouping. The batch spectral moments program of Milenkovic (1999) provided an assessment of four moments of each vocalization signal in 20-ms windows with 10-ms overlap to facilitate looking for changes in spectral content within a single call (similar to changes in sound within a single syllable). The moments are

(1) Moment no. 1: amplitude-weighted average frequency of the spectrum;

(2) Moment no. 2: standard deviation of the frequency spectrum;

(3) Moment no. 3: skewness of the frequency spectrum; and

(4) Moment no. 4: kurtosis of the frequency spectrum.

The spectral moments were then averaged across all windows to define a frame of each whole vocal production. A MATLAB 6.1 artificial neural net toolbox software (Mathworks, 2003) script was used to group all simple tonal calls together based on the weighting of the moments. The script used unsupervised network architecture. Average frequency and skewness were weighted preferentially above the standard deviation and kurtosis moments for specific tonal vocalization grouping as well as to differentiate the vocalizations from water sounds. Average frequency and skewness showed greater variability between vocalization types, giving further reason for their use in weighting by the network. A matrix of spectral moments versus frames of data comprised the input for each vocalization, and training was accomplished through a set of iterations. This constituted the initial sorting of the whistles.

Given the great variety of vocalizations made by belugas, it was important to compare like vocalizations in both the noise and no-noise conditions. Therefore, a hidden Markhov classification system (Clemins and Johnson, 2003) was used to further classify and to find specific groups of vocalizations that the St. Lawrence belugas routinely made at all sites and that could be reliably identified into four specific calls. Four vocalizations were chosen to be representative of ''typical'' acoustic communication by these animals. These four vocalizations served as the sample word list for the Lombard assessment. The HMM network used for the beluga vocalization classification was a set of *n* nodes. The ''*n*'' was either 5 or 10, which corresponded to the number of natural clusters desired. The nodes represented a type of vocalization, and each node was initialized to represent the ''average'' vocalization plus a small perturbation so that each node was slightly different. Each node was a 5-state HMM. In the end, each node represented the middle vocalization in each cluster of a ''natural'' clustering of the vocalizations. The HMM model did not use phoneme information, although each model had state transitions. The data used were mel-frequency cepstral coefficients and an energy measure across each 300-ms frame as input into the classifier. Although these were not technically geometric patterns, they represented a heavily smoothed version of the spectrum.

Training was done through a set of iterations. The vocalizations were first converted to framed data. Those data were input into the nodes, each of which used an HMM to evaluate its similarity to the vocalization the node represented. The entire data set was run to determine which cluster each vocalization belonged in (i.e., determine which node each vocalization was ''closest'' to). Then, using those labels, each node was adjusted based on the vocalizations that were assigned to it. A reestimation was then performed on each HMM (node) using the vocalizations that were assigned to it. The process was repeated until the HMMs were stable (meaning when the HMMs did not significantly vary after each iteration/reestimation). The HMM itself was not self-organizing in a technical sense. It did do automatic alignment, which, in a sense, is unsupervised. The self-organizing part was the set of HMMs that was used to construct a competitive network.

The vocalizations that were selected for analysis were those whose spectral characteristics were such that they oc-

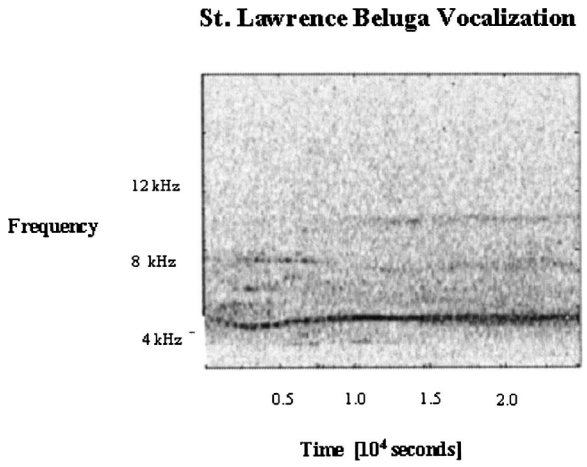


FIG. 2. Shown is a sample sonogram of one of the selected beluga vocalizations used in the Lombard vocal response testing in the St. Lawrence River Estuary.

curred in a frequency band that was above the noise band. The specific vocalizations chosen were all in the 5- to 10-kHz band. This yielded 978 individual vocalization cuts, each containing at least one of the four classified vocalizations and the noise that was occurring with that vocalization at the instant that it was made. These were the samples used in the statistical vocalization versus noise analysis. Most of the vocalizations in these recordings were of sufficient intensity to be heard.

The digital recordings of the classified beluga vocalization signals and accompanying noise from each site were fed into automated code routines that were programed using MATLAB 6.1 software (Mathworks, 2003). Low-pass and bandpass digital filtering were used to separate the vocalization signal (VL) from the noise (NL). This method assumed that the noise was in a different frequency band than the signal, so the filter cutoff frequency of 5 kHz was chosen by observing the noise frequency spectrum at each site when no vessels or animals were present and during the presence of ship noise. The frequency at which the noise level dropped by 20 dB re: 1 μPa was chosen as the cutoff frequency to use in filtering with a steep roll-off. This occurred at 5 kHz, and all signals chosen for analysis and use for detecting the phenomenon were between 5 and 10 kHz. Frequencies in the band of 5–10 kHz were considered the vocalization signal (VL) based upon the classified selected vocalizations. The frequency range of the signals was specifically chosen by the HMM classification system grouping to assure that the vocalization frequencies did not include ship noise. The particular filter method chosen in MATLAB was an elliptical forward–reverse process. The sound cut was filtered in the forward direction; the filtered sequence was then reversed and run back through the filter with the output of the second filtering operation time reversed. This ensured that the result had zero phase distortion and a magnitude modified by the square of the filter's magnitude response. The low/bandpass elliptical digital filters were designed with a roll-off of 80 dB per octave. The rms intensity averages were computed for the VL and NL and archived. A sample vocalization sonogram is shown in Fig. 2. Analyses for the groups of in the no-vessel and vessel present tests were conducted using the same techniques as described above.

C. Statistical methods

1. Signal versus noise observational analysis

The rms values of the VL and NL were statistically compared in correlation and regression analyses using SAS/STAT software Proc Mixed Model. A regression analysis using SAS Proc Mixed Model was performed on the paired noise level (NL)–vocalization level (VL) data using the values obtained by the filtering process previously described ($N=978$). Regression and correlation analyses were also performed to see whether vocalization levels changed as a function of noise and whether or not they differed from those that occurred at ambient levels at each site. The linear regression is shown in Fig. 3.

2. Noise versus no-noise treatment analysis

Confidence intervals were calculated due to the small amount of data ($N=43$) that was collected using two treatments: ship present and no-ship present. These tests were also chosen because the experimental unit could not be adequately defined as corresponding to individual subjects. These subpods were chosen at random across time of day and site. Thus, the pre- and postvessel (vessel off site) recordings were designated as the ''no-noise'' treatment, and the recordings taken while the vessel was on site were designated as the ''with-noise'' treatment. The statistical analysis consisted of calculating confidence intervals for the noise and no-noise vocalizations as further indication of the ''vocalization as a function of noise level'' phenomenon. In addition, regression analyses of vocalization level and noise level were completed for the vessel and no-vessel conditions separately.

III. RESULTS

Results were obtained on the following data:

(1) Beluga vocalization signal versus noise level observation relationship.
(2) Beluga vocalization signal versus noise level during vessel present–no-vessel present treatments.

A. Signal versus noise observational results

Results of the vocalization versus noise analysis indicated that a direct correlation exists. The coefficient of correlation had a value of 0.795. The coefficient of determination was calculated as $r^2=0.6301$. This indicates that 63% of the variability in the beluga vocalization intensity is accounted for by the background noise. These results suggest that beluga vocalization levels vary as a function of noise in the environment.

A regression analysis was conducted to further clarify the nature of the relationship between VL and the NL. The equation of the regression line was $y=0.88x+9.57$. The linear regression is shown in Fig. 3. The dashed lines indicate

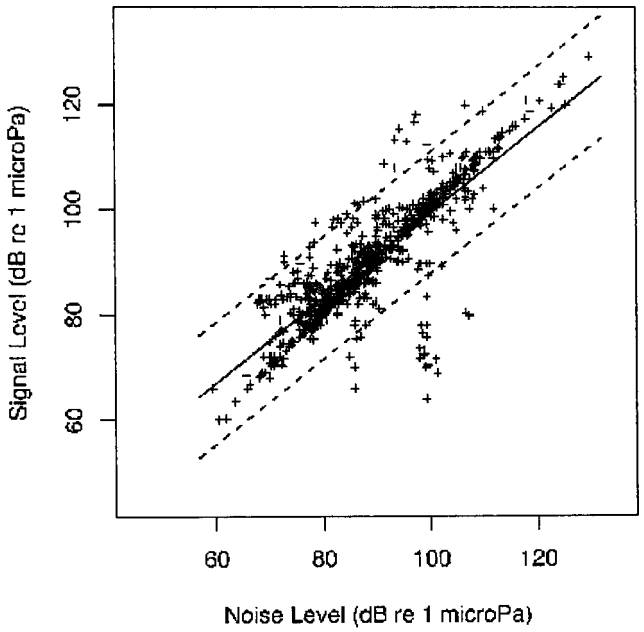


FIG. 3. Regression of beluga vocalization level (VL) versus changing noise levels from extracted beluga vocalizations at all sites in the presence of noise ($N=978$).

the 95% confidence interval for the estimation of the linear regression line. The regression shows outlying data points that are indicative of random effects.

B. Noise versus no-noise treatment results

Two tests were run on the ($N=43$) noise–no-noise treatment data: a 95% confidence interval was computed and a regression run. The confidence limits for vocalizations in the no-noise condition were 86.76 to 80.46 dB, while those for vocalizations made in noise were 99.10 to 91.74 dB. The lack of overlap in these confidence intervals confirms vocal intensities differed between the noise and no-noise conditions. The mean noise level before and during the presence of the vessel were clearly above the hearing threshold of the beluga (as reported in Au, 1993) at all sampled frequencies, and the vocalization levels of the whales in both cases ranged above the noise during all treatments as shown in Fig. 4. In these treatments the whales were recorded before, during, and after the presence of a vessel of opportunity to artificially cause the vocalization-as-a-function-of-noise phenomenon to occur.

IV. DISCUSSION

Based on the variability of ambient noise levels previously sampled at each site over 6 sample years (1996, 1998, 1999, 2001, 2002, and 2003), it is clear that the beluga whales of the St. Lawrence River Estuary are subject to relatively high noise levels, the sources of which were largely anthropogenic in these samples and a highly variable ambient acoustic environment. Results of the analysis of ambient noise at each site that the belugas visit during the summer months indicate that the noise levels vary from site to site as well as within each site. These variations are based on conditions of weather, bathymetry, tides, current regimes, and topography. The addition of anthropogenic noise exacerbates these noise fluctuations. In large baleen whales, body size allows the animals to produce low-frequency signals at high intensities that serve to increase the range over which they can be heard by conspecifics. By combining these acoustical characteristics with the selection of suitable depths and bottom types, they can match signal form with ambient medium characteristics for long-distance communication. That is not

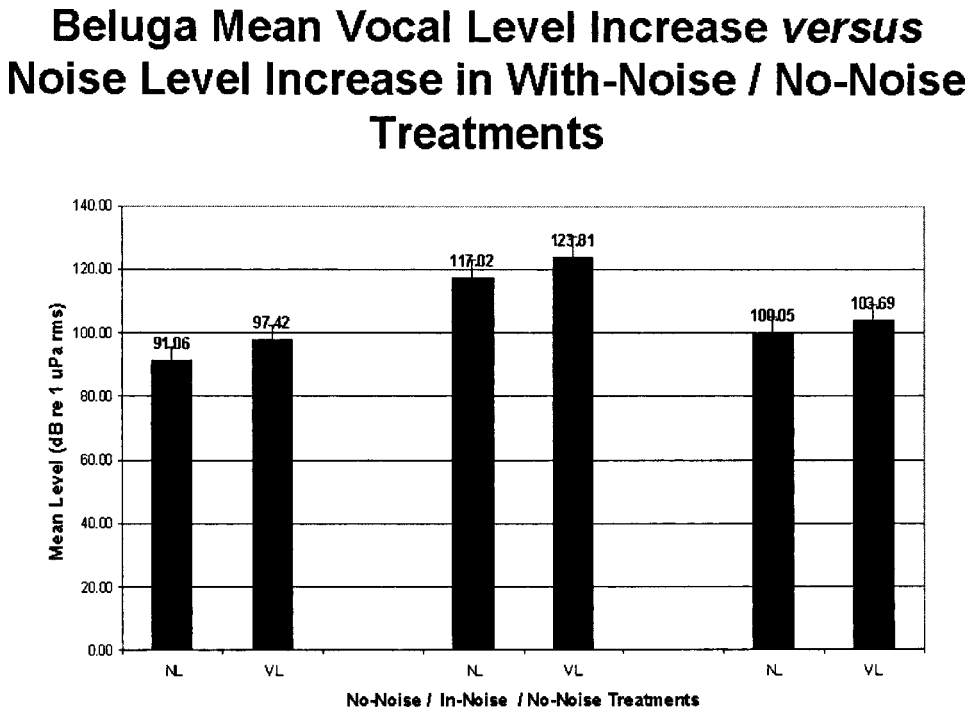


FIG. 4. Shows the response of the beluga vocalizations (VL) in no-noise (ship not present), noise (ship present), and no-noise (after the ship had left the site) conditions.

the case with the belugas of the St. Lawrence River Estuary, who constrain themselves to a considerably smaller body of water where currents and tides have a greater effect on signaling. Thus, physical environment and animal physiology (modality) both impose constraints on signal design and optimization for the beluga.

Belugas may employ a number of strategies for vocalizing that would allow them to optimize their acoustic signals to communicate with their conspecifics. These are limited to changing the frequency of their vocalizations, changing the type of call emitted, such as switching to pulsed calls instead of tonal calls, leaving the site for quieter waters, or changing vocalization intensity. Each of the former approaches has been observed in the past; however, their use of the latter tactic has not been well documented.

The belugas must select design features for their communication vocalizations that will optimize their vocalizations in noise. A design feature is a signal characteristic that is determined by environmental or other selective forces and affects the optimality of the signal (Bradbury and Vehrencamp, 1998). The optimum design scheme would use the smallest number of signal features common to most modalities that reflect both information content and transmission properties common to most signals employed by the animal. For auditory signals the maximum range, ability of the signal to be localized by sender and receiver (directionality), duty cycle, and modulation strength are all factors that have to be considered with regard to altering signal intensity. These factors are all accounted for by the Lombard response.

To increase confidence that the Lombard response exists, controlled testing on captive belugas should be evaluated. It should be noted, however, that although results from such a test may help to confirm the Lombard response, it might not accurately serve to quantify it with respect to Lombard thresholds due to the nature of the captive (pool) environment, which is reverberation limited. That is, the thresholds at which the response begins to be exhibited and at which complete masking occurs are not known. In a reverberation-limited versus noise-limited environment, these thresholds may not be able to be determined accurately.

Although this response cannot be attributed to the wild beluga with certainty, the phenomenon of ''beluga vocalization-as-a-function-of-noise'' unquestionably appears to exist based on the consistency of the phenomena throughout the recordings. In most cases the vocalization level either rose or fell in coincidence with the noise level at the moment that vocalization was made. This is highly indicative of a Lombard vocal response, as seen in humans and other animals that have been tested for the Lombard response. Less than 4% of selected vocalizations were not above the noise level. The data taken throughout the recordings support the hypothesis that the St. Lawrence belugas exhibit a vocalization-as-a-function-of-noise phenomenon based on the strong positive correlation between elevated noise and subsequent elevation of the belugas' vocalizations. The robust sample size seems to indicate that the phenomenon is not coincidental nor is it likely to be caused by physical factors. The regression data indicate a linear relationship between the beluga vocalization signal and noise at the time the signal was made, which is typical of the Lombard response.

Evidence that this phenomenon exists is further supported by the noise versus no-noise treatments, although the sample size was small. Once again, the regression relationship was linear. The data regarding the acoustical response agree well with data gathered on other animals with respect to the Lombard vocal response. That is, the vocalization level per decibel increase in noise lies between 1.0 dB VL/1 dB NL increases for humans and 3.3 dB VL/1 dB NL increase for finches.

In comparing the beluga data of vocal increase as a function of noise increase and vocal increase as a function of noise decrease with that of human subjects tested by Scheifele (2003), the values of the rate of increase and decrease per decibel compared favorably. The data from each of the tests conducted during this study strongly imply a Lombard vocal response. In addition, tests with other animals that have been shown to exhibit the response yield similar response results in their data. Overall, our findings indicate that a Lombard vocal response does exist in belugas. Thus, the data presented here suggest that environmental noise has an observable effect on the communication process of these animals. Given that elevated noise levels occur so routinely at all sites in the St. Lawrence River Estuary during summer, it is likely that observing such a response taking place so often represents a significant impact on the ability of these animals to communicate effectively with potential impact on their energetics. Since the state and stability of this threatened population of belugas is so tenuous, and since these sites represent such popular sites for the ecotourism industry, routine demonstration of a Lombard response should be viewed as a warning of potential adverse impacts of noise on these animals given that the Lombard vocal response is a first-order reaction to noise. Once the Lombard ceiling for a species or individual has been reached, the next level of noise would be masking. As such, a monitoring program should be initiated using the Lombard response as an indicator and measure for the low-level effects of noise on the St. Lawrence beluga. Further studies of the Lombard response in the beluga whale should focus on determining the floor (level at which the Lombard response begins to be exhibited) and the ceiling (level at which the Lombard response reaches its peak and where the animal's communication system cannot accommodate the noise further). This has not been determined for any species to date. Knowing these limits would provide a metric for gauging the effects of noise on populations of wild animals such as the St. Lawrence beluga whale in the future.

Au, W. W. L. (**1993**). *The Sonar of Dolphins* (Springer, New York), p. 34.

Boersma, P., and Weenink, D. (**2003**). ''PRAAT: Doing phonetics by computer,'' University of Amsterdam, Netherlands. http://www.fon.hum.uva.nl/praat/. Viewed 12 November 2003.

Bradbury, J. W., and Vehrencamp, S. L. (**1998**). *Principles of Animal Communication* (Sinauer Associates, Inc., Sunderland, Massachusetts), pp. 1–16.

Clemins, P. J., and Johnsosn, M. T. (**2003**). ''Application of speech recognition to African elephant (*Loxodonta Africana*) vocalizations,'' Proceedings ICASSP-03.

Cynx, J., Lewis, R., Tavel, B., and Tse, H. (**1998**). ''Amplitude regulation of vocalizations in noise by 2 songbird (Taeniopygia guttata),'' Animal Behavior **56**, 107–113.

Egan, J. J. (**1966**). ''Psychoacoustics of the Lombard Voice reflex,'' Ph.D. dissertation, Case Western Reserve University.
Egnor, S. E., Dingle, K., Guvench, A., Hicks, C., and Hauser, M. D. (**2003**). The Lombard Effect on the Cotton-top Tamarin (Saguinus oedipus). Presentation given at the First International Conference on Acoustic Communication by Animals, July 27–30, at the University of Maryland, College Park, Maryland.
Guenther, F. H. (**2001**). ''Neural modeling of speech production,'' Proceedings of the 4th International Nijmegen Speech Motor Conference, Nijmegen, The Netherlands, 13–16 June, 2001.
Lane, H., and Tranel, B. (**1971**). ''The Lombard Sign and the role of hearing in speech,'' J. Speech Hear. Res. **14**, 677–709.
Levelt, W. J. M. (**1989**). *Speaking: From Intention to Articulation* (MIT Press, Cambridge, MA).
Lombard, E. (**1911**). ''L'indication de l'elevation de la voix,'' Annales Des Malades de l''creille, Vol. 37, No. 2.
Manabe, K., Sadr, E. I., and Dooling, R. J. (**1998**). ''Control of vocal intensity in budgerigars (*Melopsittacus undulatus*): Differential reinforcement of vocal intensity and the Lombard Effect,'' J. Acoust. Soc. Am. **103**, 1190–1198.
MathWorks. (**2003**). ''MATLAB: The language of technical computing,'' Natick, MA. http://www.mathworks.com/
Milenkovic, P. H. (**1999**). ''Moments: Batch Spectral Moments Analysis,'' University of Wisconsin-Madison. (http://www.wisc.edu/cgi-n/wisc.ph/ph.cgi?source = brief.xml&name = Paul + H+Milenkovic&email =&division=&phone=). Created 1999 and accessed 2003.
Potash, L. M. (**1972**). ''A signal detection problem and a possible solution in Japanese quail,'' Anim. Behav. **20**, 192–195.
Scheifele, P. M. (**2003**). ''Investigation of the response of the auditory and acoustic communications systems in the beluga whale of the St. Lawrence River Estuary to noise using vocal classification,'' Ph.D. dissertation, University of Connecticut.
Sinott, J. M., Stebbins, W. C., and Moody, D. B. (**1975**). ''Regulation of voice amplitude by the monkey,'' J. Acoust. Soc. Am. **58**, 412–414.

North Pacific Acoustic Laboratory

Peter F. Worcester[a)]
Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093-0225

Robert C. Spindel
Applied Physics Laboratory, University of Washington, Seattle, Washington 98105

(Received 13 August 2004; revised 13 September 2004; accepted 13 September 2004)

A series of long-range acoustic propagation experiments have been conducted in the North Pacific Ocean during the last 15 years using various combinations of low-frequency, wide-bandwidth transmitters and horizontal and vertical line array receivers, including a 2-dimensional array with a maximum vertical aperture of 1400 m and a horizontal aperture of 3600 m. These measurements were undertaken to further our understanding of the physics of low-frequency, broadband propagation and the effects of environmental variability on signal stability and coherence. In this volume some of the results are presented. In the present paper the central issues these experiments have addressed are briefly summarized.
[DOI: 10.1121/1.1854780]

PACS numbers: 43.30.Qd, 43.30.Re, 43.30.Bp, 43.30.Cq [AIT] Pages: 1499–1510

I. INTRODUCTION

There is a rich history of low-frequency, long-range propagation experimentation over the last sixty years, starting with the discovery of the deep sound channel in 1944, and continuing today. Experiments up to the mid-1970's covered a wide range of investigation, but principally consisted of measurements of parameters in the sonar equation: transmission loss, reverberation, and ambient noise. The goals were twofold: to improve the prediction of sonar system performance and to optimize system performance under various operating conditions. These studies used wideband explosive sources and narrowband transducers. They form the basis for our understanding of propagation through a deterministic ocean and have shed some light on the statistics of its variability. But it is the development over the last two decades of low-frequency, wideband transducers driven by controlled waveforms and vertical receiving arrays that can store a year or more of data, that finally provide the means to measure the spatial and temporal statistics of fluctuations needed to advance the development of propagation theory. This technology allows individual multipaths and modes to be resolved and long time series to be collected.

Beginning in 1989 a series of experiments, mostly in the North Pacific, have been undertaken to study low-frequency acoustic fluctuations (Fig. 1 and Table I). The 1989 experiment was the first to combine a broadband transmitter designed to measure travel times with 1 ms precision with a 3 km long vertical receiving array. The 250 Hz, 1001 km transmissions revealed a traditional SOFAR arrival pattern of steep, ray-like time fronts in the early portion building to a crescendo of near-axial arrivals that cut off abruptly. The near-axial finale was found to be highly scattered, however, and included significant energy in the geometric shadow zone. The prevailing theory of scattering from internal-wave-induced sound-speed fluctuations predicted far more resolvable arrivals than were observed.[1] This inability to predict one of the most fundamental aspects of ocean propagation stimulated an intense research effort to reconcile theory with data. The next experimental opportunity occurred in 1994 in a test of long-range propagation (3252 km) in support of the Acoustic Thermometry of Ocean Climate (ATOC) program. It consisted of a 75 Hz broadband transmitter suspended from *R/P FLIP* near California and a 20-hydrophone vertical array near Hawaii.[2] (The array was supposed to be twice as long with 40 elements, but unfortunately one of the two 20-element subarrays failed.) The experiment was designed to test the feasibility of using long, ocean-basin-scale paths for acoustic thermometry. Results showed early, ray-like time fronts that were resolvable, identifiable, and stable. These arrivals were actually more stable than predicted, with relatively small values of travel-time wander and pulse spread. As in 1989, however, the near-axial finale of the arrival pattern was highly scattered. Following this successful test—early arrivals were clearly shown to be useful for thermometry—two 75 Hz ATOC transmitters were installed. One, off California, began operating in December, 1995; the other, off Kauai, began transmitting in October, 1997. Also in conjunction with ATOC, in 1995 two vertical array receivers were installed near Hawaii and Kiritimati Islands and each operated for about a year. (In 1998 another vertical array was installed at Ocean Weather Station Papa in the Gulf of Alaska, which operated for nine months.) The 75 Hz transmissions to these arrays and to bottom-mounted horizontal arrays provided much of the data discussed in the papers in this volume.

There were two other experimental opportunities. In 1996, for a short 9-day period, a transmitter suspended from a ship near the California coast transmitted simultaneous 28 Hz and 84 Hz signals to explore the frequency dependence of scattering.[3] And then in 1998 a 2-dimensional receiving array (3600 m horizontal aperture and 1400 m maximum vertical

a)Corresponding author: Electronic mail: pworcester@ucsd.edu

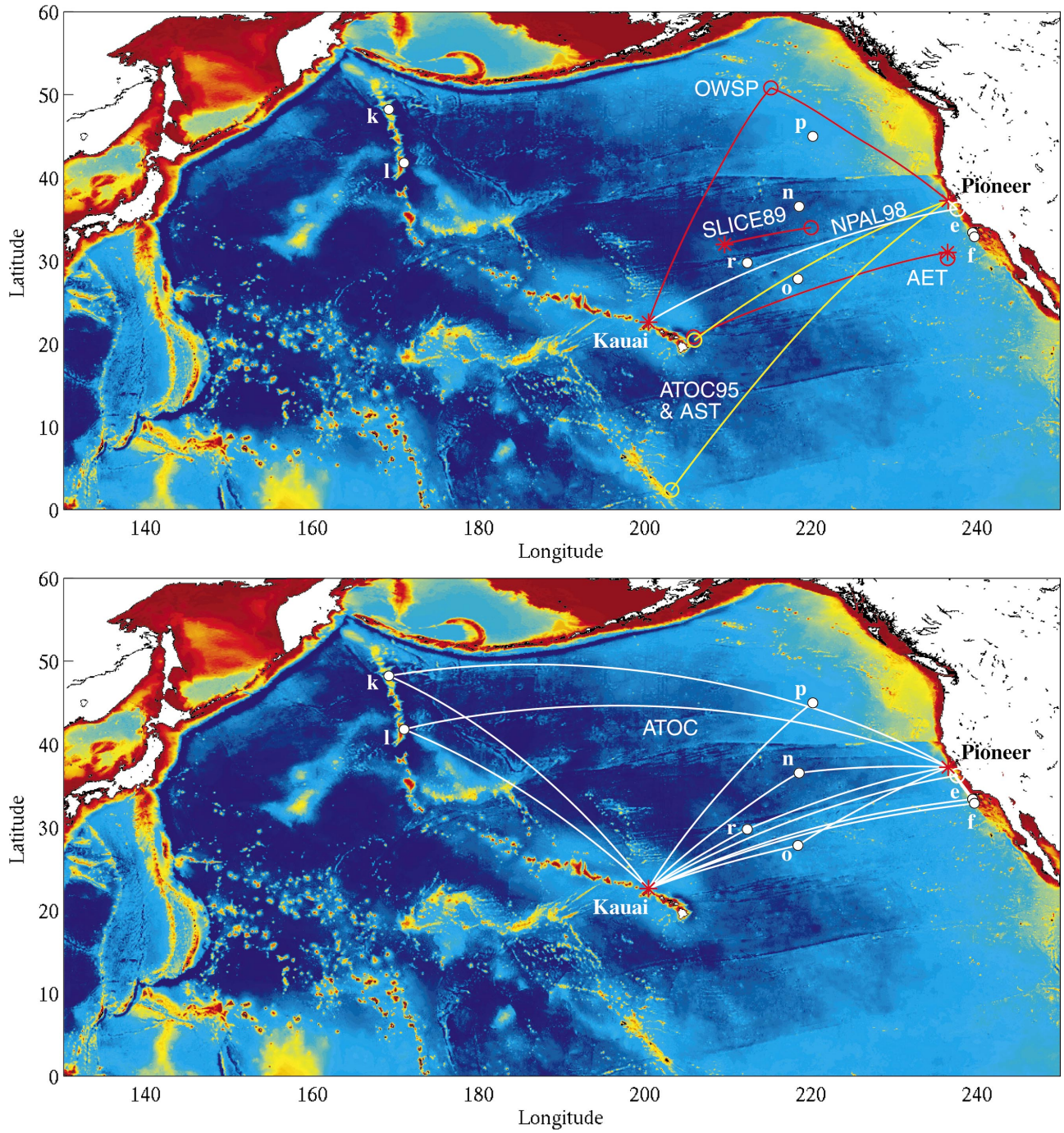


FIG. 1. Locations of various NPAL components. Acoustic paths to the vertical array receivers are shown in the upper panel, and paths to other receivers labeled with lower case letters are shown in the lower panel.

aperture) was installed at Sur Ridge off Point Sur, California, for a year (Fig. 2).

Almost all of these experiments included environmental measurements along the various transmission paths, including high-resolution conductivity, temperature and depth transects and moored time series. They were also accompanied by independent (from the acoustic scientists) marine mammal monitoring and research programs.

Although each experiment has its own individual name (SLICE89, AET, AST, ATOC), they are loosely grouped into what we call the North Pacific Acoustic Laboratory (NPAL). The variety of sources and receivers deployed during the various phases of NPAL offer a wide range of data, over propagation paths up to 5000 km long, spanning seasonal and annual cycles, and encompassing spatial receiving apertures up to 3000 m in the vertical and 3600 m in the horizontal.

These data have served to test ever more sophisticated theories of wave propagation in the ocean medium. Still, many unanswered questions remain. For example, although it is generally accepted that internal waves are the primary source of scattering resulting in random amplitude and phase variations, none of the various theories developed over the last two decades successfully predicts received signal char-

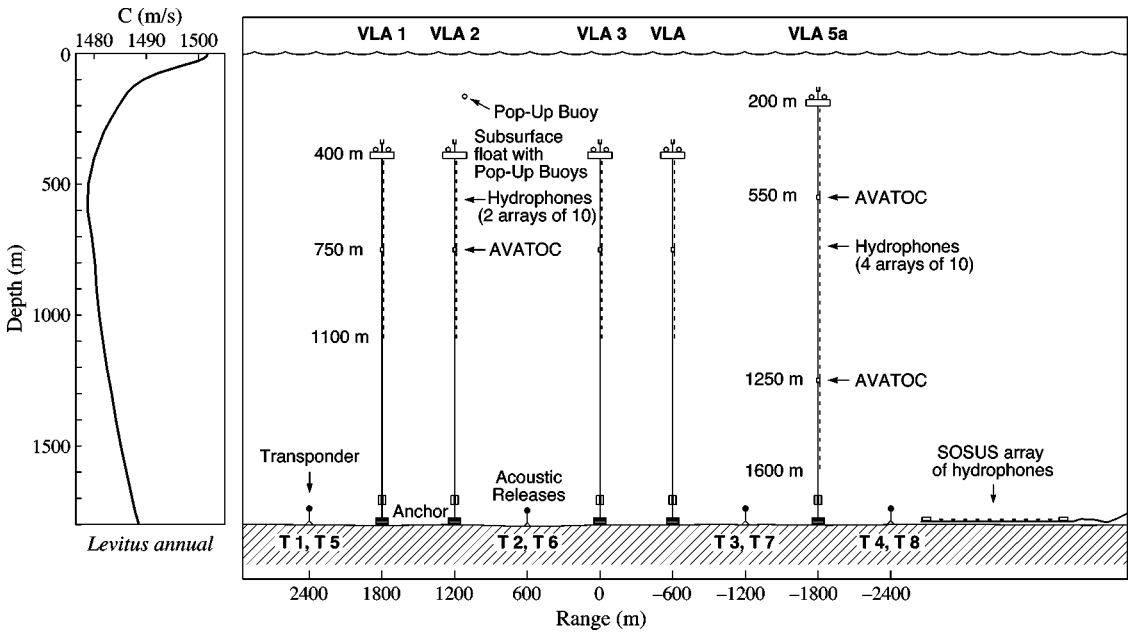


FIG. 2. Two-dimensional receiving array at Pt. Sur on the right and sound-speed profile on the left. A transponder network (T1, T5, T2 and so on) was installed for array element localization. The 2-D array was located near an existing Navy SOSUS array shown at the lower right.

acteristics as a function of range, frequency, depth or internal wave intensity. Other scattering mechanisms such as ocean ''spice'' have been proposed recently. Here the temperature and salinity of patches of hot salty water and cold fresher water compensate to yield equal density, but not equal sound speed. The spatial and temporal coherence properties of the acoustic field are still not completely predictable. Indeed, new theories have introduced the notion of ray chaos, which suggests an exponential proliferation of rays with increasing range and calls into question the entire premise of ray stability and coherence. The vertical extent and intensity of acoustic energy received on hydrophones in deep shadow zones cannot be explained by the presumed origin of the phenomenon, scattering from internal waves. Many of the properties of the ambient noise field are also poorly understood, particularly in the vertical, simply because horizontal apertures have until now been the primary technology employed in low-frequency applications.

In this paper we provide a brief summary of current NPAL results.[4] Details are found in the various reports in this volume. Although many of the experimental configuration and signal processing particulars are provided in these papers, Appendix A assembles them here for the reader's convenience. Remarks about signal processing are in Appendix B.

II. PRINCIPAL RESULTS

The papers in this volume are concerned with statistical measures of amplitude and phase fluctuations at long ranges, the spatial and temporal coherence of the received signal, signal energy redistribution through mode scattering, horizontal refraction, the implications of the theory of ray chaos, the effects of bottom interactions, the characteristics of ambient noise on ocean-basin scales and the potential impact of the low-frequency transmissions on marine mammals. The single vertical arrays allow resolution of low order modes (up to about ten at 75 Hz). The 2-D array allows the measurement of the signal wave front as a function of depth, transverse distance and time. It enables the study of the temporal, vertical, and horizontal coherence of resolved rays and modes (and ambient noise).

A. Fluctuations

Perhaps the single most engaging issue in low-frequency ocean acoustics in the last three decades has been the quest to develop a theoretical basis to explain observed amplitude and phase fluctuations. Early efforts applied various techniques borrowed from electromagnetic scattering theory, such as the method of small perturbations and Rytov's method, to calculate acoustic scattering from sound-speed fluctuations associated with homogeneous, isotropic turbulence. These theories of wave propagation in random media have all been found wanting when applied to propagation in the ocean. The difficulty is that much of the small-scale variability in the deep ocean is neither homogeneous nor isotropic, but rather is due to oceanic internal waves. Following the development of a model of oceanic internal waves by Garrett and Munk in the early 1970s, Flatté *et al.*[5] combined the model with path-integral methods to predict the fluctuations of resolved geometric ray arrivals. In simple terms internal wave scattering causes single acoustic rays to split into many micro-multipaths that more or less follow the same mean path as the unperturbed ray, and arrive at more or less the same time. This construct offers a pleasing physical explanation for observed time spread (pulse broadening—micromultipaths that cannot be resolved in time and thus look like a single broader pulse) and time wander (arrival time fluctuations). There have been numerous refinements to this theory since its development in the mid-1970's, but the essential physics has remained the same.

Virtually every observation since then has appealed to

this theory for an explanation with varying degrees of success. The NPAL series of observations with its broadband transmissions at multiple frequencies (28, 75, 84, and 250 Hz), receptions on vertical arrays, a variety of ranges and range-dependent environmental (sound speed) measurements along many of the paths, provided a unique data set to further test this theory's predictions.

The outcome is at best problematic. Some of the data provide a good match to theory; some do not. Direct comparisons with theoretical predictions can be made in the portion of the SOFAR receptions preceding the finale where multipaths can be resolved. For the two paths examined in detail in this volume[6] signal fluctuations are near the 5.6 dB value indicative of full saturation, but the scintillation index, SI, is significantly greater than one, indicating that full saturation had not yet been reached. Recent comparisons between theoretical predictions and simulations that use the parabolic equation show that the integral-approximation estimates of the fluctuations can be seriously in error at ranges in excess of a few hundred kilometers,[7] so the lack of agreement with the theory is perhaps not too surprising.

In the near-axial finale region of the SOFAR receptions, where multipaths cannot be resolved and are uncorrelated, the integral-approximation predictions are not applicable. The statistics are nonetheless of interest. Results in this volume[8] show that in this case the interference pattern is representative of a fully saturated process. Here, according to theory, the phase is uniform in $[0, 2\pi]$, the scintillation index, SI, is 1 and the variance of log intensity, $\sigma_{\ln I}$, is 5.6 dB. Over multi-day time scales the data support this. On the other hand, shorter time scales, from minutes to 40 minutes, show considerably less variability with SI$\sim$0.3–0.7, and $\sigma_{\ln I}\sim$3–4.7 dB. (The lowest values come from the 28 Hz data.) Similarly, although phase fluctuations at 28 Hz are consistently smaller than those at 84 Hz as predicted, the observed scaling is at odds with the predicted scaling. Single-path geometrical theory predicts a frequency-squared dependence, or a factor of 9, but observations show factors of only 1.5–3. There are other inconsistencies. For example, shorter paths often exhibited larger fluctuations than longer ones at the same frequency (e.g., 3252 km vs 5100 km). (The interpretation is somewhat uncertain, however, because the paths were geographically separated.) Vertical wave number spectra of phase at 75 Hz and 84 Hz are barely consistent with theoretical models of random multipaths, and frequency spectra of phase do not agree with the models at all.

The dependence of acoustic variability on bandwidth is another issue, which is especially important because it has significant consequences for signal processing. NPAL data from the finale region show no significant difference in broadband ($\sim$37 Hz at 75 Hz center frequency) and narrowband (1 Hz) fluctuation statistics, which is unexpected.[8]

These and other inconsistencies between theory and experiment have accentuated the limitations of current theory, and have reinforced the need for further theoretical development.

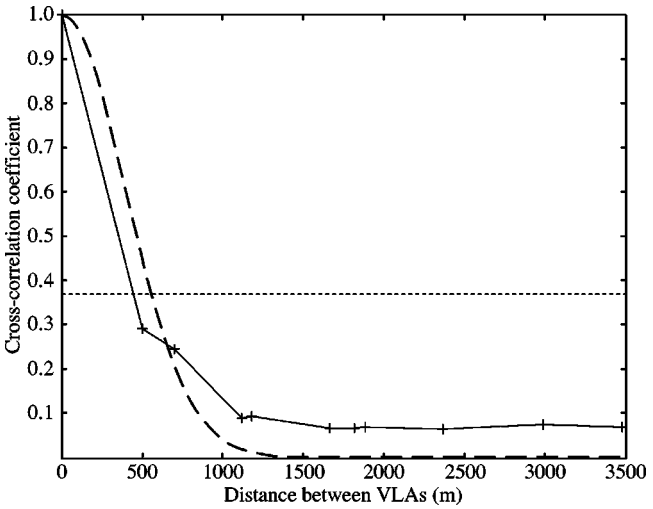


FIG. 3. Horizontal correlation at 75 Hz, 3900-km range, averaged over depth. The solid line is data; the dashed line is a model prediction (from Ref. 9).

B. Signal coherence

The spatial and temporal coherence of the acoustic signal is fundamental to achievable signal processing gains. It sets limits to gains realized by averaging in space and time, and to angular resolution. It governs the efficiency of adaptive array and matched field processors. Yet the goal of developing an accurate predictive model, or even developing dependable empirical rules of thumb, with a few exceptions, remains unachieved.

1. Resolved arrivals

NPAL data from horizontal arrays suggest horizontal coherence lengths of resolved wave fronts of 400–500 m at 75 Hz and ranges of 2000–3000 km, in accord with theoretical predictions.[6] The data from the 2-D array exhibit similar horizontal coherence lengths, even though the measurement is coarser (Fig. 3).[9] (The first lag distance is about 500 m, the distance between the two closest vertical arrays in the 2-D array.) Vertical coherence lengths in the region of resolved arrivals at all frequencies are in excess of 500 m, allowing modal filtering and ray beamforming. Temporal coherence is in excess of 10 minutes at all frequencies.

2. Near-axial finale

The coherence of the signal in the unresolved SOFAR finale has been examined in two ways. One approach uses the total acoustic field,[8] and the other decomposes the field into modes.[10] In the former case, vertical coherence lengths at 3000–5000 km ranges are about 100 m at 75 Hz and 84 Hz, and 150 m at 28 Hz, which is surprising considering the complexity of the vertical interference pattern in this region of signal reception. Coherence times are roughly 10 minutes at 75 Hz and 84 Hz, and greater than 40 minutes at 28 Hz. (The longest signal transmission was 40 minutes.)

The NPAL vertical arrays permitted resolution of the lowest ten modes, those associated with the late finale arrivals, the most energetic part of the signal. The mode signals can be used in tomography, matched field processing and other applications in much the same way that resolved mul-

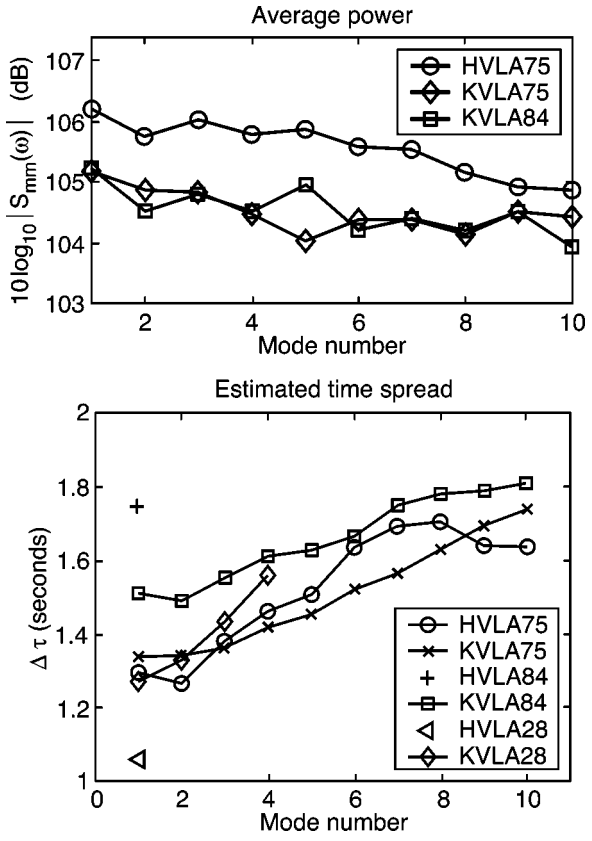


FIG. 4. Upper panel: Average power in center frequency bin (75 Hz and 84 Hz) for modes 1–10 for California to Hawaii (HVLA) and California to Kiritimati (KVLA) transmissions. Lower panel: Mode time spread for 28, 75 and 84 Hz transmissions (from Ref. 10).

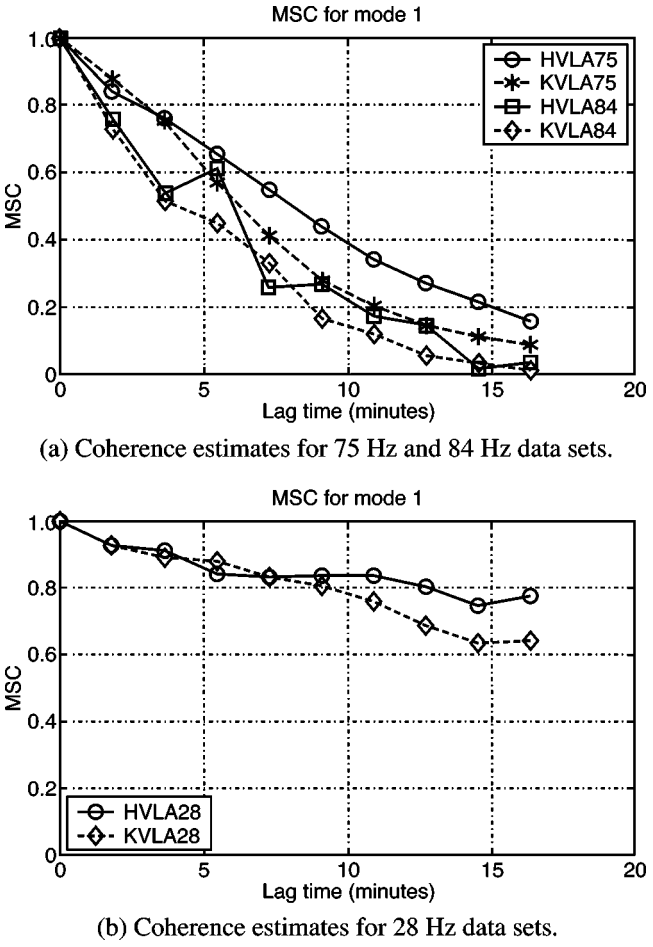


FIG. 5. Temporal coherence for mode 1 at 28 Hz (lower panel), and 75 Hz and 84 Hz (upper panel), for the same paths as Fig. 4 (from Ref. 10).

tipaths have been used. Indeed, mode travel times have been used to deduce warming in the Arctic Ocean.[11] Here the internal wave field is weak, and there is little scattering of energy among modes. Thus, each mode arrival has essentially traveled as the same mode, and it is easy to identify the path that the received energy traveled. In the deep ocean, however, the internal wave field is much stronger and causes significant scattering of energy between modes. This results in complicated arrival patterns that are difficult to interpret. Further, there is no theory comparable to the case of internal wave scattering of acoustic rays that explains how modes propagate in the presence of internal waves.

NPAL data provide an empirical picture of the mode statistics at low frequencies and long ranges. In general, the lowest ten modes have roughly equal power as seen in the upper panel of Fig. 4. They have complicated arrival patterns with frequency selective fading and high temporal variability. Mode time spread, which is a measure of the coherence bandwidth of the channel, ranges from about 1 s to 1.8 s depending on mode number, range and frequency, with increasing spread roughly proportional to these three parameters (Fig. 4, lower panel). The roughly 1.5 s spread implies a coherence bandwidth of 0.67 Hz, which is in accordance with the observed frequency selective fading. The temporal coherence is 5 to 10 min for mode 1, dropping to zero after about 15 minutes, at 75 Hz and 84 Hz, and is much longer than 15 minutes at 28 Hz (Fig. 5).[10]

C. Horizontal refraction

The magnitude of horizontal refraction induced by ocean inhomogeneities has been of interest since it was first studied by Munk in 1980,[12] and subsequently proposed as an observable for ocean acoustic tomography.[13] Rough estimates were that angular deflections of about 0.5° could be produced by an average mesoscale eddy. We believe that the first experimental study of horizontal refraction is reported here.[9] Data from the transmitter near Kauai received on the 2-D array off the California coast show a deflection of about 0.4° that occurs twice during the year-long time series. The sound is refracted southward, presumably by an average horizontal gradient of about 5.4×10^{-6} s^{-1} over the 3890 km path from the source to the receiving array. (This cannot be verified by sound-speed data since environmental measurements were obtained only along transmission paths, not perpendicular to them.) The standard deviation of the horizontal refraction, a measure of short time period effects such as those due to internal waves, was 0.37°. This implies a ''correlation radius'' of about 700 m, which is in keeping with other (more conventional) NPAL horizontal spatial coherence measurements discussed above.

D. Chaos

In recent years the limits of predictability for deterministic acoustic propagation have been challenged by the con-

cept of ray chaos, which suggests that classic ray theory is no longer useful beyond a certain range sometimes called the ''predictability horizon.'' [14] At issue is the validity of the usual ray approximation, where the linear wave equation is replaced with a series of nonlinear differential equations, in the presence of small-scale oceanographic variability, such as that due to internal waves. Ray calculations indicate that scattering by ocean inhomogeneities along the propagation path results in the production of multipaths that grow exponentially in number as a function of range,[15] thus invalidating the concept of a deterministic ray, and rendering ray methods unusable. A further consequence of chaotic behavior is that propagation is very sensitive to initial conditions. Rays from closely spaced sources, or equivalently ray paths that are initially very close to one another, diverge at an exponential rate. While ray simulations show that the scattering is predominantly along the deterministic wave front, rather than across it, the implication is that even a stable pulse reception cannot be unambiguously associated with a particular ray that has traversed a known path.

The seeming disparity between empirical observations, where stable and predictable pulse travel times are observed at very long ranges (up to 5000 km in the experiments reported here) and the implications of ray chaos, where ''predictability horizons'' as short as several hundred kilometers have been calculated, has been and continues to be the subject of intense research. Two papers in this volume offer explanations for the discrepancy between experimental evidence and modeling. One suggests that chaotic behavior is predicted unrealistically soon because inhomogeneities that do not appreciably alter the propagating wave, i.e., those smaller than the vertical projection of a wavelength, are included in ray models.[16] It is shown that excluding this fine structure by smoothing it, or filtering it out, leads to much longer predictability horizons. Another argues that while chaotic rays are expected to become random, travel times retain some of the same properties as nonchaotic rays.[17] Specifically, chaotic rays tend to arrive in clusters with a local maximum at times close to unperturbed rays, and the arrival times of chaotic rays with a particular identifier tend to be roughly the same as unperturbed rays with the same identifier. Both arguments support the conclusion that ray chaos does not render ray methods useless, and does not invalidate the use of acoustic travel times as an observable in tomography[18] and other schemes.

Ray methods are only an approximate solution to the full wave equation, of course. The extent to which the phenomenon of ray chaos is manifested in finite frequency wave fields is crucial. One paper in this volume examines wave field stability from both ray-based and mode-based points of view.[19] It is shown that the more familiar mode-based ''waveguide invariant,'' $\beta=(\delta\omega/\omega)/(\delta r/r)$, where ω is frequency, and r is the range,[20] is equivalent to a ray-based stability parameter, α, a function of the ray action variable, whose magnitude has been shown to control ray stability.[21] A second paper uses finite-frequency parabolic equation simulations of narrow acoustical beams propagating through internal wave fields to examine the range-evolution of beam properties, such as beam width, scintillation index and entropy, for comparison with ray calculations.[22] The parabolic equation simulations mimic a small angle bundle of rays. It is found that the acoustical beams in an ocean waveguide with realistic sound-speed perturbations do not expand exponentially (or explosively) as would be expected for ray chaos. Nonetheless, the rates of change of entropy as a function of range in the parabolic equation and ray calculations are found to be comparable once statistical saturation is reached, suggesting at least a limited full-wave manifestation of ray chaos.

E. Bottom interaction and travel time offset

Transmissions from the bottom-mounted NPAL Kauai source differ from those from the mid-water sources typically used in tomographic measurements in that in addition to purely refracted energy, bottom-interacting energy also contributes to the far-field arrival structure.[23] Heaney[24] and Vera *et al.*[25] in this volume show that it is necessary to include surface and bottom interaction near the source and receiver in order to correctly identify the arrivals at long ranges. They find that surface and bottom interacting rays come off the slope north of Kauai with levels comparable to the purely refracted rays, but delayed approximately 0.5 s. Rays with high launch angles that reflect off the surface and bottom in front of the source are converted to lower-angle rays due to the bottom slope. These lower angle rays then propagate as purely refracted or refracted-surface reflected rays to the NPAL 2-D array, the vertical array at Ocean Weather Station Papa and other receivers. The resulting arrivals overlap in angle with the purely refracted arrivals, but arrive slightly later in time due to the extra path length near the source. This class of rays also accounts for the earliest arriving resolved arrivals, because it includes rays with shallower turning points than are possible for purely refracted arrivals.

Once the measured arrivals were identified with specific ray paths, the travel times from the 75 Hz Kauai transmitter to the 2-D California array (3890 km) were found to differ from predictions made using extensive environmental data obtained in 1998 and 1999 by about 0.2 s and 0.4 s, respectively. Offsets between measured and predicted travel times have been found in all NPAL experiments for which contemporaneous temperature and salinity (CTD/XBT) measurements allowed accurate predictions. The offset has always had measured travel times greater than predicted travel times. Worcester *et al.*[26] noted a 0.2 s difference at 3250 km range in the AET transmissions from a ship-suspended 75 Hz source to the vertical line array near Hawaii in 1994. The offset was attributed to error in the Del Grosso sound-speed equation at high pressure, in accord with earlier work.[27–29] The 0.2 s offset observed in 1998 is comparable to that found during the AET, but the 0.4 s offset observed in 1999 is significantly larger. The measured travel times were changing rapidly during the time that the 1999 environmental measurements were being made, however, suggesting that the environmental data were not sufficiently synoptic to make accurate travel time predictions.

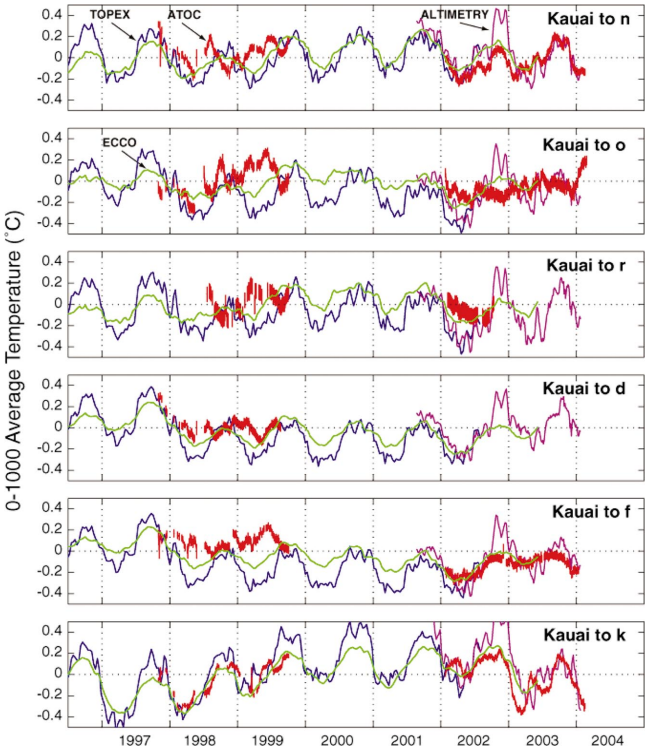


FIG. 6. Thermometry data from the NPAL Kauai transmitter. The acoustic data are red. The first permit for operation ended in 1999, and the transmitter was inoperative for two years from late 1999 to early 2002 pending the granting of a new one. The blue line is obtained from satellite altimetry, and the green line is a model prediction.

F. Thermometry

The continuing measurements in the North Pacific represent the second phase of research on the feasibility and value of large-scale acoustic thermometry [following the Acoustic Thermometry of Ocean Climate (ATOC) experiment]. Long-term trends in large-scale ocean temperature are easily visible in the acoustic time series because acoustic methods give integral measurements of large-scale ocean temperature that provide the spatial low-pass filtering needed to observe small, gyre-scale signals in the presence of much larger, mesoscale noise. Recent data along the paths from Kauai to the California coast (e.g., *f*, Fig. 6) show cooling relative to earlier ATOC data. A path to the northwest, *k*, showed modest warming until early 2003, when a rapid cooling event occurred.[30] This was followed by a much stronger annual cycle.

G. Ambient noise

Ambient noise is typically characterized by simple statistics such as mean and peak levels, and standard deviations. It is usually considered to be white and Gaussian in narrow bands, and amenable to pre-whitening in wider bands. With few exceptions, most ambient sound measurements have been short-term observations.

NPAL provided an opportunity to collect long time series of ambient sound at several fixed locations, both with single hydrophones and vertical arrays, in the deep ocean and in a coastal region. Much of this data was obtained using U.S. Navy SOSUS receivers and has been reported

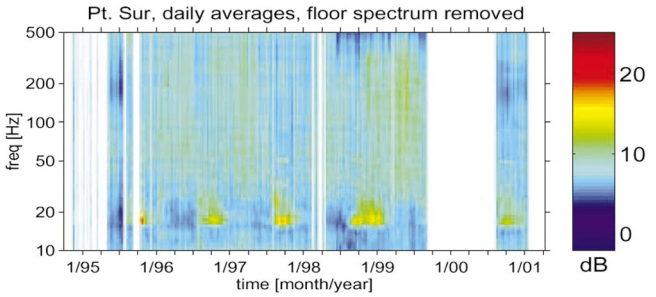


FIG. 7. Six-year record of ambient noise at Pt. Sur. Note the annual increase in spectral level near 20 Hz due to migrating whales (from Ref. 32).

elsewhere.[31,32] Single hydrophone data in the 1–500 Hz band from 13 widely separated receivers in the North Pacific show that for the open ocean at 75 Hz sound levels are 3 dB higher than the median level 10% of the time and 6 dB higher 1% of the time. For coastal areas, sound levels are 7 dB higher than the median level 10% of the time and 15 dB higher 1% of the time. The most obvious feature in the records is a strong annual signal in the 15–22 Hz band with signal levels up to 25 dB above the background. It peaks in the autumn, and is almost certainly due to migrating blue and fin whales (Fig. 7). On average, whales are detected 43% of the time, and ships are detected 55% of the time. A comparison of seven years (1994–2001) of ambient noise data with two years of data taken in 1963–1965 using the same SOSUS receiver off the California coast near Sur Ridge shows a startling 10 dB increase between 20 and 80 Hz and up to 9 dB between 100 and 400 Hz, presumably due to increased shipping. Figure 8 illustrates typical spectra for ship, whale and earthquake signals. When whales or ships are present the noise is significantly non-Gaussian (Fig. 9).

The 2-D NPAL array allowed measurement of the vertical directionality of the noise field as well as the coherence of noise sources near the array.[33] The practical implication of sensor-to-sensor noise coherence is signal gain greater than the 10 log *N* achievable in an incoherent noise field. In general, coherence is low when no identifiable noise sources are present. The degree of coherence with nearby ships and whales is difficult to estimate because the array is aliased at

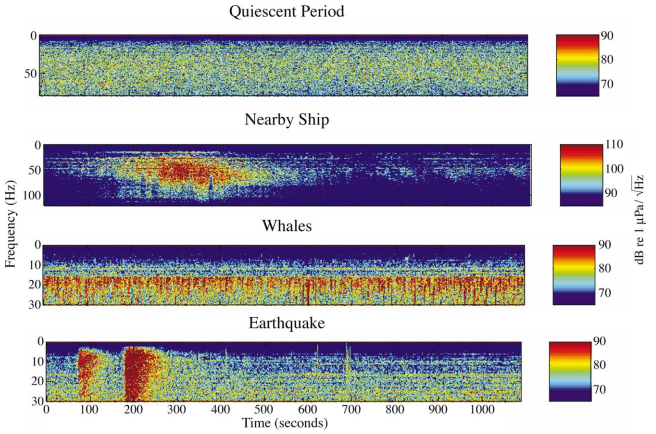


FIG. 8. Typical signals received by the hydrophone at 720 m depth on one of the 2-D receiver's vertical arrays. Average background level during the quiescent period is about 75 dB *re* 1 μPa/$\sqrt{\text{Hz}}$ (from Ref. 33).

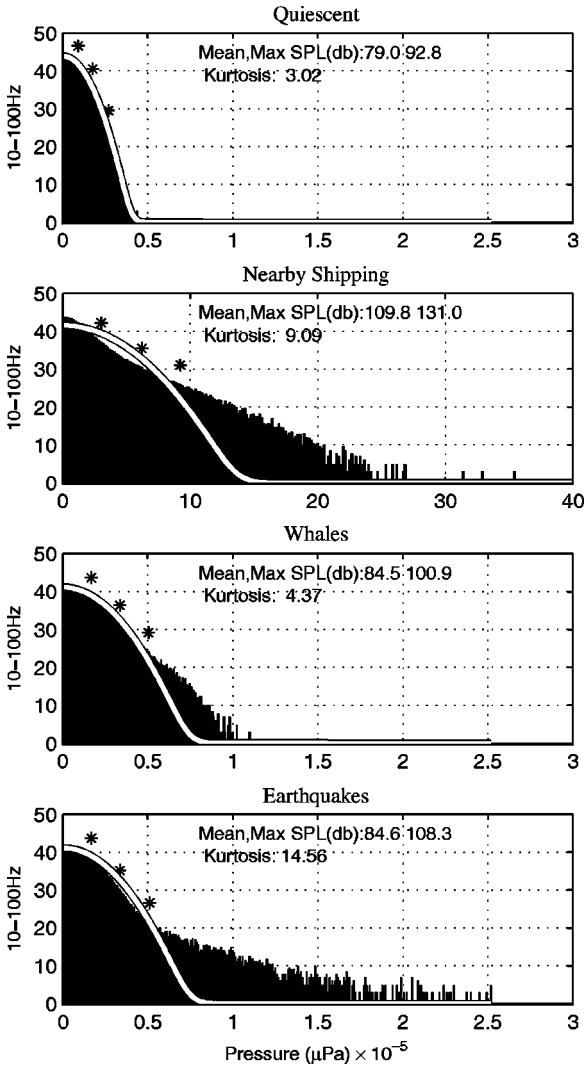


FIG. 9. Histogram of data from the hydrophone at 1020 m depth on VLA5 (Fig. 2) of the 2-D array, showing severe departure from Gaussianity during nonquiescent periods. (Gaussian noise has a kurtosis of 3.) (from Ref. 33).

1.2 Hz in the horizontal, and 20 Hz in the vertical, so that only very strong signals can be discerned unambiguously. However, when strong enough signals are present, such as 10–20 Hz whale tones, coherence was high across both dimensions of the array. Similarly, earthquake noise is strongly coherent across the array. Vertical beamforming of these strong signals indicates a highly directional spectrum at low grazing angles. There is very little directionality during periods when such signals are absent.

The long time series from so many locations make it clear that the usual characterization of ambient noise by mean and peak levels with Gaussian statistics is inadequate. More than half the time it contains components—whales, ships, earthquakes—that produce significant and rapid changes as well as departures from Gaussianity.

H. Marine mammals

The sensitivity of marine mammals to low-frequency sound has been a central issue in the series of experiments leading up to NPAL, starting with the 1990–91 Heard Island Feasibility Test (HIFT), and continuing through the Acoustic

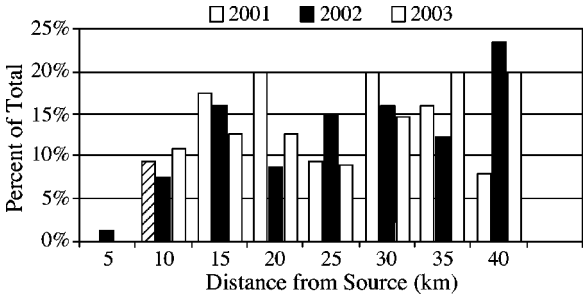


FIG. 10. Humpback whale distance from the Kauai transmitter. Although more humpbacks occupied the farthest distance category (35–40 km) during years when the source was active (2002 and 2003), an overall comparison between these years and the control year when the source was inactive (2001) showed no statistically significant differences (from Ref. 40).

Thermometry of Ocean Climate (ATOC) project.[34–39] Extensive observations at Kauai and at Pioneer Seamount, the locations of the two 75 Hz transmitters, included ground-based, aerial and ship surveys. Some were conducted when the transmitters were turned on (experimental periods); some when off (control periods). Many were ''blind'' in that the surveyors were not informed of the status of the transmitter. Scaled-amplitude playback tests were also conducted using sources suspended from vessels. Typically there was a small but discernable change in the distribution of mammals as measured by the distance from the source, although not during all years.

The most recent set of aerial surveys, conducted off Kauai during one year when the transmitter was inactive (2001) and two years when it was transmitting (2001–2003), supports that conclusion.[40] The surveys were all performed during the peak humpback whale residency period (February–April). The data indicate no significant differences in distribution across the years (Fig. 10).

The small response, and the fact that it is observed only sporadically, leads to the conclusion that these acoustic sources do not cause biologically significant changes in the behavior of whales or elephant seals, the two principal mammal species studied.

In addition to studying the effects of low-frequency sound on marine mammals, NPAL ambient noise data were used to study the distribution of blue whales in the northeast Pacific.[41]

III. SUMMARY

The overall objective of NPAL, which does not end with publication of this volume, is to add to our understanding of the basic physics of low-frequency, long-range, broadband propagation, the effects of environmental variability on signal stability and coherence, the three-dimensional character of the sound and noise fields and the fundamental limits to signal processing at long-range imposed by ocean processes. The goal is to enable advanced signal processing techniques, including matched field processing and other adaptive array processing methods, for a variety of applications such as sonars, acoustic communications, acoustic navigation systems and acoustical oceanography.

TABLE I. Chronology of NPAL experiments. All of the receivers were autonomous vertical line arrays (VLA) on subsurface moorings with 35 m spacing between hydrophones, unless otherwise indicated.

Year	Experiment	Center frequency (Hz)	Source/source depth	Receiver/receiver depth	Range (km)	Duration
1989	SLICE89	250	HLF-5 804 m (moored)	50-elements 60-m spacing 0–3000 m (*R*/*P FLIP*)	1001.1	July 1989 (9 days)
1994	ATOC Acoustic Engineering Test (AET)	75	HX-554 652 m (*R*/*P FLIP*)	Near VLA: 20-elements 900–1600 m	87.3	Nov. 1994 (7 days)
				Far VLA: 20-elements 900–1600 m	3252.4	
1995–1996	ATOC95	75	HX-554 939 m (Pioneer Seamount)	Hawaii VLA (HVLA): 40-elements 300–1700 m	3515.1	Nov.–Dec. 1995 to Aug.–Sept. 1996
				Kiritimati VLA (KVLA): 40-elements 400–1800 m	5170.9	
1996	Alternate Source Test (AST)	28 and 84	HLF-6A 652 m (*M*/*V Independence*)	HVLA: 20-elements 300–1000 m (The lower half of the array failed prior to the AST)	3501.6	June–July 1996 (9 days)
				KVLA: 40-elements 400–1800 m	5157.9	
1998–1999	NPAL98	75	HX-554 939 m (Pioneer Seamount)	2-D array (3600-m horizontal aperture): four VLAs with 20 elements 400–1100 m and one VLA with 40 elements 200–1600 m	145.8	July 1998–June 1999
			HX-554 811 m (Kauai)		3889.8	
1998–1999	OWS-Papa	75	HX-554 939 m (Pioneer Seamount)	OWS-P VLA: 20 elements 200–900 m	2217.1	Sept. 1998–June 1999
			HX-554 811 m (Kauai)		3336.2	

Here we have summarized some recent results towards that end, and have highlighted some of the still missing pieces of the puzzle.

ACKNOWLEDGMENTS

We thank Allan Pierce, Editor-in-Chief, for allowing us to publish these papers in one volume and Alexandra Tolstoy, Associate Editor, for her assistance and patience in assembling this special section. The critiques provided by the reviewers she selected have been of great help in improving the manuscripts. We also gratefully acknowledge ONR, DARPA, NSF, NOPP and the U.S. Navy for their generous support of this work. We are particularly grateful to the Office of Naval Research, which has been responsible for the overwhelmingly large fraction of research leading to what is reported here, and for its continuing commitment to ocean acoustics.

APPENDIX A: TRANSMITTERS AND RECEIVERS

1. Acoustic transmitters

A. California and Kauai bottom mounted, 75 Hz

These two transmitters were installed as part of the ATOC project. One, on Pioneer Seamount, 88 km west of Pillar Point, California, transmitted from December, 1995, to December, 1998. The second, 14.8 km north of Haena Point,

Kauai, started transmitting in October, 1997, and continues today. The locations of the transmitters were chosen to be close to shore to reduce cabling and power requirements, and to be where the seafloor depth is about the depth of the sound channel axis. Both were on the seafloor. The Pioneer Seamount source, at 37° 20.555′N, 123° 26.712′W, was 939 meters deep; the Kauai source, at 22° 20.949′N, 159° 34.195′W, was 811 m deep. Each transmitted a 1023-digit phase-modulated pseudorandom *m*-sequence repeated 44 times every four hours for a day, followed by 3 days of no transmissions.[42,43] (There have been slight variations in transmission schedule from time to time, but almost all reported results have been based on this one.) Each digit of the sequence consists of 2 cycles of 75 Hz yielding a time resolution of 26.67 ms (37.5 Hz bandwidth). The entire 1023-digit sequence lasts 27.28 seconds, and 44 repetitions take 1200.32 seconds. The source level was gradually ramped up over approximately five minutes preceding the nominal transmission time, from 165 dB *re* 1 μPa at 1 m (0.263 W) to 195 dB *re* 1 μPa at 1 m (263 W), in accord with the requirements of the various authorizations for the operation of the sources.

This combination of parameters yields a time resolution of 26.67 ms, and a theoretical signal processing gain of 46.1 dB—30.1 dB from pulse compressing 1023 digits, 16.0 dB from coherently averaging 40 out of the 44 repetitions. (In practice, the processing gain actually achieved was somewhat less than this, because the signal typically does not remain coherent for the entire signal duration.)

B. Ship suspended, 75 Hz—ATOC Acoustic Engineering Test (AET)

This transmitter was suspended at 652 m, the approximate depth of the sound channel axis, from *R/P Flip* moored at 31° 02.05′N, 123° 35.42′W off the California coast during November, 1994. Fifty-four transmissions were made of the 1023-digit sequence described above, consisting of a mix of 10, 20, and 40-min transmissions, with 2 to 4 h between transmissions.

C. Ship suspended, 28 and 84 Hz—Alternate Source Test (AST)

This source was suspended from the GPS P-Code dynamically-positioned vessel *M/V Independence* at a depth of 652 m at 37° 17.508′N, 123° 35.000′W, 13.4 km southwest of the Pioneer Seamount 75 Hz source. It transmitted 28 Hz (fundamental) and 84 Hz (third harmonic) signals simultaneously. It transmitted a 1023-digit sequence as above, but with 3 cycles/digit at 28 Hz (and therefore 9 cycles/digit at 84 Hz). A total of 42 dual-frequency broadband transmissions were completed over two 4-day periods.

2. NPAL receivers

A. 2-D receiving array

The 2-D array consisted of five individual moorings, four equipped with 700 m long, 20-element vertical arrays, and one equipped with a 1400 m long, 40-element vertical array. Water depth at the array site, Sur Ridge, was 1800 m. The 20-element arrays spanned the depth range from 400 m to 1100 m; the 40-element array spanned 200 m to 1600 m. The five moorings were deployed along a straight line nearly perpendicular to a Kauai-to-Pt. Sur geodetic path, thereby forming a 3.6 km horizontal aperture at 3900 km range. The array center was at 36° 17.000′N, 122° 28.700′W.

The position of each array element (hydrophone) was tracked by a long-baseline acoustic navigation system so that signal phase distortion caused by movement of the moorings due to currents could be distinguished from distortions due to sound-speed changes along the acoustic propagation paths. The signals received by each hydrophone were recorded in autonomous digital data acquisition systems. A common time base was provided by synchronized clocks referenced to atomic (Rb) frequency standards in each data acquisition system to allow coherent processing and beamforming across the full horizontal and vertical aperture of the array.

The 2-D array was deployed primarily to receive signals from the 3900 km distant Kauai transmitter, and was programmed to acquire data at times corresponding to the Kauai source transmission schedule.

The vertical distance between hydrophones was 35 m (1.75λ at 75 Hz), which allows resolution of modes 1–10. Rays and (some) higher-order modes are resolved by time and vertical angle of arrival.

B. Hawaii and Kiritimati vertical arrays

These vertical arrays each with 40 hydrophones at 35 m spacing were in place from November–December 1995 until August–September 1996. (The lower half—20 hydrophones—of the Hawaii array failed after about 5 months.) As above, these arrays were tracked by long-baseline acoustic navigation systems, and each was equipped with a Rb standard time base to enable beamforming and coherent processing. The Kiritimati array provided the longest NPAL transmission path, 5171 km.

C. Ocean Weather Station Papa vertical array

This vertical array in the Gulf of Alaska at 50° 04.909′N, 144° 50.718′W consisted of 20 hydrophones at 35 m spacing. It was deployed to augment the ATOC network, but turned out to be useful in understanding the arrival pattern at the 2-D array off California.

D. Horizontal arrays

A variety of horizontal arrays primarily belonging to the U.S. Navy were used to receive all the transmissions outlined above, as well as ambient noise data.

APPENDIX B: SIGNAL PROCESSING

Signals at the Hawaii, Kiritimati, and Ocean Weather Station Papa vertical arrays were received for the most part with high signal-to-noise ratios. Straightforward signal processing (pulse compression, averaging, beamforming) was effective for resolving time and vertical arrival angle for early arriving paths. These arrays were all deployed in deep water with the array center well up in the water column out of reach of bottom or surface influences. The situation at the

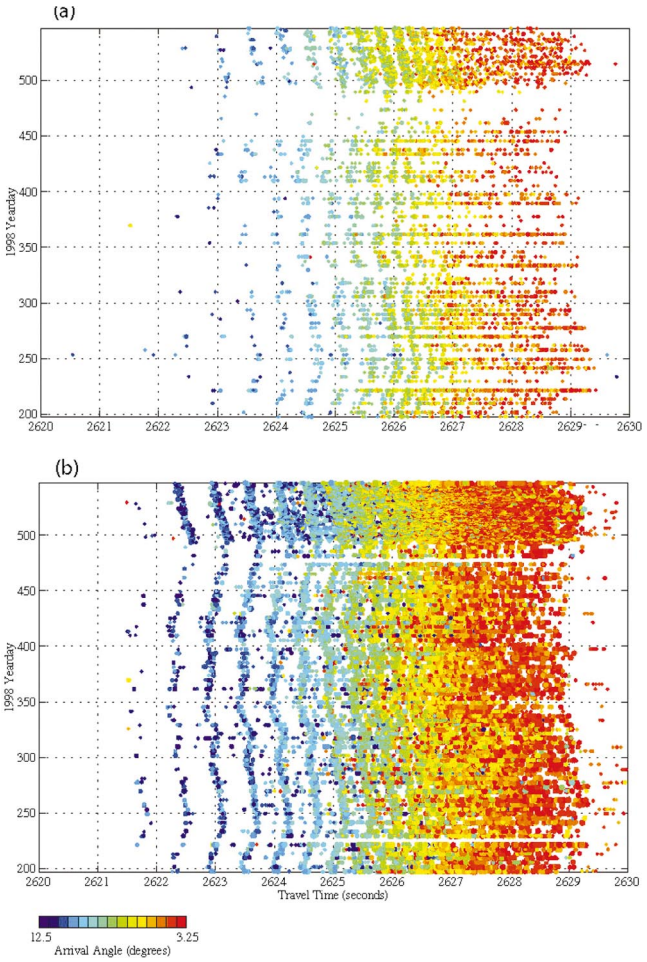


FIG. 11. Kauai transmissions received at the center vertical array (VLA3, Fig. 2) of the 2-D array. The arrivals span an interval of about 8 s. Dot size is proportional to the signal-to-noise ratio. Thus, a persistent strong arrival would appear as a vertical line, with meanders resulting from travel-time changes (sound-speed changes) along the acoustic path. The result of standard coherent processing is shown in the upper panel. The improved signal-to-noise ratio evident in the lower panel was achieved by applying a RAKE correlator with an assumed coherent bandwidth of 5 Hz, a coherence time of 600 s, and a depth coherence of 800 m (Ref. 45).

2-D array was disappointingly different. Receptions from the 3900 km distant Kauai source had low signal-to-noise ratios, and individual arrivals were difficult to resolve. The much higher ambient noise at this near-coast location than at the deep-water sites is partly responsible. In addition, however, interaction with the surface and bottom at the beginning and end of the Kauai-to-California path resulted in an unexpectedly complex arrival pattern.

For the high signal-to-noise deep water arrays a simple filter matched to the transmitted signal works well enough, even though it is not optimum because the received signal is not an exact replica of the transmitted signal—it has been randomized by ocean variability.[44] For the low signal-to-noise 2-D array a better estimator is the Rake correlator. Receptions were improved remarkably by applying this technique, which is used in many applications to combat multipath interference, especially in CDMA mobile telephone systems. It works by employing a set of correlators called Rake fingers, each delayed and scaled to a particular path. The receiver searches through the received signal to find where each finger peaks, and then sums the signals from the various paths. The improvement can be seen in Fig. 11. The lower panel shows clearly resolved early arrivals that are difficult to discern in the upper panel.

[1] P. F. Worcester, B. D. Cornuelle, J. A. Hildebrand, W. S. Hodgkiss, Jr., T. F. Duda, J. D. Boyd, B. M. Howe, J. A. Mercer, and R. C. Spindel, "A comparison of measured and predicted broadband acoustic arrival patterns in travel time-depth coordinates at 1000 km range," J. Acoust. Soc. Am. **95**, 3118–3128 (1994).

[2] This experiment has become known as the ATOC Acoustic Engineering Test (AET) and is referred to by that name in some of the papers in this volume.

[3] This experiment has become known as the Alternate Source Test (AST).

[4] A summary of many results up to 1999 is found in J. A. Colosi and the ATOC Group (A. B. Baggeroer, T. G. Birdsall, C. Clark, J. A. Colosi, B. D. Cornuelle, D. Costa, B. D. Dushaw, M. Dzieciuch, A. M. G. Forbes, B. M. Howe, D. Menemenlis, J. A. Mercer, K. Metzger, W. H. Munk, R. C. Spindel, P. F. Worcester, and C. Wunsch), "A review of recent results on ocean acoustic wave propagation in random media: Basin scales," IEEE J. Ocean. Eng. **24**, 138–155 (1999).

[5] S. M. Flatté, R. Dashen, W. H. Munk, K. M. Watson, and F. Zachariasen, *Sound Transmission Through a Fluctuating Ocean* (Cambridge University Press, Cambridge, England, 1979).

[6] R. K. Andrew, B. M. Howe, J. A. Mercer, and the NPAL Group (J. A. Colosi, B. D. Cornuelle, B. D. Dushaw, M. A. Dzieciuch, B. M. Howe, J. A. Mercer, W. H. Munk, R. C. Spindel, and P. F. Worcester), "Transverse horizontal spatial coherence of deep arrivals at megameter ranges," J. Acoust. Soc. Am., **117**, 1511–1526 (2205).

[7] S. M. Flatté and M. D. Vera, "Comparison between ocean-acoustic fluctuations in parabolic-equation simulations and estimates from integral approximations," J. Acoust. Soc. Am. **114**, 697–706 (2003).

[8] J. A. Colosi, A. B. Baggeroer, B. D. Cornuelle, M. A. Dzieciuch, W. H. Munk, P. F. Worcester, B. D. Dushaw, B. M. Howe, J. A. Mercer, R. C. Spindel, T. G. Birdsall, K. Metzger, and A. M. G. Forbes, "Analysis of multipath acoustic field variability and coherence in the finale of broadband basin-scale transmissions in the North Pacific Ocean," J. Acoust. Soc. Am., **117**, 1538–1564 (2005).

[9] A. G. Voronovich, V. E. Ostashev, and the NPAL Group (J. A. Colosi, B. D. Cornuelle, B. D. Dushaw, M. A. Dzieciuch, B. M. Howe, J. A. Mercer, W. H. Munk, R. C. Spindel, and P. F. Worcester), "Horizontal refraction of acoustic signals retrieved from North Pacific Ocean, Acoustic Laboratory billboard array data," J. Acoust. Soc. Am., **117**, 1565–1581 (2005).

[10] K. E. Wage, M. A. Dzieciuch, P. F. Worcester, B. M. Howe, and J. A. Mercer, "Mode coherence at megameter ranges in the North Pacific," J. Acoust. Soc. Am., **117**, 1565–1581 (2005).

[11] P. N. Mikhalevsky, A. Gavrilov, and A. B. Baggeroer, "The transarctic acoustic propagation experiment and climate monitoring in the Arctic," IEEE J. Ocean. Eng. **24**, 183–201 (1999).

[12] W. H. Munk, "Horizontal deflection of acoustic paths by mesoscale eddies," J. Phys. Oceanogr. **10**, 596–604 (1980).

[13] A. G. Voronovich and E. C. Shang, "A note on horizontal-refraction-modal tomography," J. Acoust. Soc. Am. **98**, 2708–2716 (1995).

[14] This term was first used in M. A. Wolfson and F. D. Tappert, "Study of horizontal multipaths and ray chaos due to ocean mesoscale structure," J. Acoust. Soc. Am. **107**, 154–162 (2000).

[15] Sometimes described as explosive growth of an acoustic beam.

[16] K. C. Hegewisch, N. R. Cerruti, and S. Tomsovic, "Ocean acoustic wave propagation and ray method correspondence: Internal wave fine structure," J. Acoust. Soc. Am., **117**, 1582–1594 (2005).

[17] I. P. Smirnov, A. L. Virovlyansky, and G. M. Zaslavsky, "Ray chaos, travel time modulation and sensitivity to the initial conditions," J. Acoust. Soc. Am., **117**, 1595–1606 (2005).

[18] W. H. Munk, P. F. Worcester, and C. Wunsch, *Ocean Acoustic Tomography* (Cambridge University Press, Cambridge, England, 1995).

[19] M. G. Brown, F. J. Beron-Vera, I. Rypina, and I. Udovydchenkov, "Rays, modes, wavefield structure and wavefield stability," J. Acoust. Soc. Am., **117**, 1606–1610 (2005).

[20] S. D. Chuprov and N. Y. Mal'tsev, "An invariant of the spatial-frequency interference pattern of the acoustic field in a layered ocean," Dokl. Akad. Nauk SSSR **257**, 2475–2479 (1981).

[21] F. J. Beron-Vera and M. G. Brown, "Ray stability in weakly range-dependent sound channels," J. Acoust. Soc. Am. **114**, 123–130 (2003).

[22] A. K. Morozov and J. A. Colosi, ''Entropy and scintillation analysis of acoustical beam propagation through ocean internal waves,'' J. Acoust. Soc. Am., **117**, 1611–1623 (2005).
[23] Bottom interactions are less of an issue for the Pioneer Seamount source, because the bottom slope on Pioneer Seamount is much steeper than the slope on Kauai.
[24] K. D. Heaney, ''The Kauai near-source test (KNST): Modeling and measurements of downslope propagation near the North Pacific Acoustic Laboratory (NPAL) Kauai source,'' J. Acoust. Soc. Am., **117**, 1635–1642 (2005).
[25] M. D. Vera, K. D. Heaney, and the North Pacific Aboustic Laboratory (NPAL) Group (J. A. Colosi, B. D. Cornuelle, B. D. Dushaw, M. A. Dzieciuch, B. M. Howe, J. A. Mercer, W. H. Munk, R. C. Spindel, and P. F. Worcester), ''The effect of bottom interaction on transmissions from the North Pacific Acoustic Laboratory (NPAL) Kauai source,'' J. Acoust. Soc. Am., **117**, 1624–1634 (2005)
[26] P. F. Worcester, B. D. Cornuelle, M. A. Dzieciuch, W. H. Munk, B. M. Howe, J. A. Mercer, R. C. Spindel, J. A. Colosi, K. Metzger, T. G. Birdsall, and A. B. Baggeroer, ''A test of basin-scale acoustic thermometry using a large-aperture vertical array at 3250-km range in the eastern North Pacific Ocean,'' J. Acoust. Soc. Am. **105**, 3185–3201 (1999).
[27] J. L. Spiesberger and K. Metzger, ''A new algorithm for sound speed in seawater,'' J. Acoust. Soc. Am. **89**, 2677–2688 (1991).
[28] B. D. Dushaw, P. F. Worcester, B. D. Cornuelle, and B. M. Howe, ''On equations for the speed of sound in seawater,'' J. Acoust. Soc. Am. **93**, 255–275 (1993).
[29] J. L. Spiesberger, ''Is Del Grosso's sound-speed algorithm correct?'' J. Acoust. Soc. Am. **93**, 2235–2237 (1993).
[30] For a more complete discussion of the ATOC measurements see ATOC Consortium (A. B. Baggeroer, T. G. Birdsall, C. Clark, J. A. Colosi, B. D. Cornuelle, D. Costa, B. D. Dushaw, M. Dzieciuch, A. M. G. Forbes, B. M. Howe, D. Menemenlis, J. A. Mercer, K. Metzger, W. H. Munk, R. C. Spindel, P. F. Worcester, and C. Wunsch), ''Ocean climate change: comparison of acoustic tomography, satellite altimetry, and modeling,'' Science **281**, 1327–1332 (1998); B. D. Dushaw, B. M. Howe, J. A. Mercer, R. C. Spindel, and the ATOC Group, ''Multimegameter range acoustic data obtained by bottom hydrophone arrays for measurement of ocean temperature,'' IEEE J. Ocean. Eng. **24**, 202–214 (1999); B. D. Dushaw, G. Bold, C.-S. Chui, J. Colosi, B. Cornuelle, Y. Desaubies, M. Dzieciuch, A. Forbes, F. Gaillard, J. Gould, B. Howe, M. Lawrence, J. Lynch, D. Menemenlis, J. Mercer, P. Mikhaelvsky, W. Munk, I. Nakano, F. Schott, U. Send, R. Spindel, T. Terre, P. Worcester, and C. Wunsch, ''Observing the ocean in the 2000's: A strategy for the role of acoustic tomography in ocean climate observation,'' in *Observing the Oceans in the 21st Century*, edited by C. J. Koblinsky and N. R. Smith (GODAE Project Office and Bureau of Meteorology, Melbourne, 2001), pp. 391–418.
[31] K. R. Curtis, B. M. Howe, and J. A. Mercer, ''Low-frequency ambient sound in the North Pacific: Long time series observations,'' J. Acoust. Soc. Am. **106**, 3189–3200 (1999).
[32] R. K. Andrew, B. M. Howe, J A. Mercer, and M. A. Dzieciuch, ''Ocean ambient sound: Comparing the 1960s with the 1990s for a receiver off the California coast,'' ARLO **3**, 65–69 (2002).
[33] A. B. Baggeroer, E. K. Scheer, and the NPAL Group (J. A. Colosi, B. D. Cornuelle, B. D. Dushaw, M. A. Dzieciuch, B. M. Howe, J. A. Mercer, W. H. Munk, R. C. Spindel, and P. F. Worcester), ''Statistics and vertical directionality of low frequency ambient noise at the North Pacific Acoustic Laboratory (NPAL) site,'' J. Acoust. Soc. Am., **117**, 1643–1665 (2005).
[34] W. W. L. Au, P. E. Nachtigall, and J. L. Pawloski, ''Acoustic effects of the ATOC signal (75 Hz, 195 dB) on dolphins and whales,'' J. Acoust. Soc. Am. **101**, 2973–2977 (1997).
[35] D. P. Costa, D. E. Crocker, D. M. Waples, P. M. Webb, J. Gedamke, D. S. Houser, P. D. Goley, B. J. Le Boeuf, and J. Calambokidis, ''The California Marine Mammal Research Program of the Acoustic Thermometry of Ocean Climate Experiment,'' in *California and the World Ocean '97*, San Diego, California, March 24–27, 1997, edited by O. T. Magoon *et al.* (American Society of Civil Engineers, Reston, VA, 1998), pp. 1542–1553.
[36] A. S. Frankel and C. W. Clark, ''Results of low-frequency playback of *M*-sequence noise to humpback whales, *Megaptera novaeangliae*, in Hawai'i,'' Can. J. Zool. **76**, 521–535 (1998).
[37] A. S. Frankel and C. W. Clark, ''Behavioral responses of humpback whales (*Megaptera novaeangliae*) to full-scale ATOC signals,'' J. Acoust. Soc. Am. **108**, 1930–1937 (2000).
[38] A. S. Frankel and C. W. Clark, ''ATOC and other factors affecting the distribution and abundance of humpback whales (*Megaptera novaeangliae*) off the north shore of Kauai,'' Marine Mammal Sci. **18**, 644–662 (2002).
[39] D. P. Costa, D. E. Crocker, J. Gedamke, P. M. Webb, D. S. Houser, S. B. Blackwell, D. Waples, S. A. Hayes, and B. J. Le Boeuf, ''The effect of a low-frequency sound source (acoustic thermometry of the ocean climate) on the diving behavior of juvenile northern elephant seals, *Mirounga angustirostris*,'' J. Acoust. Soc. Am. **113**, 1155–1165 (2003).
[40] J. R. Mobley, Jr., ''Assessing responses of humpback whales to North Pacific Acoustic Laboratory (NPAL) transmissions: Results of 2001–2003 aerial surveys north of Kauai,'' J. Acoust. Soc. Am., **117**, 1666–1673 (2005).
[41] J. C. Burtenshaw, E. M. Oleson, J. A. Hildebrand, M. A. McDonald, R. K. Andrew, B. M. Howe, and J. A. Mercer, ''Acoustic and satellite remote sensing of blue whale seasonality and habitat in the northeast Pacific,'' Deep-Sea Res., Part II **51**, 967–986 (2004).
[42] The modulation angle, $\varphi = \tan^{-1}\sqrt{1023} = 88.21°$, is chosen to achieve a spectrum with a smooth $\sin x/x$ envelope.
[43] The Kauai source transmitted from 30 October 1997 to 3 October 1999, when the permit for its operation ended. The final authorization needed to resume transmissions was received on 23 January 2002, and the source transmissions started on 24 January 2002 for a 5-year period.
[44] A matched filter produces the maximum signal-to-noise output if the input signal is known exactly, and the background noise is white.
[45] M. A. Dzieciuch (private communication, 2002).

Transverse horizontal spatial coherence of deep arrivals at megameter ranges

Rex K. Andrew, Bruce M. Howe, and James A. Mercer
Applied Physics Laboratory, University of Washington, Seattle, Washington 98105

The NPAL Group (J. A. Colosi, B. D. Cornuelle, B. D. Dushaw, M. A. Dzieciuch, B. M. Howe, J. A. Mercer, W. H. Munk, R. C. Spindel, and P. F. Worcester)[a)]

(Received 12 April 2004; revised 9 November 2004; accepted 10 November 2004)

Predictions of transverse horizontal spatial coherence from path integral theory are compared with measurements for two ranges between 2000 and 3000 km. The measurements derive from a low-frequency (75 Hz) bottom-mounted source at depth 810 m near Kauai that transmitted *m*-sequence signals over several years to two bottom-mounted horizontal line arrays in the North Pacific. In this paper we consider the early arriving portion of the deep acoustic field at these arrays. Horizontal coherence length estimates, on the order of 400 m, show good agreement with lengths calculated from theory. These lengths correspond to about 1° in horizontal arrival angle variability using a simple, extended, spatially incoherent source model. Estimates of scintillation index, log-amplitude variance, and decibel intensity variance indicate that the fields were partially saturated. There was no significant seasonal variability in these measures. The scintillation index predictions agree quite well with the dataset estimates; nevertheless, the scattering regime predictions (fully saturated) vary from the regime classification (partially saturated) inferred from observation. This contradictory result suggests that a fuller characterization of scattering regime metrics may be required. [DOI: 10.1121/1.1854851]

PACS numbers: 43.30.Re [AIT] Pages: 1511–1526

I. INTRODUCTION

The theory of wave propagation in random media (WPRM) has had considerable success in its original context involving fields propagating through homogeneous turbulent media.[1–3] The application to long-range propagation in the ocean requires modifications: the medium is a waveguide, the background sound speed is depth dependent, and the sound speed fluctuations are inhomogeneous and anisotropic. Flatté *et al.*[4] developed a theory (henceforth FDMWZ) for signal statistics that incorporated these additional features. Predictions based on FDMWZ theory have been compared to measurements from the straits of Florida,[5] Cobb seamount,[6] the Azores,[7] and the Slice89[8] experiments. These experiments involved ranges up to 1000 km and used single hydrophones or multiple hydrophones attached to vertical moorings (i.e., vertical arrays). The Acoustic Thermometry of Ocean/Climate (ATOC) Engineering Test (AET) experiment[9] tested the theory over a much longer range of 3250 km, but again only vertical line arrays were used to sample the acoustic field.

Second moment measurements available with these configurations were temporal (time–time) correlations, frequency–frequency correlations, and vertical spatial correlations. A measurement not available in these experiments due to equipment configurations was the horizontal second spatial moment, i.e., the horizontal coherence. An opportunity to measure horizontal coherence at 75 Hz over multimegameter paths arose during the ATOC and follow-on North Pacific Acoustic Laboratory (NPAL) experiments.[10] In this paper we report on the use of the data from two NPAL horizontal line array receivers at ranges from 2000 to 3000 km for measuring the transverse horizontal coherence.

The type of signals under consideration here are the long-range low-frequency ATOC signals trapped in the SOFAR channel. This is essentially a waveguide problem; as is characteristic of a waveguide problem, the acoustic field separates into faster modes with high-angle wavenumber vectors and slower modes with low-angle wavenumber vectors. The partial fields associated with the faster modes are well characterized by ray theory, and these manifest themselves as earlier, ray-like arrivals. Partial fields associated with the slower modes are more mode-like, and comprise the late crescendo[11,12] (sometimes called the finale or coda). This ray-mode duality is nicely demonstrated by Grabb.[13]

The full acoustic field at the receiver can thus be characterized roughly as

$$p = p_1 + p_2 + \cdots + p_{\text{modal}}, \tag{1}$$

where p_1 is the "first" ray-like arrival, p_2 is the next ray-like arrival, and so on. p_{modal} represents the final crescendo, and the explicit space–time dependence on location $\mathbf{r}$ and time t has been suppressed. While rays and modes have been utilized for tomography in both deep and shallow water, the early arrivals in this study are best modeled by rays.[14] The statistical parameters—notably, in this paper, the spatial coherence—of wave fronts associated with individual ray-like arrivals are quantities of fundamental importance.

[a)]J. A. Colosi is at Woods Hole Oceanographic Institution, Woods Hole, MA 02543. B. D. Cornuelle, M. A. Dzieciuch, W. H. Munk, and P. F. Worcester are at Scripps Institution of Oceanography, University of California at San Diego, La Jolla, CA 92093. B. D. Dushaw, B. M. Howe, J. A. Mercer, and R. C. Spindel are at the Applied Physics Laboratory, University of Washington, Seattle, WA 98105.

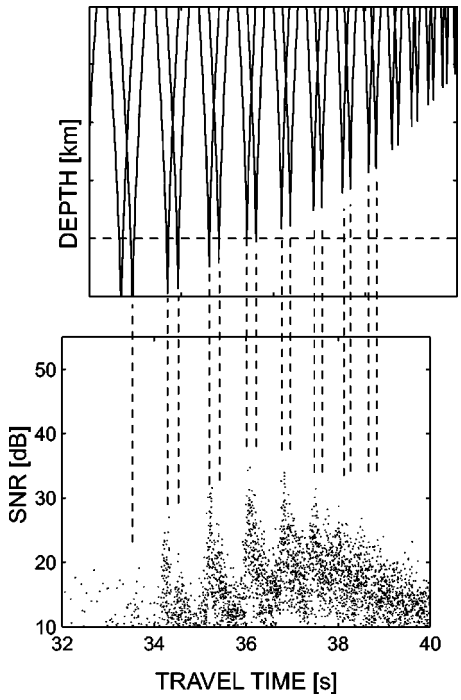


FIG. 1. Strength of time front arrivals at site N from a source mounted on a Pioneer seamount. (Site designations are shown later in this paper.) Top: time fronts at the range of site N. The receiver depth is suggested by the horizontal line. Bottom: the signal to noise ratio (SNR) versus arrival time scatterplots for 15 months of data. A correspondence between time front arrivals and strong SNR arrivals is indicated by dotted lines. The strongest SNR arrivals appear to be related to time fronts with cusps very near the receiver depth.

As described by Dushaw *et al.*,[15] the two receivers used here are generally too deep to measure the crescendo. As shown in Fig. 1, only early ray-like arrivals reach these receivers; all the later acoustic energy refracts above. The measured fields therefore contain only the first few arrivals, with no late modal contribution. These first few arrivals are termed ''deep arrivals'' in this paper. It can also be seen in Fig. 1 that there are signatures in the data for later arrivals, even though predictions of rays for later arrivals turn above the receiver depth (for these receivers). This phenomenon is not addressed in this paper.

This investigation separates out and estimates the statistics of the partial field $p_J(\mathbf{r},t)$, which was, on average, the strongest and most robust deep arrival at each receiver. The ordinal J represents one of the earliest detectable arrivals, but the poor signal-to-noise ratio of even earlier arrivals makes the exact enumeration of J impossible. An inspection of Fig. 1 appears to indicate that the strongest arrivals correspond to measurements made very near the cusps in the time front. This would agree with ray theory, which places caustics near ray turning points.

In addition to horizontal coherence, estimates are made of several intensity fluctuation statistics. These serve to identify the ''scattering regime'' of the signal fluctuations in FDMWZ theory. Intensity variances, and the regime itself, are also key predictions of FDMWZ theory that can be tested.

FDMWZ theory is based on a path-integral treatment of the stochastic wave propagation problem. It incorporates the depth-dependent sound speed profile and waveguide characteristics of the ocean by constructing the deterministic eigenray from source to receiver and considering sound speed fluctuations along that path to be the primary contribution to the path integral calculation. (A ''deterministic eigenray'' is conceptualized here as an eigenray traced through an ocean devoid of sound speed fluctuations.) Inhomogeneous and anisotropic sound speed fluctuations are represented by the Garrett–Munk interval wave spectrum.[16–18] Fluctuation calculations have been implemented numerically by the code ''Calculations of Acoustic Fluctuations due to Interval Waves'' (CAFI).[19,20] Empirical measurements in this paper are compared to FDMWZ predictions calculated by CAFI.

The remainder of the paper is organized as follows. In Sec. II we present the definitions of the mathematical entities that will represent the coherence and the intensity measures predicted by theory. In Sec. III we describe the NPAL experiment, including the *m*-sequence signal parameters. The data processing is described in Sec. IV: this concerns the calculation of quantities that are comparable to theoretical quantities. Experimental results are shown in Sec. V for the strongest deep arrival over each of two long-range paths. CAFI calculations for fluctuation statistics over representative paths of similar length are given in Sec. VI, with an analysis of the findings in Sec. VII. The findings indicate that the horizontal coherence length and scintillation index estimates are found to agree with CAFI predictions. Scattering regime classifications do not agree, however: the regime is inferred from actual measurements to be partially saturated, but CAFI identifies the regime as fully saturated. This contradictory result suggests that a fuller characterization of scattering regime metrics may be required. The main results are summarized in Sec. VIII.

II. THEORY

A. Coherence

In general, coherence is a measure of correlation, with values near one indicating a high correlation and values near zero poor correlation. However, there are subtle variations in the interpretation of coherence among investigators from different specialities. Motivated by these variations, the following discussion reviews several notions of coherence as a background for the definitions used in this analysis.

Coherence is generally a normalized second moment. Letting $\psi(\mathbf{r}_l,t_l)$ represent a general complex scalar field (e.g., an acoustic field) at location $\mathbf{r}_l$ and time t_l, and using the shorthand notation $\psi(l)=\psi(\mathbf{r}_l,t_l)$, the second moment is

$$M_{\psi\psi}(l,m)=\langle \psi^*(l)\psi(m)\rangle, \tag{2}$$

where $(\cdot)^*$ denotes conjugation and $\langle\cdot\rangle$ expectation. This is the complex cross-correlation from elementary statistics. In the statistical optics literature,[21,22] the *degree of complex coherence* is defined as Eq. (2), normalized by the ''self-'' (or autocorrelation) terms:

$$\gamma_{lm}=\frac{M_{\psi\psi}(l,m)}{\sqrt{M_{\psi\psi}(l,l)M_{\psi\psi}(m,m)}}. \tag{3}$$

This is essentially a complex correlation coefficient with a magnitude in the range [0,1]. $|\gamma_{lm}|=1$ implies full coher-

ence, and $|\gamma_{lm}|=0$ implies complete incoherence.

In the theoretical literature of wave propagation in random media (WPRM), it is common to assume that the fields are normalized (i.e., the self-terms are 1), and

$$\gamma_{lm}=\Gamma(\mathbf{r}_l,\mathbf{r}_m,t_l,t_m), \tag{4}$$

is usually referred to simply as the *mutual coherence*. This could, in general, be complex, but often only its magnitude is of interest: in this paper, it is necessary to consider the full complex coherence. Note also that this quantity can involve *any* two stochastic processes, $\psi(\mathbf{r}_l,t_l)$ and $\psi(\mathbf{r}_m,t_m)$, and is a *time–domain* definition.

A quantity similar to Eq. (3) appears in time series analysis,[23] where it is simply called the *coherence*. Here, the moments are cross-spectra ($G_{xy}(f)$) and autospectra ($G_{xx}(f)$ and $G_{yy}(f)$) between two time series $x(t)$ and $y(t)$;

$$\gamma_{xy}(f)=\frac{G_{xy}(f)}{\sqrt{G_{xx}(f)G_{yy}(f)}}.$$

By comparison with Eq. (3), this is clearly seen to be a frequency domain version of the degree of complex coherence. One often sees the *magnitude squared coherence* $|\gamma_{xy}(f)|^2$ (MSC) in the literature.

Again, a similar quantity appears in statistical array processing. The elements of the array *cross-spectral density matrix*[24] are precisely the second moments between sensors l and m, with ψ representing the space–time acoustic field. The moment is again considered to be a narrow band or single-frequency concept. The complexity here must be retained: beamforming depends on it. Cox[25] has shown that the loss of signal coherence across an array is a fundamental limiting factor in array system performance.

The fundamental link between spatial coherence and array system performance motivates this analysis to adopt the frequency–domain, ''single-frequency'' interpretation of coherence. This choice enables a simpler and more direct mathematical discussion in this section, and also facilitates a direct comparison with a theoretical prediction in Sec. VI. There is unfortunately one main disadvantage: the data involve wideband signals and are commonly considered in the time–domain. (Indeed, after pulse compression, these kinds of experiments are usually considered pulse-propagation experiments.) The procedures and consequences for evaluating time–domain observations in a frequency–domain context are discussed in Sec. IV.

Following FDMWZ, let the acoustic pressure field at frequency $\overline{\omega}$ be written as

$$p(\mathbf{r},t;\overline{\omega})=p_0(\mathbf{r}_0,t_0;\overline{\omega})e^{i[\overline{\mathbf{k}}\cdot(\mathbf{r}-\mathbf{r}_0)-\overline{\omega}(t-t_0)]}\psi(\mathbf{r},t;\overline{\omega}). \tag{5}$$

Here, p_0 is the deterministic part of the propagation to some reference position $\mathbf{r}_0$ at time t_0. $\mathbf{k}$ is a ''mean'' incident wave number vector and $\overline{\omega}$ is a ''mean'' frequency. The exponential term provides a phase correction to p_0 to yield the deterministic field at $\mathbf{r}$ at time t. All the fluctuations are represented by ψ; this factor converts the deterministic field at $\mathbf{r}$ and t to a stochastic field $p(\mathbf{r},t)$. ψ can contain randomness in amplitude and phase, including perturbations to $\mathbf{k}$ (due, e.g., to refracting media or extended sources) and $\overline{\omega}$ (due to Doppler effects). The exponential term is required for beamforming and coherent summing, so we retain it in its full form where necessary to show how it affects various intermediate quantities. By definition, ψ is normalized such that $\langle|\psi|^2\rangle=1$. The second moment expression [Eq. (2)] now becomes (dropping the $\overline{\omega}$ notation)

$$\begin{aligned}M_{pp}(l,m)&=\langle p^*(\mathbf{r}_l,t_0)p(\mathbf{r}_m,t_0)\rangle\\&=|p_o(\mathbf{r}_0,t_0)|^2e^{i\overline{\mathbf{k}}\cdot(\mathbf{r}_m-\mathbf{r}_l)}\\&\quad\times\langle\psi^*(\mathbf{r}_l,t_0)\psi(\mathbf{r}_m,t_0)\rangle.\end{aligned} \tag{6}$$

In this and in the following development, the fields are considered only at the same time, i.e., both at $t=t_0$, and the time notation is dropped. The mutual coherence is therefore

$$\begin{aligned}\Gamma_{lm}&=\frac{\langle p_l^*p_m\rangle}{\sqrt{\langle|p_l|^2\rangle\langle|p_m|^2\rangle}}\\&=\frac{e^{i\overline{\mathbf{k}}\cdot(\mathbf{r}_m-\mathbf{r}_l)}\langle\psi_l^*\psi_m\rangle}{\sqrt{\langle|\psi_l|^2\rangle\langle|\psi_m|^2\rangle}}\\&=e^{i\overline{\mathbf{k}}\cdot(\mathbf{r}_m-\mathbf{r}_l)}\langle\psi_l^*\psi_m\rangle,\end{aligned} \tag{7}$$

where the last line follows because $\langle|\psi_l|^2\rangle=\langle|\psi_m|^2\rangle=1$ by definition. When the phase fluctuations of the product $\psi_l^*\psi_m$ are zero mean, the expectation in the final expression above will be real. In the more general case, the expectation may provide a residual phase and hence be complex. Residual phase and the deterministic phase correction are eliminated by considering in this paper only the mutual coherence magnitude. This expression is formally valid at the single frequency $\overline{\omega}$, although, in practice, there will be some finite bandwidth associated with actual measurements.

Some assumptions can be invoked to simplify the problem. The mutual coherence is generally a six-dimensional function of spatial coordinates. When the field ψ is homogeneous, the mutual coherence is only a function of the difference $\mathbf{R}=\mathbf{r}_l-\mathbf{r}_m$:

$$\Gamma(\mathbf{r}_l,\mathbf{r}_m)=\Gamma(\mathbf{R}). \tag{8}$$

In general, ψ is definitely not homogeneous everywhere in a waveguide. The sound speed fluctuations themselves are not homogeneous; in the vertical, for example, they are largest near the surface and decrease rapidly with depth. However, since the horizontal scales of the sound speed fluctuations are orders of magnitude larger than the vertical scales,[26] it is therefore reasonable to assume that homogeneity holds, at least approximately, for horizontal sensor configurations of modest size compared to horizontal sound speed fluctuation correlation lengths, as is the case here.

While the self-term $\Gamma(\mathbf{r},\mathbf{r})=1$ by definition, the incoherent limit may not be 0. Conceptually, at great separation, (i.e., where $R=|\mathbf{R}|\to\infty$),

$$\begin{aligned}\lim_{R\to\infty}\langle\psi^*(\mathbf{r}_l,t)\psi(\mathbf{r}_m,t)\rangle&=\langle\psi^*(\mathbf{r}_l,t)\rangle\langle\psi(\mathbf{r}_m,t)\rangle\\&=|\langle\psi(\mathbf{r},t)\rangle|^2\equiv\Gamma_\infty.\end{aligned} \tag{9}$$

The term Γ_∞ is real. Thus, the coherence at great separation should reach an asymptotic value proportional to the squared magnitude of the mean field (excluding attenuation). In the saturated regime, the mean field is identically zero, and it will turn out that, for these measurements, the mean field estimate is insignificantly different from zero. But it is worthwhile to note that it is incorrect to assume *a priori* that the mean field is zero, especially if the fluctuation regime is partially saturated or unsaturated.

Inasmuch as the receivers used here are nearly perfect horizontal line arrays, $\mathbf{R}$ will be modeled as some $R\hat{\mathbf{a}}$, where $\hat{\mathbf{a}}$ is a unit vector aligned along the mean bearing of the array. Thus

$$\Gamma(\mathbf{R})=\Gamma(R;\hat{\mathbf{a}}). \tag{10}$$

Furthermore, for all *horizontal* line arrays, $\hat{\mathbf{a}}\perp\hat{\mathbf{z}}$ (where $\hat{\mathbf{z}}$ is the vertical unit vector.) For transverse measurements of coherence, $\hat{\mathbf{a}}$ must also be perpendicular to $\mathbf{k}$. This is not the case for the two arrays used here, so a transverse separation u is defined that is related to the actual separation along the array via the projection

$$u=R\hat{\mathbf{a}}\cdot\hat{\mathbf{t}}, \tag{11}$$

where $\hat{\mathbf{t}}$ is a unit vector transverse to the propagation such that $\hat{\mathbf{t}}\perp\mathbf{k}$ and $\hat{\mathbf{t}}\perp\hat{\mathbf{z}}$.

Combining Eqs. (8), (9), (10), and (11), a general expression for the transverse mutual coherence magnitude can be written as

$$|\Gamma(u)|=(1-\Gamma_\infty)g(u)+\Gamma_\infty. \tag{12}$$

This expression involves a real, unknown "shape" function $g(u)$ with the properties

$$g(u)=\begin{cases}1, & u=0,\\ 0, & u\to\infty.\end{cases} \tag{13}$$

The approach taken here is to model the unknown shape functions with the family $\hat{g}(u;\alpha)$, based on the single parameter α. A model $\hat{g}$ is fit in the least-square sense to the data, yielding best-fit parameter $\alpha_{\mathrm{LSE}}=\alpha_D$, and this function is shown for comparison against the spatial coherence function prediction produced by CAFI. A "transverse horizontal coherence length" L_h is defined such that

$$\hat{g}(L_h;\alpha_{\mathrm{LSE}}=\alpha_D)=e^{-1/2}. \tag{14}$$

[Another common definition of coherence length would be the "*e*-folding" length, the value at which the coherence is e^{-1}, but the definition in Eq. (14) will prove more convenient in Sec. VI.] The estimate $\hat{L}_h$ using α_D will be compared to L_h predicted by CAFI.

B. Scintillation index

The scintillation index SI is defined as

$$\mathrm{SI}=\mathrm{var}(I)/\langle I\rangle^2=\frac{\langle I^2\rangle-\langle I\rangle^2}{\langle I\rangle^2}=\frac{\langle I^2\rangle}{\langle I\rangle^2}-1, \tag{15}$$

and characterizes the variability of the acoustic intensity I. Although apparently a simple formula, it is important to recognize that the variance cannot be simply computed from the available time series if the measurements are correlated in time. Moreover, the transmissions were spaced irregularly in time, and this leads to additional interpretational complexities described below.

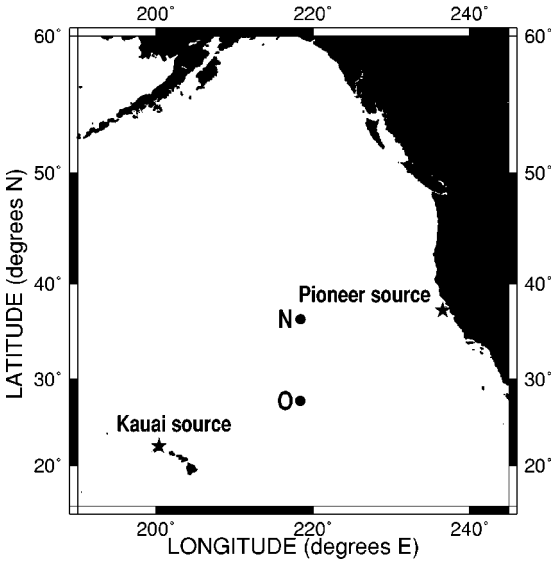


FIG. 2. Source and receiver configuration, NPAL experiment. Receiver locations are designated as N and O. Only transmissions from the Kauai source are presented in this paper.

III. EXPERIMENT

The locations of the instruments germane to the analysis presented in this paper are shown in Fig. 2. The Kauai source and receivers N and O were components of the larger NPAL experiment. The Pioneer source is shown for reference, but no further results for Pioneer transmissions beyond Fig. 1 are presented here.

The acoustic source located at Kauai was one of two very-low-frequency sources used in the ATOC/NPAL experiments.[27] It was deployed in July 1997 on the northeast slope of Kauai at latitude 22°20.949′N and longitude 159°34.195′W at 810 m depth, and is cabled to shore at the Pacific Missile Range Facility. Built by Alliant Techsystems (Mukilteo, Washington), it is a pressure-compensated bender-bar/barrel-stave transducer designed for operation at depths to 1300 m. At 810 m depth, it has a main resonant frequency of 65 Hz and a -3 dB bandwidth of 14 Hz.

The source transmitted from August 1997 to October 1999, when the permit for its operation ended. Transmissions were nominally scheduled for every 4 h every fourth day, but this scheme varied widely. During the first one-third of the experiment, some transmissions were only two hours apart. There were several system outages that precluded data acquisition for more than a month at either receiver site (e.g., June 1998). A new permit was established in 2001, and the source resumed transmissions starting 1 January 2002 for a 5 year period. The data analyzed here correspond to transmissions acquired from October 1997 to the end of permitted operation in October 1999.

The transmissions were standard ATOC *m*-sequence signals. The signal parameters are provided in Table I. Data acquisition was timed to begin after the first *m*-sequence and conclude prior to the last *m*-sequence, thereby ensuring that the acoustic fields at the receivers were fully involved during

TABLE I. Signal parameters for the Kauai source. The primary scientific transmission utilized 40 *m* sequences, but prefixed and suffixed guard sequences resulted in a total full-power transmission duration of 1227.28 s, which equates to 44.98 *m*-sequence periods.

Law	3471
Bits	1023
Cycles/bit	2
m-sequence period	27.28 s
Carrier frequency	75 Hz
−3 dB bandwidth	33.4 Hz
Sequences/transmission	40 (42)
Source power	260 W

data collection. The ''principle transmission'' consisted of 40 end-to-end sequences lasting 1091.20 s. The temporal and frequency characteristics are shown in Fig. 3.

Receivers N and O are bottom-mounted horizontal line arrays. The receivers are cabled to shore to equipment designed, built, and operated by the Applied Physics Laboratory, University of Washington (APL/UW). The signals were filtered between 2 (ac coupling) and 113.636 Hz (two-stage cascaded 6-pole Butterworth switched-capacitor filter), digitized at 300 Hz and spooled to disk. To reduce data storage requirements for each reception, every four consecutive *m*-sequences were coherently added, producing 10 ''4-period sums'' from each 40 *m*-sequence collection. The raw data was then discarded. All further data processing was conducted at APL/UW.

A total of 1204 receptions were collected at site N, and 1197 at site O. The temporal distributions of receptions at both sites are shown in Fig. 4.

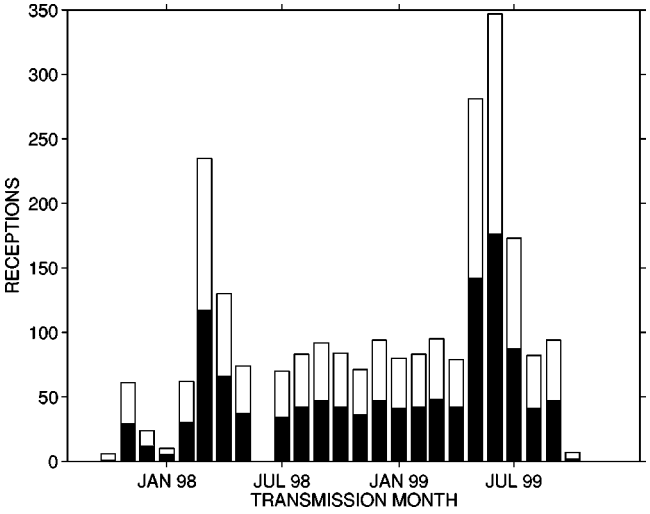


FIG. 4. Receptions per month, Kauai to sites N (white) and O (black). 1204 total receptions at site N, 1197 at site O.

IV. DATA PROCESSING

A. Preliminary processing

The first processing step (which occurred on the receiver computers) for each reception accumulated every four *m*-sequences into 4-period sums. For perfectly coherent signals in uncorrelated additive noise, this results in a processing gain of $10\log_{10}(4)\approx 4$ dB. Since, from Table I, a single *m*-sequence has a duration of 27.28 s, this step represents a time average of the fluctuations over 109.12 s. Coherent summing enhances correlated features and deemphasizes short-duration changes, and therefore this step may result in a bias toward overestimating the coherence.

In the second step, the 4-period sums were pulse-compressed[28] to achieve an additional theoretical gain of $10\log_{10}(1023)\approx 30$ dB.

Since the quantity ψ is usually defined in terms of solutions ''at a single frequency'' of the Helmholtz equation,[29] the next step in processing was to develop ''narrow band'' quantities to represent ψ. Conceptually, after pulse compression, the arrival structure that corresponded to the partial field $p_J(\mathbf{r},t)$ would be isolated with a ''time gate'' and transformed into the Fourier domain with a discrete Fourier transform (DFT). ψ is then represented by the single Fourier coefficient at the carrier frequency. The ideal time gate would contain all of the pulse arrival structure for the ray-like arrival under investigation, but not contributions from adjacent arrivals.

The procedure described above would be reasonable if the wideband (time–domain) field contained individually identifiable pulses, as is the case for the early ray-like arrivals in a deterministic ocean. However, in an actual stochastic ocean, particularly at the ranges considered here, each single deterministic arrival pulse is replaced by a complicated arrival structure spread consisting of multiple arrivals, which slowly fade in and out, flanking one or two (or no) ''main'' arrivals. Furthermore, as suggested in Fig. 1, the arrivals were grouped in pairs, with adjacent pairs separated by approximately 0.8 to 1.0 s. (One ray in the pair launches upward from the source, the other downward.) Very often, the pulse structure from each arrival in the pair had considerable overlap with that of the other.

The spreading of each pulse and the proximity of pulse pairs to one another therefore rendered infeasible the construction of a time gate small enough to isolate one arrival from the other in a pair. It was, however, usually possible to

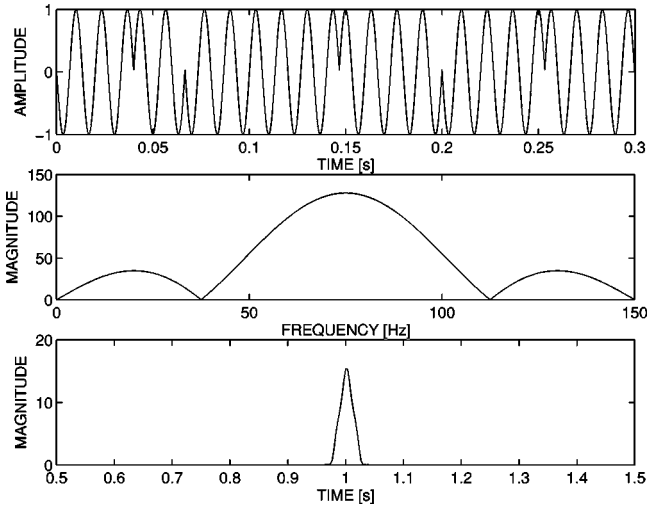


FIG. 3. Features of *m*-sequence processing. Top: a section of a simulated *m*-sequence carrier at 75 Hz. (A complete *m*-sequence is 27.28 s in length.) Middle: envelope of the Fourier transform magnitude. Bottom: output magnitude after pulse compression (designed with an arrival time of 1.0 s).

isolate one pair from another. A time gate width of 0.5 s was adequate for this task, but this gate width unavoidably comingled the fluctuations from both pulses. This may result in a bias toward underestimating the coherence (of a single partial field).

The effective bandwidth of the Fourier coefficient is the reciprocal of the time gate width and is therefore 2 Hz. This is probably a fair representation of a single-frequency theoretical quantity, particularly as the coherent bandwidth of the channel is thought to be substantially larger.

Over the entire experiment, the data displayed considerable wander. Wander can be defined on several scales for these paths.[30] Internal-wave-induced wander is roughly on the order of 10 ms; mesoscale-induced wander is roughly on the order of 100 ms; and gyrescale wander can produce wander on scales of 1000 ms.

If the small-scale wander produces a delay of τ across all channels, this will introduce an additional factor of $e^{-i\omega\tau}$ in Eq. (5). This factor, however, cancels in the second moment expression if the delay does not cause the pulse arrival structure to move out of the time gate.

The large-scale wander can move the pulse spread realization out of a fixed time gate. To automate processing, either the time gate must follow the large-scale wander, or the data must be synthetically shifted to correct for the large-scale wander. The gross behavior of the large-scale wander is called the *track*. All the data used here have been manually tracked,[30] and the track solution used as an input to shift the pulse structure to the center of the time gate.

An obliquely incident acoustic field will introduce a factor of $e^{i\mathbf{k}\cdot(\mathbf{r}_j-\mathbf{r}_0)}$ into the Fourier coefficient for channel j. Here $\mathbf{k}$ is the wave number vector of the incident field. As long as the entire arrival structure falls within the same time gate for all sensor channels, and if $\mathbf{k}=\mathbf{k}$ is deterministic, Eq. (7) shows that $|\Gamma|$ will be independent of obliquity.

In practice, for obliquely incident fields, a narrow time gate may exclude arrivals at the far ends of the array. The data in each channel have therefore been delayed (in the time domain) by a factor $\mathbf{k}_0\cdot(\mathbf{r}_j-\mathbf{r}_0)/\overline{\omega}$, where $\mathbf{k}_0$ is in the geodesic direction from Kauai to the point $\mathbf{r}_0$ on a WGS84 spheroid.[31] (This makes the signals appear as if they were incident from broadside.)

(*N.B.*: If the incident wavenumber vector $\mathbf{k}$ is varying with time, due, say, to time-varying horizontal refraction, the data contain *arrival angle wander*. Unlike time domain wander, this influence does not factor out in calculations of $|\Gamma|$. Averaging (i.e., expectation) in the presence of arrival angle wander will result in additional spatial decoherence.)

In summary, the processing steps from raw time domain data to ψ surrogate involved (1) four-period sums, (2) pulse compression, (3) large-scale wander correction, (4) obliquity correction, (5) time gate isolation, (6) Fourier transformation, and (7) coefficient extraction.

B. Coherence

It is possible to infer spatial coherence indirectly by examining the beamformed outputs of subarrays.[32] However, receivers N and O measure individual sensor signals that are strong enough, after four-period summing and pulse-compression, to stand out from the background noise. This makes it possible to estimate the spatial coherence directly by cross-correlating ψ from pairs of individual sensors, which is the approach taken here.

The processing steps described in Sec. IV A yielded, for an array with N sensors, an $N\times 1$ vector $\mathbf{v}_j$ of complex Fourier coefficients for the four-period sum group j. This procedure was then repeated to obtain a vector $\mathbf{v}_{j+1}$ for the four-period sum group $j+1$. These vectors $\mathbf{v}_j$ and $\mathbf{v}_{j+1}$ were then incorporated in an algorithm presented in Appendix A to form a coherence matrix estimate $\hat{\mathbf{G}}$ with elements

$$[\hat{\mathbf{G}}]_{jk}=|\hat{\Gamma}(\mathbf{r}_j-\mathbf{r}_k)|. \tag{16}$$

In FDWMZ theory, the transverse horizontal coherence magnitude is modeled as

$$|\langle\psi^*(x)\psi(x+\Delta x)\rangle|=\exp\left\{-\frac{1}{2}D(\Delta x)\right\}, \tag{17}$$

where $D(\Delta x)$ is a structure function (usually identified purely as a "phase structure function") and Δx is the transverse horizontal spatial separation. For small separations, the structure function can be approximated[33] by

$$D(\Delta x)\approx(\Delta x/L_h)^{3/2}, \tag{18}$$

where L_h is the transverse horizontal coherence length. Small separations are those for which

$$\Delta x\ll X_h\omega_i/\omega_X,$$

where X_h is the typical horizontal correlation size of sound speed inhomogeneities, ω_i is the local inertial frequency, and $\omega_X^2=\omega_i^2+n^2\tan^2\beta$ depends on the local buoyancy frequency n and the inclination β of the local eigenray. Flatté and Stoughton suggest $X_h\approx 12$ km, and, using typical Northern Pacific ocean values for the other parameters, $X_h\omega_i/\omega_X\approx 3.4$ km. The sensor separations used in this analysis are generally less than this, and hence Eq. (18) is a valid approximation for the structure function here. Based on Eqs. (17) and (18), the model shape function $\hat{g}(\Delta x)$ [cf., Eq. (13)] is

$$\hat{g}(\Delta x;\alpha)=\exp\left\{-\frac{1}{2}\left(\frac{\Delta x\hat{\mathbf{a}}\cdot\hat{\mathbf{t}}}{\alpha}\right)^{3/2}\right\}; \tag{19}$$

Eq. (19) is combined with Eq. (12) to model the magnitude mutual coherence, and this is compared to Eq. (16). The model is fit to the data by minimizing over α the sum of the squared errors between the model and the estimated coherence matrix. This produces the best-fit parameter $\alpha_{\mathrm{LSE}}=\alpha_D$, and the transverse spatial coherence length estimate $\hat{L}_h$ is then the root of

$$(1-\Gamma_\infty)\hat{g}(\hat{L}_h;\alpha_D)+\Gamma_\infty-e^{-1/2}=0. \tag{20}$$

C. Scintillation

Data processing for the scintillation index computation was slightly different. Additive ambient noise will increase the apparent variability in the data, and this will bias the scintillation index high. This bias can be reduced by beamforming, but beamforming will reduce the variability over that of a point measurement due to the averaging inherent in

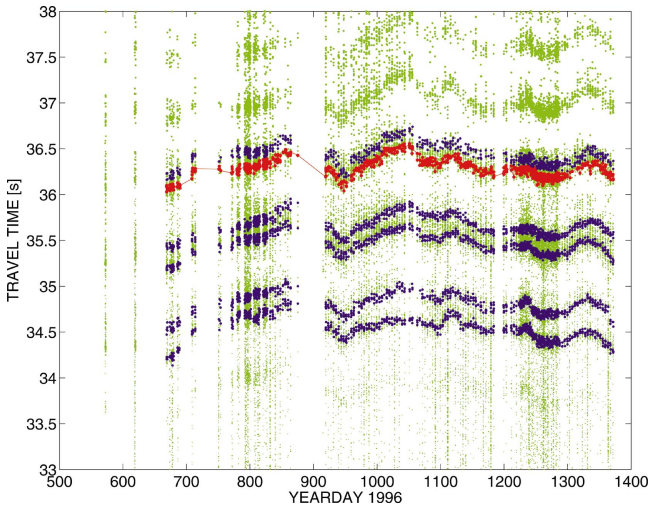


FIG. 5. Arrival structure visualizations, site N. Manual tracking can consistently identify six tracks, which are highlighted. The track for "Eigenray 5" is interconnected with line segments and shown in red.

the beamforming process. The advantage gained by beamforming generally outweighs the disadvantages, as long as the signal component across channels has a reasonably high cross-coherence. Therefore, each four-period sum was pulse compressed, corrected for large-scale wander, and delay-and-sum beamformed. The beamformer was steered to the geodesic path bearing from receiver to source assuming a WGS84 spheroid.[31] The beamformer used only those channels with cross-coherence greater than 0.5. The appropriate pulse arrival structure was extracted and Fourier transformed. The amplitude of ψ was then represented by the modulus of the carrier Fourier coefficient and the intensity by the modulus squared. The log-amplitude and log-intensity were then computed accordingly.

The data sample therefore consists of the amplitude (intensity, log-amplitude, log-intensity) of the carrier Fourier coefficient for each four-period sum, beamformed to increase the signal-to-noise ratio. The mean intensity and variances are determined from this set. Ensemble averages are interpreted as arithmetic means over the sample population.

V. RESULTS

A. Site N

1. Data description

The ray-like arrival analyzed for site N had, on average, the best signal-to-noise ratio (SNR) of all the ray-like arrivals for this site. An estimated track for this arrival structure is shown in Fig. 5. The arrival time variation of roughly 0.5 s over the dataset can be attributed to meso- and gyrescale ocean processes.

The data at site N had generally poorer SNR than the data at site O, and therefore a further description of the appearance of the raw data is deferred to Sec. V B.

2. Coherence length

The asymptotic value Γ_∞ was estimated [from Eqs. (7) and (9)] using

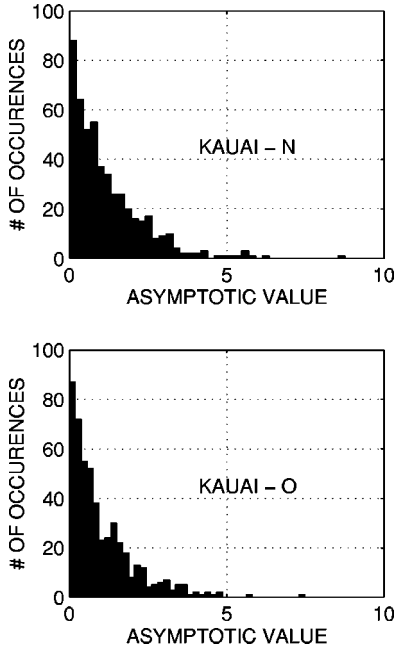


FIG. 6. Histograms of bootstrap replicates of the asymptotic value Γ_∞. Left: Kauai to site N. The estimated value of 0.13 is not significantly different from 0.0. Right: Kauai to site O. The estimated value of 0.042 is not significantly different from 0.0.

$$\Gamma_\infty = \frac{|\langle p \rangle|^2}{\langle |p|^2 \rangle}. \tag{21}$$

For this calculation, the second four-period sum group in each reception was used. Motivated by the same argument advanced in Sec. IV C, the multichannel four-period sum data was beamformed to improve SNR. The bearing chosen was the bearing of the geodesic path to Kauai. The beamformed data was pulse compressed, time gated, and Fourier transformed. The estimated asymptotic value using the full two-year dataset was 0.13. The samples were bootstrapped,[34] using 500 replicates, to approximate the distribution of this estimate. The histogram of the bootstrap replicates is shown in Fig. 6. Here 90% of the replicates were greater than the full sample estimate, so the asymptotic value $|\Gamma|_\infty$ was judged to be not significantly different from zero.

The data covariance matrix was estimated following a procedure described in the Appendix that utilized four-period sum groups, two and three from each reception. The sample mutual coherence matrix was fit to Eq. (12) with $\Gamma_\infty = 0$. The coherence length estimate was 528 m. The error in this estimate, represented by the bootstrap standard error using 200 bootstrap replicates, was ± 152 m.

3. Scintillation index

The scintillation index requires the variance of the intensity fluctuations, but care must be used in estimating the variance of a correlated sequence. The existence of a sample-to-sample correlation will generally bias the standard calculation. One alternative is to estimate the degrees of freedom in the signal and apply a corresponding correction. This would be a difficult undertaking because the raw four-period sums represent irregular clustered sampling in time. The

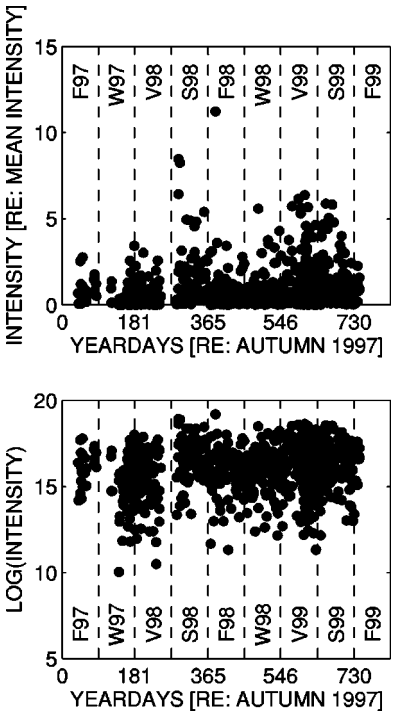


FIG. 7. Sample intensities, Kauai to site N. Top: intensity I (normalized by the arithmetic mean intensity). Bottom: log-intensity. The panels are partitioned by season, and annotated "W" for Winter, "V" for Spring, "S" for Summer, and "F" for Fall. The autumnal equinox in 1997 was on September 22.

shortest sampling interval is from one four-period sum to the next within a transmission, or about 109 s. The clusters are generally spaced four hours apart on transmission days, sometimes less. Estimating correlation measures for this kind of irregular clustered data is nontrivial.

An alternative approach was used here. A new sample population was derived from the original by using only a single four-period sum in each reception. (These results used the second four-period sum). This time series is shown in Fig. 7. The nonparametric run test[35] can be used to accept or reject the hypothesis that this new sample is random. The results are shown in Table II. When the new sample used all the available receptions, randomness was rejected at the 5% significance level. During the first third of the experiment, there were clusters of receptions only two hours apart. When receptions closer than four hours to the nearest neighbor were eliminated, the corresponding run test found the modified sample population to be random at the 5% level. Interestingly, this implies that the intensity was partially correlated on time scales of two hours.

Intensity fluctuation statistics (scintillation index, log-intensity variance, and decibel intensity variance) are shown in Table III for the modified sample population. In all cases,

TABLE II. Run test for randomness of the intensity fluctuations at site N. The randomness hypothesis is accepted at the 5% level of significance when the run test statistic is less than 1.965.

Sample	Sample size	Run statistic
Entire dataset	1204	3.29
Modified dataset	766	0.94

TABLE III. Intensity statistics: scintillation index (SI), variance of log intensity, and variance of intensity in decibels for site N. The top line is the full sample, and the bottom group is partitioned by season. Annotated "W" for Winter, "V" for Spring, "S" for Summer, and "F" for Fall. The errors are bootstrap standard errors for 200 replicates.

Population	SI	var(log(I))	var(I_{dB}) (dB^2)	Sample size
Full	1.53±0.13	2.02±0.14	$(6.17)^2 \pm (1.56)^2$	766
F97	0.83±0.22	1.06±0.25	$(4.47)^2 \pm (2.11)^2$	23
W97	1.69±0.36	2.38±0.51	$(6.70)^2 \pm (3.06)^2$	62
V98	1.44±0.24	2.34±0.39	$(6.65)^2 \pm (2.71)^2$	80
S98	1.13±0.20	1.57±0.29	$(5.44)^2 \pm (2.36)^2$	70
F98	2.25±0.89	1.98±0.38	$(6.11)^2 \pm (2.78)^2$	82
W98	1.43±0.25	1.76±0.30	$(5.77)^2 \pm (2.37)^2$	85
V99	1.47±0.16	2.02±0.24	$(6.18)^2 \pm (2.11)^2$	205
S99	1.00±0.12	1.43±0.20	$(5.20)^2 \pm (1.95)^2$	145
F99	0.43±0.20	1.29±0.55	$(4.93)^2 \pm (3.33)^2$	14

the quoted error is the bootstrap standard error. The variance of the decibel intensity was not significantly different than the Dyer value of 5.6 dB,[36] but the scintillation index was significantly greater than the saturation value of 1.0. The log-intensity variance was approximately 2.0. The sample population was also partitioned by season to determine if there were significant seasonal effects, but the sample sizes were too small to yield statistically significant differences.

Based on the intensity statistics, signal fluctuations at site N were much stronger than would be expected for the Rytov regime.[37] On the other hand, the scintillation index, which was greater than 1.0, shows that full saturation has not been achieved, even though the decibel-intensity variance was approximately the Dyer value. The regime was therefore classified as neither unsaturated nor saturated, but partially saturated.

B. Site O

1. Data description

The ray-like arrival analyzed for site O had, on average, the best SNR of all the ray-like arrivals at this site. An esti-

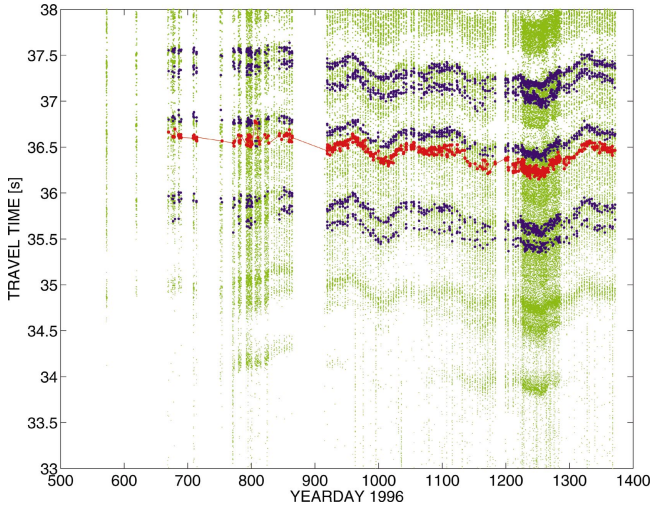


FIG. 8. Arrival structure visualization, site O. Manual tracking can consistently identify six tracks, which are highlighted. The track for "Eigenray 3" is interconnected with line segments and shown in red.

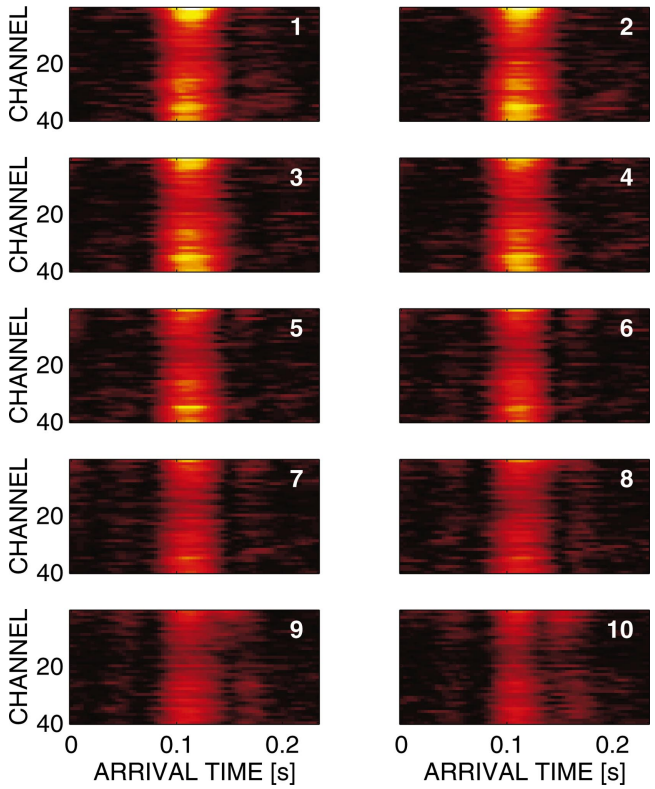


FIG. 9. The four-period sum data—Kauai to site O, transmission 00:00:00 yearday 305 1997. All panels use the same linear intensity color scale. The inset number is the four-period sum index. The signals on each channel have been artificially aligned to appear to arrive from the broadside.

mated track for this arrival structure is shown in Fig. 8. Again, the arrival time variability of roughly 0.5 s over the dataset can be attributed to meso- and gyrescale ocean processes.

Representative raw (individual channel) pulse-compressed four-period summed arrivals are shown in Fig. 9 over a single reception (ten, four-period sums.) A slow intensity change can be observed across the array over the ten pulses. The lower channels (say 1 to 5) had consistently higher signal power levels than the remaining channels. Since this was observed in data collected throughout the experiment, this was suspected to be due to gain mismatch. That this will not bias the results can be seen as follows. Let the generalized gain be represented by a complex channel-dependent coefficient h_l for channel l. Then from Eq. (7), Γ_{lm} would contain a term,

$$h_l^* h_m / \sqrt{|h_l|^2 |h_m|^2}.$$

The magnitude of this term is 1, and therefore generalized gain mismatch has no influence on $|\Gamma_{lm}|$.

Representative pulse arrival structures are shown in Fig. 10. These figures show the variations in the individual realizations, ranging from a single clearly identifiable peak (where there should be a pair) to a cluster of low-level peaks where none are uniquely identifiable.

2. Coherence length

The asymptotic value Γ_∞ was estimated using the second four-period sum group in each reception. The full population estimate was 0.042. A histogram of the bootstrap replicates is shown in Fig. 6. Here 98% of the replicates were greater than the full population estimate, so this value was also not significantly different from zero.

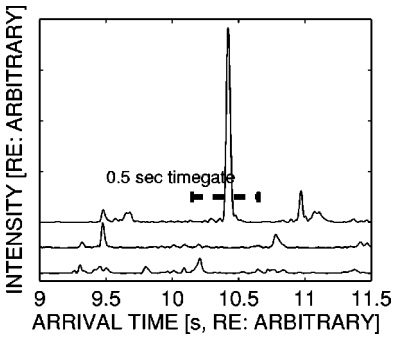


FIG. 10. Pulse structure, Kauai to site O. The data have been beamformed and incoherently averaged over ten, four-period sums. Top: transmission 00:00:00 1997 yearday 305. Middle: transmission 06:00:00 1998 yearday 189. Bottom: transmission 22:00:00 1998 yearday 341. All pulses shown on the same linear intensity scale: the top and middle pulses are translated up for visualization purposes. The arrival structure under analysis is centered approximately at 10.45 s (top), 10.15 s (middle), and 10.2 (bottom). A 0.5 s time gate is superimposed for reference: this location is appropriate for the top arrival.

The data covariance matrix was estimated using four-period sum groups two and three for each transmission. The sample mutual coherence matrix was fit to Eq. (12) with $\Gamma_\infty = 0$. The coherence length estimate was 410 m, and the bootstrap error in this estimate using 200 replicates was ± 25 m.

It is interesting to note that the coherence estimate from site N, which generally had data with a poorer SNR than that from site O, has a much greater variability, as evidenced by the bootstrap error. This is consistent with the convergence properties of the estimation algorithm in the Appendix: even though the estimator is unbiased, convergence still depends on SNR.

3. Scintillation index

The full sample population, derived from the original by using only the second four-period sum in each reception, is shown in Fig. 11. The nonparametric run test results are shown in Table IV. When the new sample population used all the available receptions, randomness was rejected at the 5% significance level. When receptions closer than 4 h to the nearest neighbor were eliminated, the corresponding run test found the modified sample population to be random at the 5% level.

Intensity fluctuation statistics are shown in Table V for the entire sample population. The quoted error is again the bootstrap standard error. The decibel intensity variance was again not significantly different than the Dyer value, but the scintillation index was significantly greater than the saturation value of 1.0. The log-intensity variance was again approximately 2.0. Again, there were no significant seasonal trends.

Based on the intensity statistics, signal fluctuations at site O were similar to site N and much stronger than those in the Rytov regime. Full saturation had not been achieved at

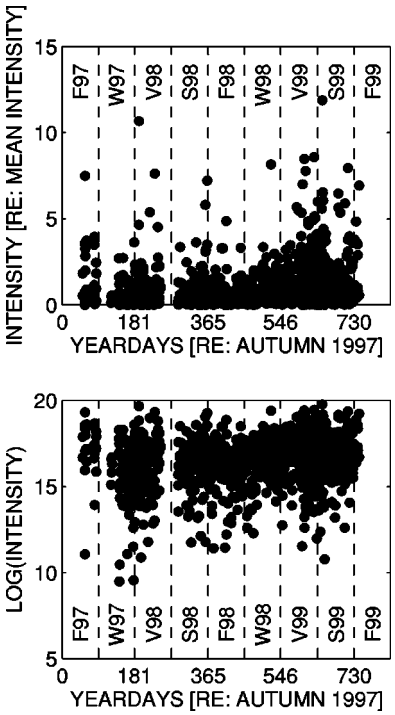


FIG. 11. Sample intensities, Kauai to site O. Top: intensity I (normalized by the arithmetic mean intensity). Bottom: log intensity. The panels are partitioned by season, and annotated "W" for Winter, "V" for Spring, "S" for Summer, and "F" for Fall. The autumnal equinox in 1997 was September 22.

this range, even though the decibel-intensity variance is approximately the Dyer value. The regime was also classified as neither unsaturated nor saturated, but partially saturated.

C. Summary of data results

A summary of the signal fluctuation estimators for the Kauai-N and Kauai-O paths is shown for reference in Table VI. In both cases, the asymptotic value Γ_∞ was judged by statistical inference to be negligible. The transverse coherence lengths were 528 ± 152 m and 410 ± 25 m at sites N and O, respectively. Given that both paths (Kauai-N and Kauai-O) involve similar ocean properties, and both receivers are at similar depths, one would expect the longer path (Kauai-N) to exhibit greater decoherence, i.e., a smaller coherence length. The data, however, suggests the opposite: the shorter path (Kauai-O) seems to have greater decoherence. Unfortunately, the measurements possess too much variability to establish this result at standard (95%, 99%) levels of significance. Nevertheless, the validity of this inference is supported by theoretical calculations presented in Sec. VI B and discussed further in Sec. VII.

TABLE IV. Run test for randomness of the intensity fluctuations at site O. The randomness hypothesis is accepted at the 5% level of significance when the run test statistic is less than 1.965.

Sample	Sample size	Run statistic
Entire dataset	1197	4.37
Modified dataset	766	0.51

TABLE V. Intensity statistics: scintillation index (SI), variance of log intensity, and variance of intensity in decibels at site O. The top line is the full sample, and the bottom group is partitioned by season. Annotated "W" for Winter, "V" for Spring, "S" for Summer, and "F" for Fall. The errors are bootstrap standard errors for 200 replicates.

Population	SI	var(log(I))	var(I_{dB}) (dB^2)	Sample size
Full	1.69 ± 0.16	2.01 ± 0.14	$(6.28)^2\pm(1.62)^2$	766
F97	0.61 ± 0.24	0.78 ± 0.18	$(3.83)^2\pm(1.88)^2$	13
W97	2.02 ± 0.46	2.67 ± 0.76	$(7.10)^2\pm(1.63)^2$	61
V98	2.47 ± 0.75	2.16 ± 0.31	$(6.38)^2\pm(2.48)^2$	81
S98	1.95 ± 0.45	2.37 ± 0.40	$(6.69)^2\pm(2.70)^2$	69
F98	1.63 ± 0.36	2.10 ± 0.38	$(6.30)^2\pm(2.61)^2$	83
W98	1.62 ± 0.67	1.66 ± 0.32	$(5.60)^2\pm(2.44)^2$	88
V99	1.30 ± 0.18	1.54 ± 0.19	$(5.39)^2\pm(1.82)^2$	213
S99	1.42 ± 0.30	1.98 ± 0.32	$(5.98)^2\pm(2.45)^2$	146
F99	0.93 ± 0.27	0.67 ± 0.20	$(3.56)^2\pm(1.99)^2$	12

The variance of decibel intensity for both datasets was not significantly different from the Dyer value; however, in both cases, the scintillation index was also considerably more than 1.0. The first finding suggests that the fields at both sites may have reached saturation, but the second finding shows that this is not true. A fuller theoretical understanding of these metrics is warranted.

VI. COMPARISONS WITH PATH INTEGRAL STATISTICS

A. Theoretical quantities

1. Horizontal coherence

FDMWZ theory provides an expression for the structure function $D(\Delta x)$,

$$D(\Delta x)=2k_*^2J(\Delta x), \tag{22}$$

where k_* is the wave number at the sound speed profile minimum and $J(\Delta x)$ is an integral expression derived from a path integral treatment; $J(\Delta x)$ is an integral of decoherence effects along a deterministic eigenray from the source to the receiver.

This integral is implemented numerically in CAFI using expansions and approximations described in Flatté and Stoughton.[33] As discussed in Sec. IV B, the computation for the separations under consideration in this paper is well modeled by the "small separation" approximation, given already in Eq. (18), which is repeated below:

$$D(\Delta x)\approx(\Delta x/L_h)^{3/2},$$

where L_h is again the transverse horizontal coherence length. Thus, the CAFI coherence length prediction can be determined directly from the separation at which the coherence falls to $e^{-1/2}$. This predicted coherence length is then compared to the experimentally measured coherence length.

TABLE VI. Summary of estimated transverse horizontal coherence length, scintillation index, log-intensity variance, and decibel-intensity variance over both paths, and the modified sample population size.

	Kauai-N	Kauai-O
Coherence length (m)	528 ± 152	410 ± 25
SI	1.53 ± 0.13	1.69 ± 0.16
var(log(I))	2.02 ± 0.14	2.01 ± 0.14
var(I_{dB}) (dB^2)	$(6.17)^2\pm(1.56)^2$	$(6.28)^2\pm(1.62)^2$
Modified sample size	766	766

TABLE VII. CAFI input parameters.

Parameter	Value
Source depth	810 m
Carrier frequency	75 Hz
Adiabatic gradient γ_A	1.2×10^{-5} s^{-1}
Ref. buoyancy freq. n_a	3.0 cph
Modal bandwidth j_*	3.0
rms displacement ζ_0	7.3 m
Ray launch angle	12.37°

2. Scintillation

CAFI models the intensity moments as

$$\langle I^n\rangle \approx n!\langle I\rangle^n\left[1+\frac{n(n-1)}{2}\gamma\right], \tag{23}$$

where γ is the "microray focusing parameter." Using Eq. (15), the CAFI prediction of SI is

$$\mathrm{SI}\approx 2\gamma+1.$$

B. Modeling results

Fluctuation calculations were conducted for two paths intended to represent the paths Kauai-N and Kauai-O. The primary CAFI input parameters are shown in Table VII. The calculations used representative ranges between 2000 and 3000 km for the two scenarios.

CAFI requires range-independent sound speed and buoyancy profiles. Range-independent profiles were generated from the annual Levitus climatology[3] by extracting temperature and salinity profiles every 100 km along each path, converting into sound speed[39] and Brunt–Väisälä buoyancy frequency,[40,41] and then averaging over the full range to yield the required input profiles. This averaging will smooth out local extremes in the buoyancy profile that might cause enhanced scattering, but is more consistent with the processing applied to the data, which averaged over nearly two full years.

CAFI also requires a profile of the (depth-dependent) sound speed variance. Denoting the random nondimensionalized sound speed fluctuations at depth z as $\mu(z)$, the variance profile is

$$\langle\mu^2(z)\rangle=\zeta_0^2\frac{n_a}{n(z)}\left(\frac{1}{c}\frac{dc}{dz}-\gamma_A\right)^2, \tag{24}$$

where ζ_0 is a rms internal wave vertical displacement, $n(z)$ is the buoyancy frequency as a function of depth, n_a is a reference buoyancy frequency, and γ_A is the adiabatic gradient. This variance profile was computed directly from the averaged sound speed and buoyancy frequency profiles using spline interpolation.

CAFI associates fluctuation calculations with deterministic rays launched at user-prescribed angles from the transmitter. For this comparison, the range-dependent code EIGENRAY[42] was used to calculate deterministic eigenrays from transmitter to receiver. The calculations used the same annual Levitus climatology as above, with sound speed profiles extracted again every 100 km along the path. Bathymetry was ignored, except that only upward-going rays at the bottom-mounted transmitter were considered.

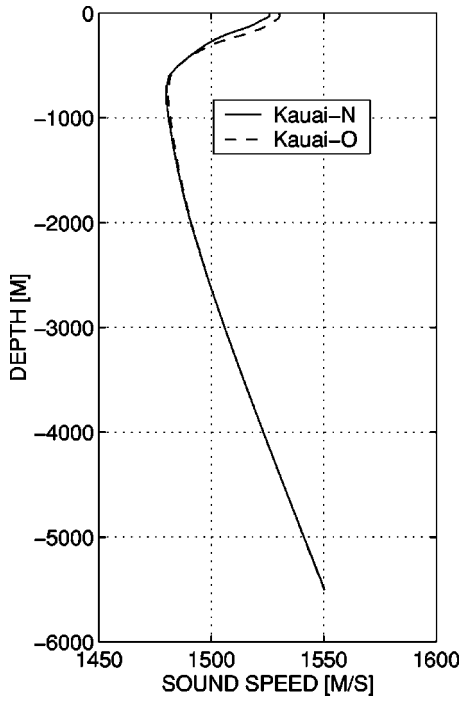


FIG. 12. Averaged sound speed profile derived from annual Levitus climatology profiles and averaged over the source–receiver path.

For both paths, EIGENRAY could not find wholly refracted–refracted rays from transmitter to receiver: the receivers were slightly too deep. As a compromise, the receiver depth was raised until an eigenray solution could be formed. This solution had a lower turning point directly above the true receiver. Such a representative early-arriving eigenray was chosen for each Kauai-N and Kauai-O path, and the initial launch angle of this ray was input to CAFI. The final CAFI results were not sensitive to small changes (of order 0.5°) in this initial launch angle.

The averaged sound speed and buoyancy frequency profiles, and the corresponding $\langle\mu^2\rangle$ profile for both paths are shown in Fig. 12, Fig. 13, and Fig. 14. A summary of CAFI calculations are shown in Table VIII. Statistics for both paths are similar, as might be expected, since these paths have similar lengths over similar regions of the ocean. The CAFI horizontal coherence functions are shown in Fig. 15. A least-squares estimate of L_h based on Eq. (18) yields $L_h=414.1$ m and $L_h=366.8$ m for Kauai-N and Kauai-O, respectively. The value γ yields via Eq. (23) scintillation indices of approximately 1.47 and 1.32 for the Kauai-N and Kauai-O paths, respectively. For reference, a diagnostic parameter ϕ is supplied by CAFI; the path-integral treatment is considered valid when this parameter is less than 1.0. For both runs, this validity condition holds.

Signal fluctuations are characterized in FDMWZ theory as belonging to one of several regimes, depending on the values of a "strength parameter" Φ and a "diffraction parameter" Λ. The strength parameter characterizes the average variance of the sound speed fluctuations along the path. The diffraction parameter characterizes the amount of diffraction from sound speed inhomogeneities.

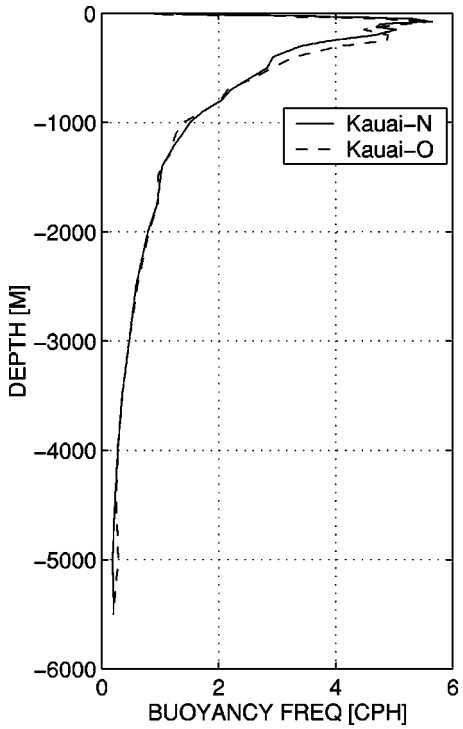


FIG. 13. Averaged buoyancy profile derived from annual Levitus climatology profiles and averaged over the source–receiver path.

For $\Phi>1$, the boundary $\Lambda\Phi=1$ in Λ, Φ space separates the ''fully saturated'' regime from the ''partially saturated'' regime. In both regimes, using the language of ray theory, the acoustic field separates into micromultipaths. In the partially saturated regime, the micromultipaths are partially correlated, but in the fully saturated regime, the micromultipaths are completely uncorrelated.

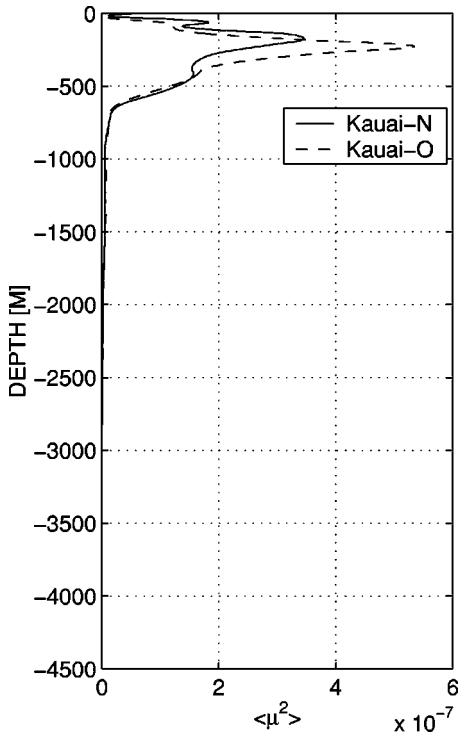


FIG. 14. Sound speed variance profile derived from the averaged sound speed and buoyancy profiles for each source–receiver path.

TABLE VIII. CAFI output parameters, range-independent paths representative of the Kauai-N and Kauai-O paths.

Parameter	Kauai-N	Kauai-O
Scattering regime	Fully saturated	Fully saturated
Diffraction parameter Λ	0.13	0.23
Strength parameter Φ	8.57	8.35
$\Lambda\Phi$	1.11	1.92
Transverse horizontal coherence L_h	414.1 m	366.8 m
Microray focusing γ	0.24	0.16
Scintillation index SI	≈1.47	≈1.32
ϕ (anisotropy)	0.027	0.023
ϕ (inhomogeneity)	−0.80	−0.79

The computed values of Λ and Φ, and specifically the product $\Lambda\Phi$, predict that both these rays are in the fully saturated regimes.

VII. DISCUSSION

Predictions and measurements show remarkably good agreement, even though the theory, which is based on the ray acoustics paradigm, is not thought to be valid at or below about 100 Hz, nor near caustics.[43] The agreement found in this paper suggests that FDMWZ theory makes accurate predictions of horizontal spatial coherence for deep arrivals at 75 Hz, at least for these scenarios.

Theoretical calculations indicate that the Kauai-O path, which is shorter, should have less coherence (i.e., more scattering) than the Kauai-N path. This surprising result is supported, at least to within a standard deviation, by measure-

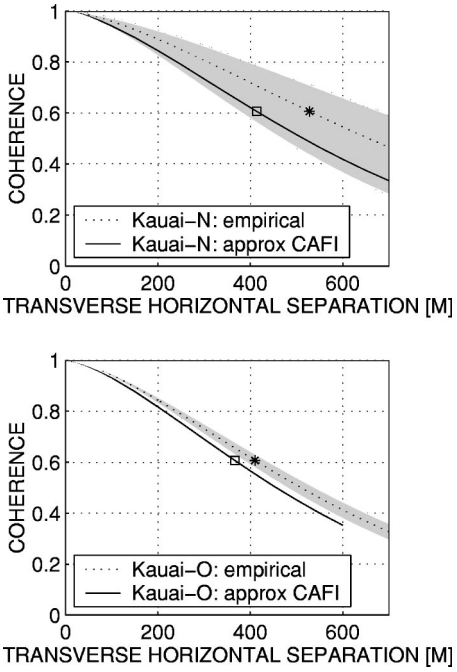


FIG. 15. A comparison of horizontal coherence models: top, Kauai-N path, bottom, Kauai-O path. The dotted lines represent the model based on empirical data, the solid lines represent the CAFI ''small separation'' approximation. ∗=coherence length estimate from data; □=coherence length from CAFI. The shaded region represents approximately the standard deviation error in the empirical model.

TABLE IX. Apparent angular widths of the Kauai source, over the entire experiment.

Path	L_h (m)	$\Delta\theta$ (°)
Kauai-N	410	1.46
Kauai-O	528	1.13

ments. One might expect the reverse to be true, based solely on path length. The sound speed variance profiles in Fig. 14 may explain this: there is more intense sound speed variability around 200 to 300 m in the Kauai-O path than in the Kauai-N path. This increased variability would result in more scattering over the Kauai-O path and therefore less coherence at site O. Thus, the path length alone does not provide a full characterization of spatial coherence—path-dependent sound speed fluctuation strength must also be considered.

The empirical measures are not precisely equivalent to the theoretical quantities. Given the individual contributions of each signal processing step, as discussed in Sec. IV A, it is reasonable to suspect that the empirical mutual coherence probably underestimates the true mutual coherence.

It is interesting to consider the consequences of arrival angle wander. By the Van Cittert-Zernike theorem,[44] the mutual coherence is the Fourier transform of the spatial spectrum of an extended incoherent source. Let the angle of arrival of radiation from the source be uniformly distributed between azimuthal angles $\bar{\theta}-1/2\Delta\theta$ and $\bar{\theta}+1/2\Delta\theta$ during the experimental period. The corresponding wave numbers are related by

$$k=\frac{2\pi}{\lambda}\sin\theta.$$

Then, to first order in $\Delta\theta$, the source spatial spectrum has the distribution

$$B(k)=\begin{cases}(\pi\Delta\theta)^{-1}, & k\in\left[\bar{k}-\dfrac{\pi\Delta\theta}{\lambda},\bar{k}+\dfrac{\pi\Delta\theta}{\lambda}\right]\\ 0, & \text{otherwise,}\end{cases}$$

where $\bar{k}=(2\pi/\lambda)\sin\bar{\theta}$ (valid when $\Delta\theta\ll 1$). The magnitude of the Fourier transform of $B(k)$ is therefore proportional to

$$\operatorname{sinc}\frac{x\pi\Delta\theta}{\lambda}.$$

From Eq. (14), there is a measure of $|\Gamma(x)|$ at $x=L_h$ equal to $e^{-1/2}$. Hence, set

$$\operatorname{sinc}\frac{L_h\pi\Delta\theta}{\lambda}=e^{-1/2}.$$

This is a transcendental equation with a principal root at $\pi L_h\Delta\theta/\lambda=1.6443$, or

$$\Delta\theta=\frac{\lambda}{\pi}\frac{1.6443}{L_h}.$$

The corresponding angular ''source widths'' are given in Table IX. If arrival angle variability were the only factor contributing to decoherence across the array, the variation in the incident field arrival angle could be attributed to seasonal horizontal sound speed gradients. However, since other factors contribute to decoherence, this angular estimate can only serve as an upper bound on the apparent influence of horizontal sound speed effects.

Two investigations into horizontal refraction along the path from the Kauai source to a receiver near Pt. Sur on the California coast (a path length of approximately 3847 km) inferred horizontal refraction angles introduced over that path that were slightly smaller than the ''angular width'' found above.[45,46] Although those analyses involved a different ocean path, the path lengths and ocean regions are similar to those reported here, and consequently the horizontal refraction angle effects might be of a similar order of magnitude in the Kauai-O and Kauai-N data. If this is true, then the spatial coherence at sites N and O due solely to scatter (i.e., the ''coherence of a single incident field'') would be slightly better than the measurements reported in Sec. V C, which are estimates containing the influence of both scatter and arrival angle wander.

VIII. SUMMARY

Predictions of transverse horizontal mutual spatial coherence from path integral theory were compared to measurements made on two U.S. Navy receiving arrays during the NPAL experiment in the North Pacific. The source was a 75 Hz bottom-mounted projector moored near the SOFAR channel axis at a depth of 810 m on the northern slope of Kauai. The source transmitted *m*-sequences with a period of 27.28 s for an average duty cycle of about 2% over the experiment.

Two receivers at ranges between 2000 and 3000 km had enough SNR after pulse compression and 4-period summation to allow the deep arriving signals to stand out from the background noise on individual channels. This made the direct computation of mutual coherence possible for these two receivers.

The data processing was designed to yield quantities comparable to theoretical single frequency quantities. Time gating was used after pulse compression to isolate partial fields corresponding to deterministic eigenrays, and Fourier coefficients with 2 Hz bandwidth were substituted for single-frequency quantities. A special algorithm was applied to estimate the coherence matrix in the presence of additive ambient noise without incurring the bias of ambient noise coherence. Realizations were averaged over the two-year span of the dataset. Statistics at both receivers were similar: the transverse horizontal coherence length measured about 400–500 m, the scintillation index about 1.5 to 1.6, and the intensity variance about equal to the 5.6 dB Dyer value. There was no significant seasonal trend in the intensity statistics.

Overall, the comparisons between the measurement and theoretical prediction of the second moment were quite good. Computations from CAFI were remarkably close, with predictions of horizontal coherence lengths of about 400 m and a scintillation index of about 1.3 to 1.5. An apparent decrease in the measured coherence on the shorter Kauai-O path was successfully predicted by CAFI. Measurements similar to

these, but in the fully-ensonified portion of the water column, preferably at multiple ranges, would augment the results reported here.

Empirical intensity statistics yield a parametrization of the scattering regime that is not well characterized in current theories. This is due to weaknesses in the treatment of the multiple forward scattering problem. The empirical scintillation indices, around 1.5 to 1.6, were quite a bit above the ''saturation'' limit of 1.0, but CAFI still identified the fluctuation regime as ''fully saturated.'' Further exploration and characterization of these scattering regimes is still needed.

An interesting auxiliary finding was the apparent correlation of intensity levels over two hour time spans. Future experiments should schedule a suite of transmissions at various intervals of about two hours to investigate the character of this correlation.

ACKNOWLEDGMENTS

We would like to acknowledge discussions with F. Henyey, T. Ewart, M. Wolfson, and B. Dushaw. We would also like to thank B. Dushaw for providing the eigenray arrival tracks, and the data for Figs. 1, 5, and 8. This work was supported by ONR code 321.

APPENDIX: UNBIASED COHERENCE MATRIX ESTIMATOR

Let the measured data on channel j be

$$x_j(t)=s_j(t)+n_j(t), \tag{A1}$$

where the signal $s(t)$ is a scintillating phase-encoded m sequence and $n(t)$ is additive ambient noise. The Fourier transform of Eq. (A1) is

$$X_j(\omega)=S_j(\omega)+N_j(\omega), \tag{A2}$$

where $X(\omega)=\mathcal{F}\{x(t)\}$ is the Fourier transform of $x(t)$, etc. In the following, it will be convenient to adopt vector-matrix notation. Fourier components at frequency ω (henceforth omitted) for each channel are organized into a vector ''snapshot'' across the array, leading to

$$\mathbf{x}=\mathbf{s}+\mathbf{n}, \tag{A3}$$

where $\mathbf{x}=[X_1,X_2,\ldots,X_N]^T$ for an N-sensor array. The data covariance matrix is then

$$\mathbf{C}_{xx}=\langle \mathbf{x}\mathbf{x}^{\mathrm{H}}\rangle \tag{A4}$$

$$=\langle \mathbf{s}\mathbf{s}^{\mathrm{H}}\rangle+\langle \mathbf{n}\mathbf{n}^{\mathrm{H}}\rangle \tag{A5}$$

$$=\mathbf{C}_{ss}+\mathbf{C}_{nn}, \tag{A6}$$

where $(\cdot)^{\mathrm{H}}$ denotes Hermitian transposition and $\langle\cdot\rangle$ expectation. Equation (A5) follows because the signal and noise processes are independent. If $\mathbf{x}_k$ denotes the kth-independent realization of the process $\mathbf{x}$, then a standard estimator for the data covariance matrix is the sample covariance matrix, defined as

$$\hat{\mathbf{C}}_{xx}=\frac{1}{K}\sum_{k=1}^{K}\mathbf{x}_k\mathbf{x}_k^{\mathrm{H}}, \tag{A7}$$

over K realizations.

Consider solely the signal covariance matrix $\mathbf{C}_{ss}$. This has elements

$$[\mathbf{C}_{ss}]_{jk}=\begin{cases}\sigma_j\sigma_k\rho_{jk}, & j\neq k,\\ \sigma_j^2, & j=k.\end{cases} \tag{A8}$$

By definition, σ_j^2 is real and positive, and ρ_{jk} is, in general, complex. $\mathbf{C}_{ss}$ is also called the cross-spectral density matrix. Using the transform

$$[\mathbf{G}]_{jk}=|[\mathbf{C}]_{jk}|/\sqrt{[\mathbf{C}]_{jj}[\mathbf{C}]_{kk}}, \tag{A9}$$

one obtains a matrix of correlation coefficients. Allowing that $s_j(t)$ is the acoustic field measured at sensor location $\mathbf{r}_j$, and using Eq. (7), $\mathbf{G}$ is simply seen to be a matrix notation for the samples of the mutual spatial coherence function magnitude at separations $\mathbf{r}_j-\mathbf{r}_k$. $\mathbf{G}$ will therefore be called the *coherence matrix*, and it contains all the information available regarding the functional shape of $|\Gamma(\mathbf{r}_j-\mathbf{r}_k)|$.

Inasmuch as the coherence matrix depends on the signal covariance matrix, it becomes centrally important to develop an unbiased estimator for the signal covariance matrix. Unfortunately, the signal covariance matrix is not available separately: the signal is measured in the presence of noise. The noise vector $\mathbf{n}$ can usually be well approximated by a complex multivariate Gaussian process with $\langle\mathbf{n}\rangle=\mathbf{0}$ and covariance matrix $\mathbf{C}_{nn}$. This noise covariance matrix possesses an intricate and usually unknown structure. This structured matrix assumes the role of a nuisance parameter, which, in a conventional estimation problem, will have to be estimated separately and removed. This cannot be accomplished for the current problem, but failure to account for this nuisance parameter will result in biased estimates of both the signal covariance matrix and hence the coherence matrix.

An alternative technique for constructing an unbiased estimator for the signal covariance matrix has been suggested by Wong *et al.*[47] In this technique, two snapshots are considered:

$$\mathbf{x}^{(a)}=\mathbf{s}^{(a)}+\mathbf{n}^{(a)},$$

$$\mathbf{x}^{(b)}=\mathbf{s}^{(b)}+\mathbf{n}^{(b)}.$$

Snapshots a and b are collected at two times t_a and t_b far enough apart that the two noise vectors have decorrelated. Then the covariance between these two vectors is ${}_N\mathbf{0}_N$. This allows

$$\mathbf{C}_{xx}=\mathbf{C}_{ss}.$$

This situation is shown schematically in Fig. 16.

The structure of $\mathbf{C}_{ss}$ in this situation is no longer the same as Eq. (A8) because $\mathbf{s}^{(a)}\neq\mathbf{s}^{(b)}$. Now,

$$[\mathbf{C}_{ss}]_{jk}=\begin{cases}\sigma_j\sigma_k\rho(\Delta r_{jk},\Delta t_{ab}), & j\neq k,\\ \sigma_j^2\rho(\Delta r_{jj}=0,\Delta t_{ab}), & j=k,\end{cases} \tag{A10}$$

where $\Delta t_{ab}=t_a-t_b$ and $\Delta r_{jk}=|\mathbf{r}_j-\mathbf{r}_k|$. The temporal separation between snapshots a and b introduces additional decorrelation. In addition, $\mathbf{C}_{ss}$ is, in general, no longer Hermitian.

If the total correlation $\rho(\Delta r_{jk},\Delta t_{ab})$ is approximated by a spatiotemporal separable form $\rho_r(\Delta r_{jk})\rho_t(\Delta t_{ab})$, the normalization of Eq. (A9) will recover a coherence matrix con-

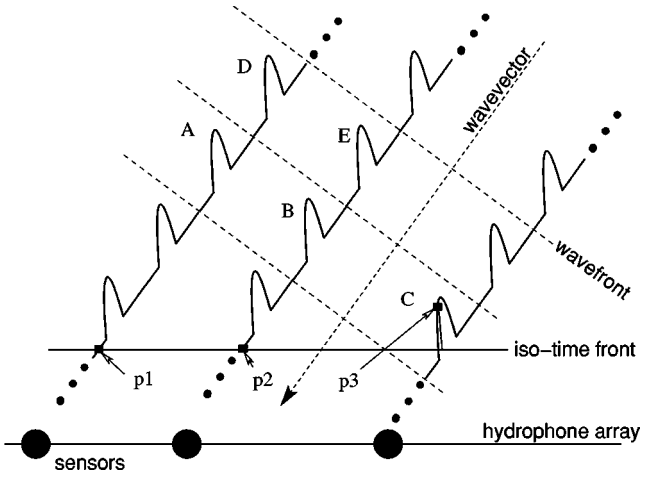


FIG. 16. Spatiotemporal diagram of *m*-sequence pulse signals relative to the sensors. In general, the acoustic field wave vector is incident on the line array at an arbitrary angle. At a single instant in time, the array measures nonidentical (p1, p2, and p3) parts of each pulse. True transverse measurements would use pulse pairs *A*, *B*. The Wong technique uses time-adjacent pulses, thus pair *A*, *E*.

taining only the terms $|\rho_r(\Delta r_{jk})|$, i.e., the magnitude of the spatial coherence. In matrix notation, Eq. (A9) is

$$\mathbf{G}=\mathcal{T}_{\times}(\mathbf{C}_{ss})=[\mathbf{D}^{-1/2}(\mathbf{C}_{ss}\otimes\mathbf{C}_{ss}^{*})\mathbf{D}^{-1/2}]^{1/2}, \tag{A11}$$

where

$$\mathbf{D}=\mathrm{diag}(\mathrm{diag}(\mathbf{C}_{ss}\otimes\mathbf{C}_{ss}^{*})),$$

and $\mathbf{X}\otimes\mathbf{Y}$ is the Hadamard product (element by element multiplication) between two matrices $\mathbf{X}$ and $\mathbf{Y}$. This new transformation eliminates the dependence of the matrix elements on the temporal decorrelation, leaving a matrix containing only spatial coherence terms.

(The matrix $\mathbf{G}$ calculated this way is not intrinsically Hermitian since $\mathbf{s}^{(a)}\neq\mathbf{s}^{(b)}$. However, any estimated coherence should be Hermitian since $|\Gamma(\mathbf{r}_j-\mathbf{r}_k)|=|\Gamma(\mathbf{r}_k-\mathbf{r}_j)|$. It therefore becomes appropriate to use instead

$$\mathbf{G}_{\mathrm{H}}=\frac{1}{2}(\mathbf{G}+\mathbf{G}^{T}). \tag{A12}$$

The matrix $\mathbf{G}_{\mathrm{H}}$ is used in subsequent computations.)

The technique does not magically eliminate the effect of the additive ambient noise: the sample data covariance matrix based on this technique, for example, will converge at a rate depending on statistical properties of both the signal and the noise. If the SNR is poor, the convergence will be dominated by the influence of the additive noise. However, the sample data covariance matrix will converge to the signal covariance matrix, and so this technique provides unbiased statistics.

[1] L. A. Chernov, *Wave Propagation in a Random Media* (McGraw-Hill, New York, 1960) (English translation, R. A. Silverman).
[2] V. I. Tatarskii, *Wave Propagation in a Turbulent Medium* (McGraw-Hill, New York, 1961) (English translation, R. A. Silverman).
[3] A. Ishimaru, *Wave Propagation and Scattering in Random Media* (IEEE Press, Piscataway, NJ, 2000).
[4] S. M. Flatté, R. Dashen, W. Munk, K. Watson, and F. Zachariasen, *Sound Transmission Through a Fluctuating Ocean* (Cambridge University Press, Cambridge, 1979).
[5] J. C. Steinberg and T. G. Birdsall, "Underwater sound propagation in the straits of Florida," J. Acoust. Soc. Am. **39**, 301–315 (1966).
[6] T. E. Ewart, "Acoustic fluctuations in the open ocean—a measurement using a fixed refracted path," J. Acoust. Soc. Am. **60**, 46–59 (1976).
[7] S. A. Reynolds, S. M. Flatté, R. Dashen, B. Buehler, and P. Maciejewski, "AFAR measurements of acoustic mutual coherence functions of time and frequency," J. Acoust. Soc. Am. **77**, 1723–1731 (1985).
[8] T. F. Duda, S. M. Flatté, J. A. Colosi, B. D. Cornuelle, J. A. Hildebrand, W. S. Hodgkiss, Jr., 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).
[9] J. A. Colosi, E. K. Scheer, S. M. Flatté, B. D. Cornuelle, M. A. Dzieciuch, W. H. Munk, P. F. Worcester, B. M. Howe, J. A. Mercer, R. C. Spindel, K. Metzger, T. G. Birdsall, and A. B. Baggeroer, "Comparisons of measured and predicted acoustic fluctuations for a 3250-km propagation experiment in the eastern North Pacific Ocean," J. Acoust. Soc. Am. **105**, 3202–3218 (1999).
[10] The NPAL Group (J. A. Colosi, B. D. Cornuelle, B. D. Dushaw, M. A. Dzieciuch, B. M. Howe, J. A. Mercer, R. C. Spindel, and P. F. Worcester), "The North Pacific Acoustic Laboratory (NPAL) Experiment," J. Acoust. Soc. Am. **109**, 2384 (2001).
[11] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (Princeton University Press, Princeton, NJ, 1968), Sec. 9.2, pp. 492–522.
[12] L. Brekhovskikh and Yu. Lysanov, *Fundamentals of Ocean Acoustics* (Springer-Verlag, Berlin, 1982), Chap. 6, pp. 109–138.
[13] M. L. Grabb, S. Wang, and T. G. Birdsall, "Deterministic three-dimensional analysis of long-range sound propagation through internal-wave fields," IEEE J. Ocean. Eng. **21**, 260–272 (1996).
[14] W. Munk, P. Worcester, and C. Wunsch, *Ocean Acoustic Tomography* (Cambridge University Press, Cambridge, 1995), Secs. 1.1 and 1.2, pp. 2–16.
[15] B. D. Dushaw, B. M. Howe, J. A. Mercer, and R. C. Spindel, "Multimegameter-Range Acoustic Data Obtained by Bottom-Mounted Hydrophone Arrays for Measurement of Ocean temperature," IEEE J. Ocean. Eng. **24**, 202–214 (1999).
[16] C. Garrett and W. Munk, "Space–time scales of internal waves," Geophys. Fluid Dyn. **3**, 225–264 (1972).
[17] C. Garrett and W. Munk, "Space–time scales of internal waves: a progress report," J. Geophys. Res. **80**, 291–297 (1975).
[18] C. Garrett and W. Munk, "Internal waves in the ocean," Annu. Rev. Fluid Mech. **11**, 339–369 (1979).
[19] "Computation of acoustic fluctuations from internal waves," (CAFI): http://www.es.ucsc.ed/~smf/cafi/cafi.html, 2000.
[20] S. M. Flatté and G. Rovner, "Calculations of internal-wave-induced fluctuations in ocean acoustic propagation," J. Acoust. Soc. Am. **108**, 526–534 (2000).
[21] E. Wolf, "A macroscopic theory of interference and diffraction of light from finite sources. II. Fields with a spectral range of arbitrary width," Proc. R. Soc. London, Ser. A **230**, 246–265 (1955).
[22] M. J. Beran and G. B. Parrent, Jr., *Theory of Partial Coherence* (Prentice–Hall, Englewood Cliffs, NJ, 1964), Sec. 1.1, pp. 2–4.
[23] G. C. Carter, "Tutorial overview of coherence and time delay estimation," in *Coherence and Time Delay Estimation*, edited by G. C. Carter (IEEE Press, Piscataway, NJ, 1993), pp. 1–28.
[24] N. L. Owsley, "Sonar array processing," in *Array Signal Processing*, edited by S. Haykin (Prentice-Hall, Englewood Cliffs, NJ, 1985), pp. 115–193.
[25] H. Cox, "Line array performance when the signal coherence is spatially dependent," J. Acoust. Soc. Am. **54**, 1743–1746 (1973).
[26] S. M. Flatté, R. Dashen, W. Munk, K. Watson, and F. Zachariasen, *op. cit.*, p. 61.
[27] B. M. Howe, S. G. Anderson, A. Baggeroer, J. A. Colosi, K. R. Hardy, D. Horwitt, F. W. Karig, S. Leach, J. A. Mercer, K. Metzger, Jr., L. O. Olsen, D. A. Peckham, D. A. Reddaway, R. R. Ryan, R. P. Stein, K. van der Heydt, J. D. Watson, S. L. Weslander, and P. F. Worcester, "Instrumentation for the Acoustic Thermometry of Ocean Climate (ATOC) prototype Pacific Network," in *OCEANS '95 Proceedings*, San Diego, CA, 9–12 October, 1995, pp. 1483–1500.
[28] T. G. Birdsall and K. Metzger, Jr., "*M*-sequence signal tutorial," Naval Oceanographic Office presentation, Communications and Signal Processing Laboratory, EECS Dept., University of Michigan, Ann Arbor, MI, April 1988.
[29] For example, L. Brekhovskikh and Yu. Lysanov, *op. cit.*, Chap. 10, pp. 208–216.

[30] B. Dushaw (private communication).
[31] National Imagery and Mapping Agency, *World Geodetic System 1984 (WGS 84)—Its Definition and Relationships with Local Geodetic Systems*, 3rd. ed., Washington, DC, National Imagery and Mapping Agency, 1997.
[32] W. M. Carey, ''The determination of signal coherence length based on signal coherence and gain measurements in deep and shallow water,'' J. Acoust. Soc. Am. **104**, 831–837 (1998).
[33] S. M. Flatté and R. B. Stoughton, ''Predictions of internal wave effects on ocean acoustics coherence, travel-time variance, and intensity moments for very long-range propagation,'' J. Acoust. Soc. Am. **84**, 1414–1428 (1988).
[34] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap* (Chapman & Hall, New York, 1993), Sec. 6.2, pp. 45–49. The bootstrap is a computational procedure for estimating the error of an estimator by reusing the sample data. Briefly, let an estimator $\hat{t}$ use all measurements x in a sample $X=\{x_1,x_2,\ldots,x_N\}$. ''New'' sample sets X_1, $X_2,\ldots$, are generated such that each sample set contains N elements selected randomly with replacement from the original set. Bootstrap replicates of $\hat{t}$ are then computed from each new sample set, and the bootstrap error in $\hat{t}$ is then the standard deviation of the bootstrap replicates.
[35] J. S. Bendat and A. Piersol, *Random Data Analysis and Measurement Procedures*, 2nd ed. (Wiley, New York, 1986).
[36] I. Dyer, ''Statistics of distant shipping noise,'' J. Acoust. Soc. Am. **53**, 564–570 (1973).
[37] S. M. Flatté, R. Dashen, W. Munk, K. Watson, and F. Zachariasen, Ref. *op. cit.*, Chap. 10, pp. 165–188.
[38] S. Levitus and T. Boyer, *World Ocean Atlas 1994 Vol. 4: Temperature*, NOAA Atlas NESDIC Vol. 4; U.S. Dept. of Commerce, Washington, DC. See also http://ingrid.ldeo.columbia.edu/SOURCES/.LEVITUS94
[39] V. A. Del Grosso, ''New equation for the speed of sound in natural waters (with comparisons to other equations),'' J. Acoust. Soc. Am. **56**, 1084–1091 (1974).
[40] S. Pond and G. L. Pickard, *Introductory Dynamical Oceanography*, 2nd ed. (Pergamon, Elmsford, NY, 1983), Sec. 5.13, p. 30.
[41] *UNESCO 10th Report of the Joint Panel on Oceanographic Tables and Standards*, UNESCO Technical Papers in Marine Science, 1981, No. 36, p. 24.
[42] B. D. Dushaw and J. C. Colosi, ''Ray tracing for ocean acoustic tomography,'' APL-UW Technical Memorandum APL-UW TM 3-98, December 1998.
[43] S. M. Flatté (private communication, December 2002).
[44] A. R. Thompson, J. M. Moran, and G. W. Swenson, Jr., *Interferometry and Synthesis in Radio Astronomy* (Wiley, New York, 1986), Appendix 3.1, pp. 73–77.
[45] A. Voronovich, V. Ostashev, and the NPAL Group (J. A. Colosi, B. D. Cornuelle, B. D. Dushaw, M. A. Dzieciuch, B. M. Howe, J. A. Mercer, W. H. Munk, R. C. Spindel, and P. F. Worcester), ''Horizontal refraction of acoustic signals retrieved from the North Pacific Acoustic Laboratory (NPAL) billboard array data,'' J. Acoust. Soc. Am. **117,** 1527–1537 (2005).
[46] M. A. Dzieciuch and P. F. Worcester, ''Horizontal refraction of ocean basin scale acoustic signals,'' submitted to J. Acoust. Soc. Am.
[47] K. M. Wong, Q. Wu, and P. Stoica, ''Generalized correlation decomposition applied to array processing in unknown noise environments,'' pp. 219–323, in *Advances in Spectrum Analysis and Array Processing*, edited by S. Haykin (Prentice–Hall, Englewood Cliffs, NJ, 1995).

Horizontal refraction of acoustic signals retrieved from the North Pacific Acoustic Laboratory billboard array data

A. G. Voronovich[a] and V. E. Ostashev
NOAA/Environmental Technology Laboratory, 325 Broadway, Boulder, Colorado 80305

The NPAL Group (J. A. Colosi, B. D. Cornuelle, B. D. Dushaw, M. A. Dzieciuch, B. M. Howe, J. A. Mercer, W. H. Munk, R. C. Spindel, and P. F. Worcester)[b]

(Received 15 March 2004; revised 2 July 2004; accepted 6 July 2004)

In 1998–1999, a comprehensive low-frequency long-range sound propagation experiment was carried out by the North Pacific Acoustic Laboratory (NPAL). In this paper, the data recorded during the experiment by a billboard acoustic array were used to compute the horizontal refraction of the arriving acoustic signals using both ray- and mode-based approaches. The results obtained by these two approaches are consistent. The acoustic signals exhibited weak (if any) regular horizontal refraction throughut most of the experiment. However, it increased up to 0.4 deg (the sound rays were bent towards the south) at the beginning and the end of the experiment. These increases occurred during midspring to midsummer time and seemed to reflect seasonal trends in the horizontal gradients of the sound speed. The measured standard deviation of the horizontal refraction angles was about 0.37 deg, which is close to an estimate of this standard deviation calculated using 3D modal theory of low-frequency sound propagation through internal gravity waves. [DOI: 10.1121/1.1854435]

PACS numbers: 43.30.Es, 43.30.Qd, 43.30.Re [AIT] Pages: 1527–1537

I. INTRODUCTION

Sound waves propagating over long ranges in the ocean exhibit horizontal refraction in the presence of horizontal gradients of the sound-speed transversal to the direction of propagation. Relatively long-period (weeks-to-months) variations of the horizontal refraction angle (HRA) can be caused by mesoscale eddies. Variations on even longer time scales can be due to seasonal trends in the sound-speed gradients. Short-period variations are caused by sound scattering by internal waves.

Thus, the dependence of the HRA on time contains important information about ocean inhomogeneities. If long-period variations of HRA could be measured, this would give an opportunity for acoustic monitoring of large-scale inhomogeneities in the ocean.[1–3] Measurements of the statistical characteristics of short-period variations may be used for remote sensing of internal waves.

So far, the horizontal refraction of sound waves propagating over long ranges in the ocean has been studied mainly theoretically.[4–9] For example, in Ref. 4, the HRA of a sound wave passing through a mesoscale eddy was estimated as 1 deg, and in Ref. 5, the HRA due to a model ocean eddy was calculated as 0.5 deg. These are small values of the HRA so a special instrumentation is required to measure them. The HRA and horizontal coherence of a sound wave propagating in shallow water were studied numerically elsewhere.[10]

In this paper, the dependence of the HRA on time is studied by processing the data recorded by a billboard acoustic array of the North Pacific Acoustic Laboratory (NPAL). The experiment was carried out in 1998–1999 and is briefly described in Sec. II. Two approaches for signal processing were used. The first approach, which is presented in Sec. III, is based on the assumption that a sound field impinging on the array is a superposition of plane waves arriving from a particular horizontal direction. This approach is called the ray-based approach. The second approach employs a modal representation of the sound field (Sec. V). Using both approaches, we obtained the dependence of the HRA on time and studied its short-period and long-period variations. The results obtained are summarized in the Conclusions.

Preliminary results of signal processing the NPAL data using these two approaches have been reported elsewhere.[11,12]

II. OUTLINE OF THE EXPERIMENT

In 1998, a billboard acoustic array of NPAL was installed near the coast of California at 36°17.1′N, 122°28.78′W. For about a year, the array recorded broadband acoustic signals transmitted by a source located near Kauai, Hawaii at a depth of 807 m with coordinates 22°20.94′N, 159°34.18′W. The distance between the source and the array was $R=3.9\times10^3$ km. The transmitted signal consisted of a 75-Hz carrier which was phase modulated according to a pseudorandom 1023-digit maximal-length M sequence. The bandwidth of the signal was about 30 Hz. The duration of one sequence was 27.28 s. The sequence was successively transmitted 40 times so that the total duration of one session

[a]Electronic mail: alexander.voronovich@noaa.gov
[b]J. A. Colosi is at Woods Hole Oceanographic Institution, Woods Hole, MA 02543. B. D. Cornuelle, M. A. Dzieciuch, W. H. Munk, and P. F. Worcester are at Scripps Institution of Oceanography, University of California at San Diego, La Jolla, CA 92093. B. D. Dushaw, B. M. Howe, J. A. Mercer, and R. C. Spindel are at the Applied Physics Laboratory, University of Washington, Seattle, WA 98105.

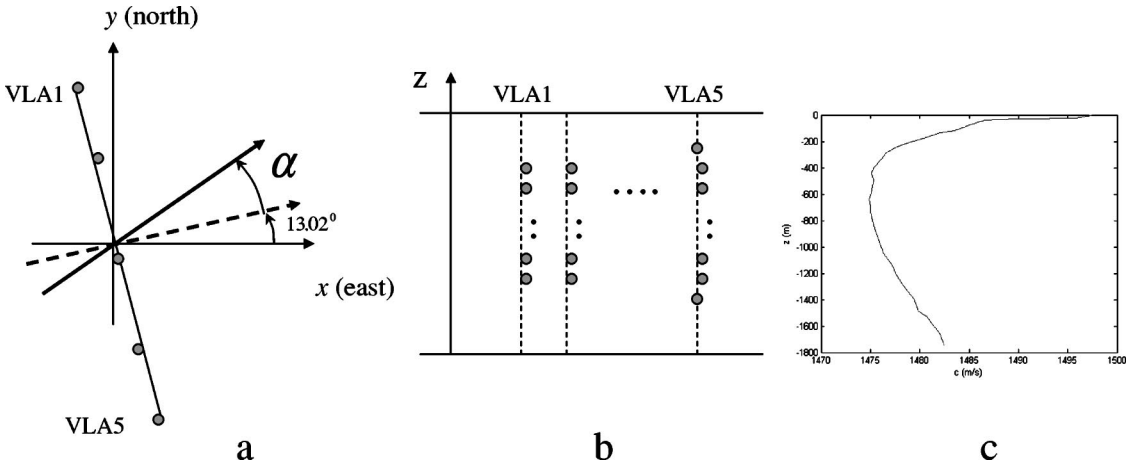


FIG. 1. (a) View of the NPAL billboard array in the horizontal plane. The dashed line corresponds to the direction of arrival in the absence of horizontal refraction, and the solid line corresponds to the direction of actual arrival. The angle between these directions is the HRA α. (b) View of the NPAL billboard array in the vertical plane. (c) Summer sound-speed profile at the array location.

was 1091.2 s. These sessions were repeated according to a schedule with intervals varying from 4 h to 4 days.

The billboard array consisted of five vertical line arrays (VLAs) which were positioned along the line approximately perpendicular to the acoustic propagation path as shown in Fig. 1(a). In the figure, the x axis is directed east, and the y axis is directed north. The horizontal distances between consecutive VLAs (starting from the most northern VLA1) were approximately 0.5, 1.2, 0.7, and 1.1 km. The angle $\alpha_0 = 13.02$ deg indicates a nominal direction of sound propagation in a homogeneous ocean along the unrefracted geodesic. The HRA α indicates a deviation of sound propagation from this direction due to horizontal refraction. Positive values of α correspond to bending of acoustic rays towards south.

The first four VLAs consisted of 20 hydrophones separated vertically by approximately 35 m; see Fig. 1(b). The overall length of each array was close to 665 m. Hydrophones of the first four VLAs were located at approximately the same depth z_n, where the index $n = 1,2,...,20$. The fifth (southernmost) VLA5 was twice the size of the other arrays and consisted of 40 hydrophones. The ocean depth at the location of the billboard array was 1800 m. The motion of the arrays was monitored by a set of high-frequency transponders which ensured the measurement of hydrophone positions with sufficiently high accuracy. A special timekeeping system allowed one to obtain coherent record of the acoustic signals at different VLAs. The summer sound-speed profile $c(z)$ near the array is shown in Fig. 1(c). A detailed description of the transmitted signal and the billboard acoustic array can be found in Refs. 13, 14.

Along with digital records of raw acoustic signals, preprocessed data were also available for further analysis. The preprocessing was accomplished for each hydrophone independently and included complex demodulation and removal of M sequence. This removal was done by division of the signal spectrum by a spectrum of M sequence that made a transmitted signal effectively a δ function within a given frequency band. (Such preprocessed data are used in Sec.

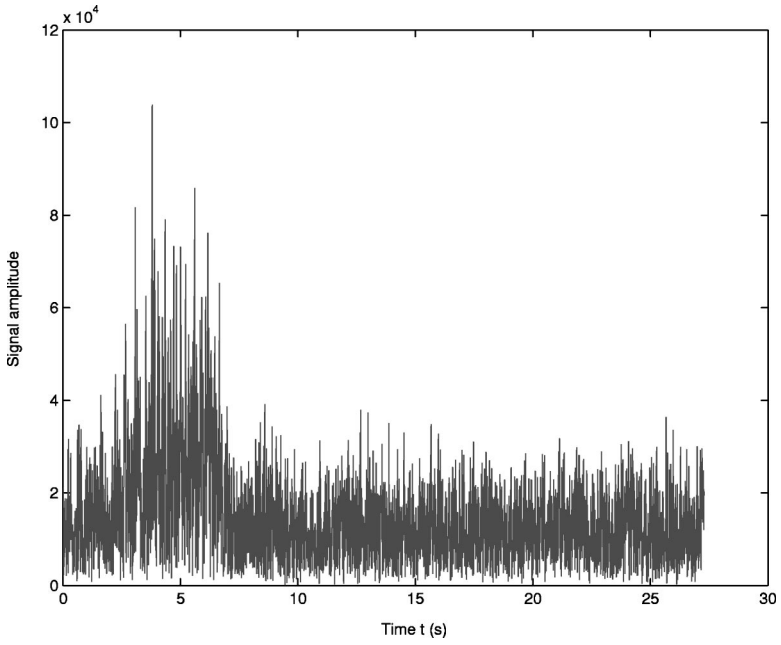


FIG. 2. Amplitude of the sound signal (in arbitrary units) as a function of time at the hydrophone of VLA3 with the index $n = 3$ for $t = 277.1206$ day of the experiment.

III C.) Then, 40 successive transmissions were coherently averaged. Such coherently averaged signals were used in ray- and modal-based processing (Secs. III A and B, and Sec. V).

An example of the coherently averaged signal recorded by the hydrophone of VLA3 with the index $n=3$ for $t=277.1206$ day of the experiment is shown in Fig. 2. (Note that $t=277$ day of the experiment corresponds to 2 October 1998). The duration of the record is 27.28 s. The signal starts at about 1 s from the beginning of the record and ends at about 7–8 s, with the rest of the record being dominated by noise. The ratio of the signal intensity at its maximum to the noise level is about 18 dB.

III. RAY-BASED PROCESSING

A. Approach

The ray-based processing employed cross correlation of signals recorded by hydrophones located at approximately the same depth and belonging to different VLAs. This approach is the most direct, and it seems to be robust and does not include any implicit or explicit assumptions which are used in more sophisticated processing approaches (e.g., in the mode-based approach considered in Sec. V).

Before performing the cross correlation, a frequency filter with a bandwidth of 30 Hz was applied to the signals and they were resampled at a higher rate. (Resampling allowed us to more accurately determine position of a maximum of the cross-correlation function of two signals.) Furthermore, the coherently averaged signals were truncated to the first 11.67 s since the rest of the signals (with a total duration of 27.28 s) consisted of noise. This ensured a better accuracy in measuring the HRA α. (Note that we also tried to truncate the coherently averaged signals to first arrivals or finale only. Such truncations resulted in almost the same values of the HRA as in the 11.67-s truncation but larger variances of the HRA, and, therefore, were not used in the ray-based approach.)

Figure 3(a) shows a cross-correlation function of signals recorded by the hydrophones with the index $n=20$ belonging to VLA3 and VLA4 for $t=277.1206$ day. The correlation between the signals is high and, hence, HRA can be reliably determined. On the other hand, Fig. 3(b) shows a poor correlation between signals recorded by the hydrophones with the index $n=20$ belonging to VLA2 and VLA4 at the same time of the experiment; the HRA cannot be determined. [Note the difference in the vertical scales in Figs. 3(a) and (b).] Figures 3(a) and (b) are typical examples of high and poor cross correlation between signals recorded by various pairs of hydrophones.

For a given transmission, we considered different pairs of VLAs (e.g., VLA3 and VLA4). We cross correlated the signals recorded by the hydrophones of two VLAs located at the same level *n*. The locations of the maxima of the cross-correlation functions allowed us to obtain the time differences Δt_n of signal arrival at 20 pairs of hydrophones. Assuming that plane waves were impinging on the billboard array from a particular horizontal direction, HRAs α_n were expressed in terms of the time differences Δt_n

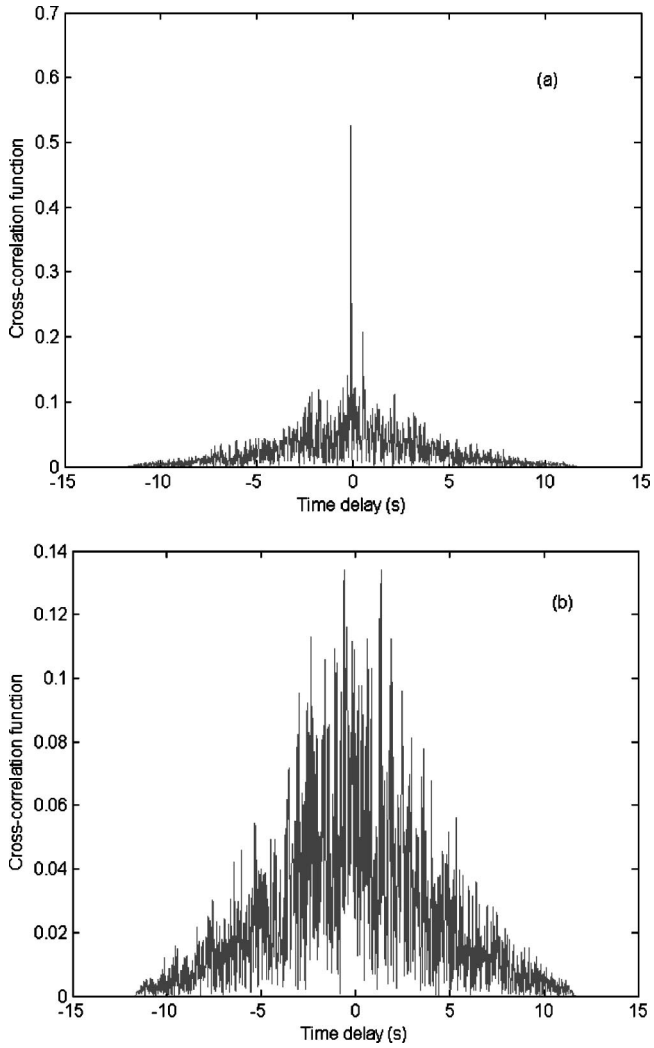


FIG. 3. Cross-correlation function of the acoustic signals recorded by two hydrophones with the index $n=20$ and belonging to (a) VLA3 and VLA4; (b) VLA2 and VLA4. The figure corresponds to $t=277.1206$ day.

$$\alpha_n = \arcsin\left(\frac{(\Delta t_n+\tau)c_n}{L_n}\right) + \arccos\left(\frac{y_n-y_n'}{L_n}\right) - \alpha_0. \qquad (1)$$

Here, $L_n=\sqrt{(x_n-x_n')^2+(y_n-y_n')^2}$ is the horizontal distance between the hydrophones of two VLAs; (x_n,y_n) and (x_n',y_n') are the horizontal coordinates of these hydrophones; $c_n=c(z_n)$ is the sound speed; and τ is a time correction to internal clocks of two VLAs. Accurate positions of all hydrophones and times corrections were given in the NPAL database.

When calculating α_n, we assumed that the hydrophones of two VLAs were located at the same depth. This assumption is valid since, in most cases, the difference z_n-z_n' in the vertical coordinates of two hydrophones was only a few meters. If this difference were taken into account in Eq. (1), it would result in the change of α_n which is much smaller than the standard deviation σ of the HRA, considered below. Also note that uncertainties in clock correction $d\tau$ and horizontal coordinates of the hydrophones, dx and dy, resulted in errors $d\alpha_\tau$, $d\alpha_x$, and $d\alpha_y$ in determining the HRA α_n. The value of $d\tau$ varied during the experiment; its average value was about 1.3 ms for VLA3 and VLA4. The value of

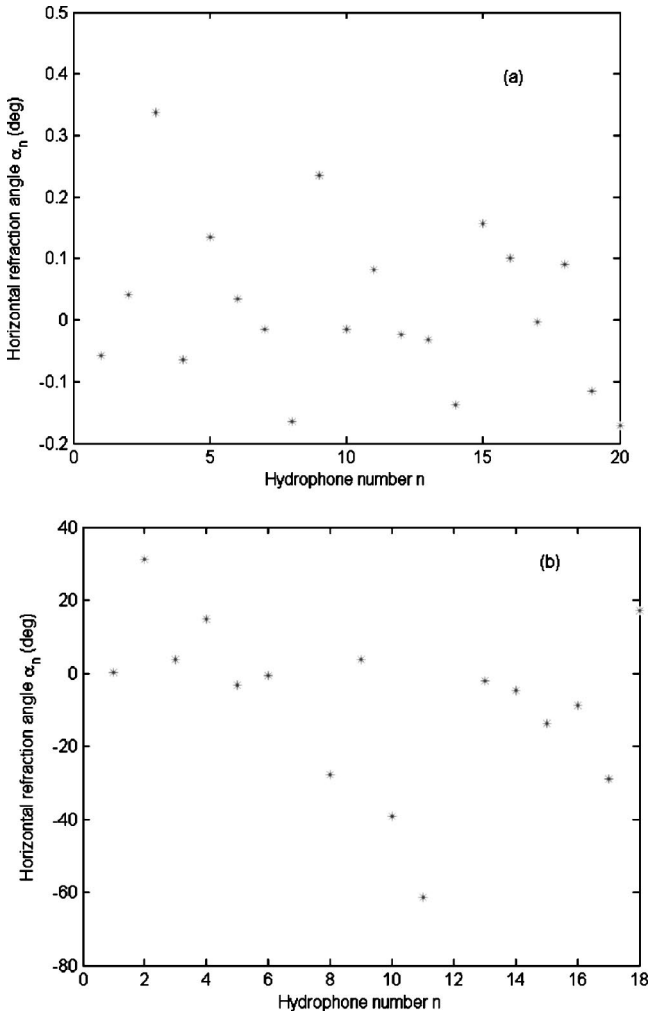


FIG. 4. Horizontal refraction angle α_n versus hydrophone number n for $t=277.1206$ day. (a) Data obtained with the use of VLA3 and VLA4. (b) Data obtained with the use of VLA2 and VLA4.

dx was about 0.6 m. Since the billboard array was nearly perpendicular to the propagation path, the error $d\alpha_y$ can be ignored while the errors $d\alpha_\tau$ and $d\alpha_x$ can be estimated as follows: $d\alpha_\tau \sim d\tau c/L$ and $d\alpha_x \sim dx/L$. For $L=700$ m (the distance between VLA3 and VLA4), we have $d\alpha_\tau \sim 0.16$ deg and $d\alpha_x \sim 0.05$ deg.

Figure 4(a) shows the dependence of α_n on n for $t=277.1206$ day computed for VLA3 and VLA4. All values of α_n are close to each other and range from -0.2 to 0.35 deg. This indicates high coherence between acoustic signals recorded by VLA3 and VLA4. The mean value of the HRA α and its standard deviation σ were computed using the values of α_n satisfying the following constraint:

$$|\alpha_n| \leqslant |\Delta\alpha|, \tag{2}$$

where $\Delta\alpha=2$ deg. This constraint was used to eliminate unrealistically large values of α_n which occurred due to the occasional poor correlation of signals recorded by some pairs of hydrophones. The value $\Delta\alpha=2$ deg is chosen such that it exceeds significantly the resulting standard deviation σ. For the data shown in Fig. 4(a), $\alpha=0.02$ deg, $\sigma=0.13$ deg, and $N=20$, where N is the number of hydrophone pairs satisfying the constraint Eq. (2). Note that α and σ calculated in such a way correspond to the averaged values of the refraction angle and its variance over the depth and 40 successive sequences (i.e., over 1091.2 s). Also note that in Fig. 4(a) and similar figures for other days of the experiments, we did not see any systematic dependence of α_n on n.

The HRAs α_n were similarly computed for other pairs of VLAs, e.g., VLA1 and VLA2, VLA1 and VLA3, etc. For VLA5 we used only 20 hydrophones located at depths close to those for hydrophones of VLA1 through VLA4. For adjacent pairs of VLAs (i.e., for VLA1 and VLA2, VLA2 and VLA3, VLA3 and VLA4, VLA4 and VLA5) the dependence of α_n on n was usually (but not always) similar to that shown in Fig. 4(a).

On the other hand, for nonadjacent pairs of VLAs (e.g., VLA1 and VLA3, VLA1 and VLA4), the dependence of α_n of n almost always had a different character. For example, Fig. 4(b) shows the dependence of α_n of n obtained with the use of VLA2 and VLA4 for the same day $t=277.1206$ as that for Fig. 4(a). One can see that the values of α_n randomly vary in a much broader interval from -60 to 30 deg. [Note that the scale of vertical axes in Figs. 4(a) and (b) differs by nearly 2 orders of magnitudes. Therefore, if the data from Fig. 4(a) were plotted in Fig. 4(b), they would lie nearly perfectly along a horizontal line.] At levels $n=7$ and $n=12$ the cross-correlation technique failed to produce physically meaningful real values of α_n. Furthermore, it follows from Fig. 4(b) that the number N of hydrophones satisfying the condition Eq. (2) is only $N=2$. Apparently, this is not enough to estimate a mean refraction angle α.

For a given transmission, values of α, σ, and N were calculated for all available pairs of VLAs. The cases corresponding to $N \leqslant 14$ and $\sigma \geqslant 0.6$ deg were considered as failures and were discarded. Finally, the pair of VLAs with a minimal value of σ was chosen. (In most cases, these were VLA3 and VLA4.) This value of σ was considered as the standard deviation $\sigma(t)$ of the HRA for transmission time t, and the corresponding HRA α was considered as $\alpha(t)$. For example, for $t=277.1206$ day, the minimal value of σ was for VLA3 and VLA4; therefore, $\alpha(277.1206)=0.02$ deg and $\sigma(277.1206)=0.13$ deg.

Using the NPAL data, we also calculated the cross-correlation coefficient between signals at the hydrophones of two VLAs with the same index n. The obtained cross-correlation coefficients were then averaged over n (i.e., over depth). Figures 5(a) and (b) show the dependence of the averaged cross-correlation coefficients on the distance between VLAs. Figure 5(a) corresponds to $t=257.2873$ day, while Fig. 5(b) corresponds to $t=285.2873$ day. The figures are typical for other days of the experiment. The dashed lines in Figs. 5(a) and (b) are theoretical curves $\exp(-l^2/r_c^2)$ which were used to fit experimental data. Here, l is the distance between the VLAs and r_c is the horizontal coherence radius. The value of r_c was an adjustable parameter for the best fit of the theoretical curves to the experimental data. The intersection of the dashed lines with the horizontal dotted lines gives the value of the coherence radius r_c. It follows from Figs. 5(a) and (b) that r_c varies in the range of 500–1000 m. Note that r_c was considerably less than the distance between nonadjacent pairs of VLAs. This explains why the nonadja-

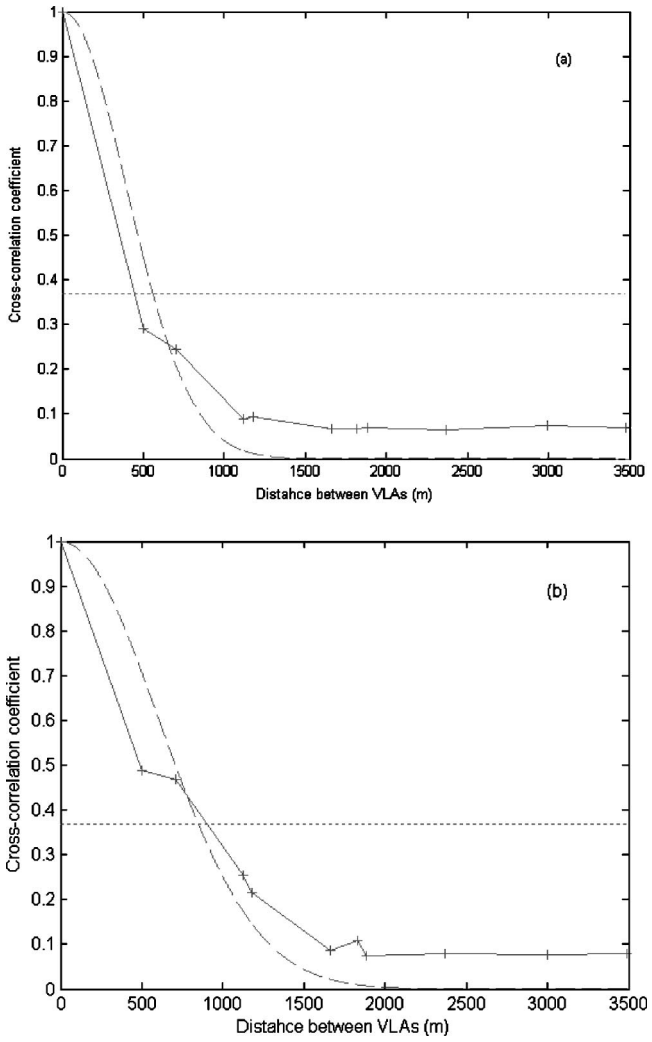


FIG. 5. Cross-correlation coefficient of acoustic signals recorded by different VLAs versus the distance between VLAs. (a) $t=257.2873$ day; (b) $t=285.2873$ day.

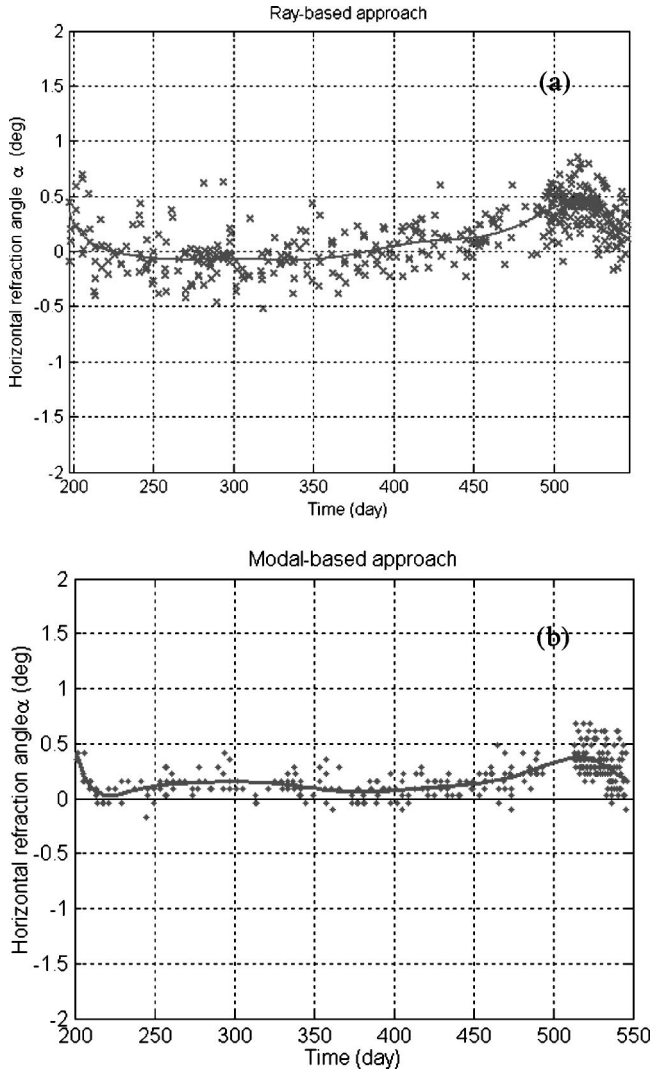


FIG. 6. Dependence of the horizontal refraction angle α on time calculated by: (a) the ray-based approach, (b) the mode-based approach. The solid lines are the least-square data fit to a polynomial of the order 10.

cent pairs could not usually be used to measure the HRA, e.g., see Fig. 4(b).

B. The time dependence of the HRA

The algorithm described in the previous subsection allowed us to obtain the dependence of the HRA α on time for the period of the NPAL experiment from $t=197$ to $t=546$ day. This dependence is shown in Fig. 6(a), where crosses correspond to computed values of α. In the figure, each HRA α was computed together with its standard deviation σ, which varied in the range from 0.13 to 0.6 deg. The mean value of all standard deviations of α shown in Fig. 6(a) is $\overline{\sigma}=0.37$ deg. Note that 67% of the data in Fig. 6(a) was obtained by processing signals of VLA3 and VLA4, and 23% was obtained with the use of VLA1 and VLA2. The horizontal distances between these pairs (700 and 500 m, respectively) were of the order of the coherence radius r_c. The distances between other adjacent pairs of VLAs were greater than 1 km.

It is seen from Fig. 6(a) that short-term (hours to days) variations of the HRA α are of the order 0.3–0.5 deg. These variations are probably due to sound scattering by internal gravity waves (see Sec. IV below).

Long-period (weeks to months) variations of the HRA α are not seen in Fig. 6(a). If they do exist, they are masked by short-term variations.

The values of the HRA α averaged over 1 day are shown in Fig. 7. These data show less variance than those in Fig. 6(a). The solid lines in Figs. 6(a) and 7 are the least-square data fit by a polynomial of the order 10. These solid lines are close to each other and we interpret them as seasonal variations of α. In the midspring to midsummer of 1998 and 1999 (197–210 and 470–545 days), the value of α is positive and reaches 0.4 deg. Note that this value is close to the value of $\overline{\sigma}=0.37$ deg. However, it seems that a systematic increase of the HRA α in Figs. 6(a) and 7 does exist. Also note that, for 197–210 and 470–545 days and VLA3, the averaged uncertainties $d\tau$ were 0.30 and 0.62 ms, respectively, and they were less than the averaged uncertainty $d\tau=1.3$ ms during the whole period of the experiment. (For VLA4, these uncertainties were of the same order of magnitude.) The errors $d\alpha_\tau$ corresponding to these uncertainties are given by 0.04

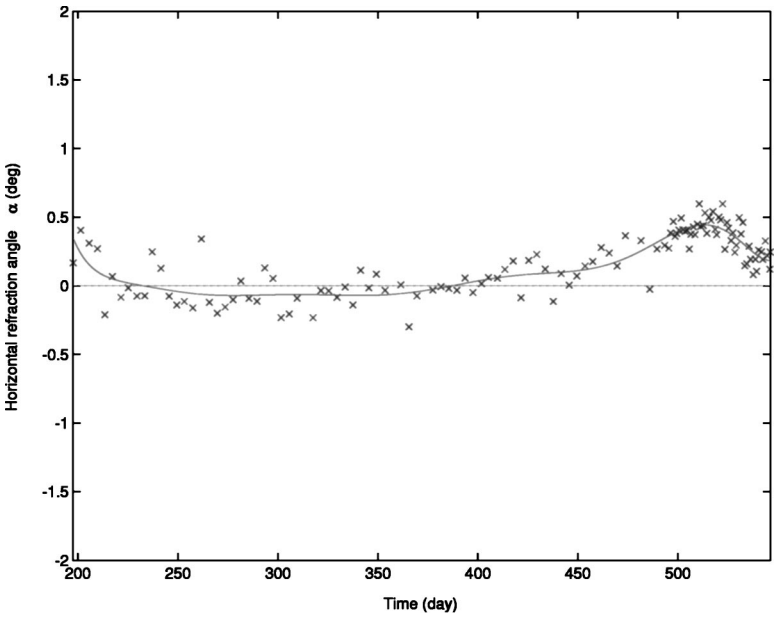


FIG. 7. Same as in Fig. 6(a) but for the horizontal refraction angles α averaged over 1 day.

and 0.08 deg. These errors are small enough to have a significant effect on a systematic increase of the HRA in 197–210 and 470–545 days.

The average horizontal gradient of the sound speed across the sound propagation path can be estimated as[1]

$$\frac{1}{c}\frac{dc}{dy}=\frac{2\alpha}{R}, \tag{3}$$

where R is the path length. A maximum deviation from zero of the solid line in Figs. 6(a) and 7 is about 0.4 deg. Substituting the value $\alpha=0.4$ deg into Eq. (3), we estimate the horizontal sound-speed gradient as

$$\frac{1}{c}\frac{dc}{dy}\sim 3.6\times 10^{-6}\ \ \mathrm{km}^{-1}. \tag{4}$$

This gradient is of the order of a typical horizontal gradient of large-scale variations of sound-speed fields in the ocean.[15] We also tried to compare the value of this gradient with *in situ* measurements. Unfortunately, according to Ref. 16, no measurements close to the sound propagation path were done during the NPAL experiment.

C. Signals from passing ships

The ray-based approach described above was also employed for signal processing of NPAL data that were not coherently averaged over 40 identical transmissions. Nonaveraged signals had smaller signal-to-noise ratio than the averaged ones. As a result, in most cases the dependence of α_n on n had a random character similar to that shown in Fig. 4(b) so that the HRA could not be determined. Occasionally, however, the processing revealed sufficiently strong, stable signals that were coming from directions significantly different from the nominal direction towards the source. We interpreted such signals as being due to passing ships. Figure 8 shows the angular position of a ship as a function of time for

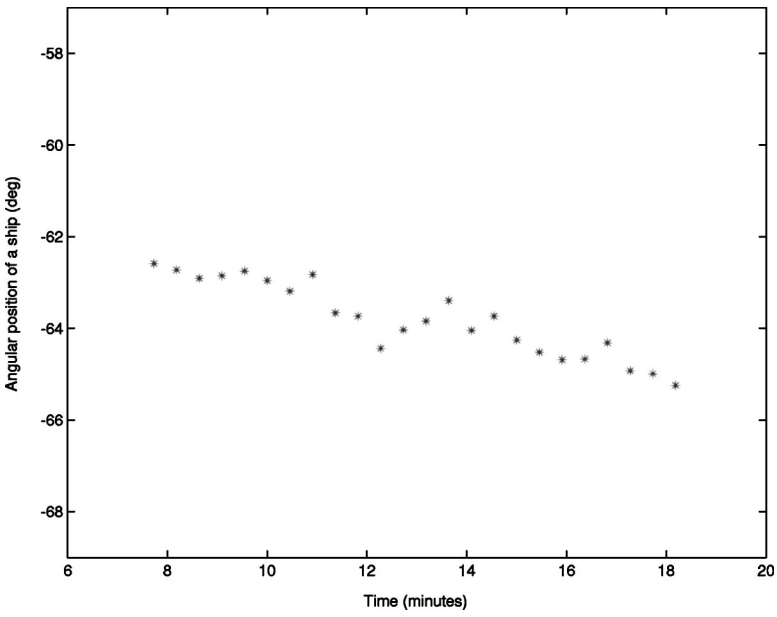


FIG. 8. Dependence of the angular position of a ship on time obtained for $t=500.7873$ day.

$t=500.7873$ day, obtained with the use of VLA1 and VLA4. According to Fig. 8, the angular speed of a ship is equal to 0.245 deg/min. If the ship were 50 nm from the NPAL array, its transverse speed would be 13 kn. Note that, since a ship emits a broadband signal, no aliasing was present when determining its angular position. Furthermore, the horizontal coherence of signals from ships was larger than that of the NPAL signals since the ships were much closer to the VLAs. This allowed us to use VLA1 and VLA4 separated by 2.4 km for calculation of the cross correlation of signals from passing ships.

IV. STANDARD DEVIATION OF THE REFRACTION ANGLE

The standard deviation $\overline{\sigma}$ of the HRA calculated in Sec. III B results from sound scattering by internal waves, the uncertainty dx, and the effect of ambient noise on signal processing. The uncertainty $d\tau$ does not contribute to $\overline{\sigma}$ since it was the same for every pair of hydrophones which was used to calculate the HRA and its standard deviation. The uncertainties dx were measured for every hydrophone. Since about 20 hydrophone were used when calculating the HRA, the contribution of $d\alpha_x$ to $\overline{\sigma}$ can be estimated as $d\alpha_x/\sqrt{20}\sim 0.01$ deg and can be safely ignored.

The signal (at its maximum) to noise ratio was relatively large (see Fig. 2) and the correlation between the signals at adjacent VLAs was usually high [see Fig. 3(a)]. Therefore, one could expect that the effect of noise on $\overline{\sigma}$ is small. Note that this effect is difficult to estimate quantitatively since the ray-based approach is nonlinear. At any rate, we can conclude that the calculated value of $\overline{\sigma}$ is an upper limit of the standard deviation of the HRA due to sound scattering by internal waves.

In Ref. 17, another approach was used to calculate the HRA using the NPAL data. That approach employed vertical beamforming and a turning point filter[18] and also resulted in the standard deviation of the HRA of about 0.4 deg.[17]

We also obtained two theoretical estimates of the standard deviation $\overline{\sigma}$ of the HRA due to sound scattering by internal waves. The first estimate uses the relationship between the coherence radius r_c and the standard deviation $\overline{\sigma}$ of angles at which the wavefront is impinging on the array

$$\overline{\sigma}\sim\frac{1}{k_0 r_c}, \tag{5}$$

where $k_0=\omega/c_0$ is a reference wave number and ω is the frequency. [It can be shown that, for a plane wave propagating through random inhomogeneities with a Gaussian spectrum, Eq. (5) takes the form $\overline{\sigma}=\sqrt{2}/(k_0 r_c)$.] For $f=75$ Hz and r_c in a range of 500–1000 m, it follows from Eq. (5) that $\overline{\sigma}$ is in a range of 0.26–0.51 deg. These values are close to $\overline{\sigma}=0.37$ deg which was obtained in the ray-based approach.

The second estimate employes a 3D, modal theory of low-frequency sound propagation through internal gravity waves developed in Ref. 19. Using this theory, it can be shown that the energy distribution with respect to horizontal angles obeys a diffusion equation with the following diffusion coefficient:

$$D=k_0^2 l_*. \tag{6}$$

Here, l_* is a parameter with the dimension of length

$$l_*=\frac{\pi}{2}\sum_{n,m,j}\int k^2 P_j(\xi_n-\xi_m,k)[N_{nm}^j(k)]^2 dk. \tag{7}$$

In this equation, $P_j(q_x,q_y)$ is the internal wave spectrum, q_x and q_y are components of the wave vector, j is the internal wave mode index, n and m are the acoustic mode indices, ξ_n is the nth acoustic mode wave number, and the matrix N_{nm}^j describes the interaction between the acoustic and internal wave modes

$$N_{nm}^j(k)=\int u_n(z)u_m(z)\Phi_j(\sqrt{(\xi_n-\xi_m)^2+k^2},z)dz. \tag{8}$$

Here, $\Phi_j(q,z)$ is a profile of the jth mode of the internal wave with the horizontal wave number q, and u_n, u_m are the acoustic mode profiles. Numerical estimates made for the Garrett–Munk spectrum of internal waves and the canonical Munk sound-speed profile results in $l_*=5.4\cdot 10^{-11}$ m. The standard deviation $\overline{\sigma}$ of the angles at which the wavefront is impinging on the array is related to the diffusion coefficient D by the following formula:

$$\overline{\sigma}=\sqrt{DR}=k_0\sqrt{l_* R}. \tag{9}$$

Using the value of l_* given above, we have $\overline{\sigma}=0.26$ deg. This value of $\overline{\sigma}$ is of the same order of magnitude as the value $\overline{\sigma}=0.37$ deg obtained in the ray-based approach.

All these support the conclusion that the standard deviation $\overline{\sigma}=0.37$ deg obtained in the ray-based approach is mainly due to scattering of the acoustic signals by internal waves.

V. MODAL PROCESSING

To independently check that a seasonal trend in the HRA shown in Figs. 6(a) and 7 does exist, we used a different approach for signal processing of the NPAL data which is based on a modal representation of a sound signal impinging on the billboard array. A modal approach works particularly well for low-frequency sound fields. The results obtained by this approach are presented below.

A. Approach

Modal analysis of acoustic signals measured by a single VLA located in the deep ocean was considered in Refs. 20–24. However, the approaches developed in these references cannot be directly used in coherent signal processing of acoustic signals recorded by a billboard array. Based on the ideas presented in these references, we developed a modal approach for coherent processing of acoustic signals recorded by five VLAs of the NPAL billboard array.

Let (x_b,y_b,z_b) be the coordinates of hydrophones of five VLAs. Hereinafter, the subscript $b=1,2,...,120$ indicates a particular hydrophone. In the modal approach, the complex sound pressure at a particular hydrophone and at a particular frequency ω_i is expressed in the following form:

$$p_b(\omega_i)=\sum_{m=1}^{M(\omega_i)} u_m(z_b,\omega_i)\exp(-ik_{x,m}(\omega_i)x_b-ik_y y_b)$$
$$\times a_m(\omega_i). \tag{10}$$

Here, the subscript m indicates the mode number, $M(\omega_i)$ is the total number of modes, which was equal to 10 in the numerical implementation of the modal approach, $u_m(z,\omega)$ is the acoustic mode profile, $a_m(\omega_i)$ is the mode amplitude, and $k_{x,m}$ and k_y are the components of the sound wave vector in the direction of the x- and y axes, respectively. Note that, in this section, the x axis is chosen in the direction of the nominal sound propagation. In numerical implementation, 10 adjacent frequencies were included into coherent processing so that in Eq. (10) $i=1,2,...,10$. A frequency increment Δf_0 between adjacent frequencies ω_i and ω_{i+1} was determined by the signal duration: $\Delta f_0=1/(27.28\,\mathrm{s})=0.0367\,\mathrm{Hz}$. Thus, the modal processing was coherent within a narrow frequency bin $10\Delta f_0=0.367\,\mathrm{Hz}$. The total number of frequency bins was 80, covering a frequency range of $75\pm14.66\,\mathrm{Hz}$.

Note that in Eq. (10) there is no summation over k_y; i.e., we assume that there is only one dominant direction of signal arrival in the horizontal plane. (The attempts to resolve different arrival angles in the horizontal plane were not successful.) The HRA α is related to k_y by the formula: $k_y=k_0\sin\alpha$. Furthermore, in Eq. (10) $k_{x,m}=\sqrt{\xi_m^2-k_y^2}$, where $\xi_m(\omega)$ is the horizontal wave number of the mth mode. Neglecting small terms of order α^2, we have $k_{x,m}=\xi_m$. Finally, in Eq. (10), the acoustic mode profiles $u_m(z,\omega)$ and the horizontal wave numbers $\xi_m(\omega)$ were numerically calculated as a solution of the standard differential equation for acoustic modes $d^2u_m/dz^2+(\omega^2/c^2-\xi_m^2)u_m=0$ with appropriate boundary conditions.

For a given value of the HRA α (say $\alpha=0.1$ deg) and a chosen frequency bin, Eq. (10) with $b=1,2,...,120$ and $i=1,2,...,10$ is a set of equations for determining the mode amplitudes $a_m(\omega_i)$ which are convenient to represent as a vector **a** with a dimension $10\times M$. In matrix notations, this set can be written as

$$Ta=p. \tag{11}$$

For a given value of α and chosen frequency bin, the matrix T was computed numerically using the positions of hydrophones and the values of $u_m(z_b,\omega_i)$ and $\xi_m(\omega_i)$.

In the presence of ambient noise, a least-square solution (pseudoinversion) of Eq. (11) is well known

$$a=(T^+T)^{-1}T^+p. \tag{12}$$

Here, the superscript "+" indicates a Hermitian conjugation. The matrix T^+T is a non-negative square Hermitian matrix, and in principle, Eq. (12) can be solved directly. The problem arises when the matrix T^+T has very small eigenvalues, which was the case in the present study. To regularize the solution, we used the following constraint:

$$\|a\|^2=\sum_{m,i}|a_m(\omega_i)|^2=a_0^2, \tag{13}$$

where a_0 is a positive constant. The physical meaning of this constraint is that sound scattering in the ocean mainly results in a redistribution of energy of an acoustic signal between different modes, with the total energy being constant. A concrete value of the constant a_0 was determined by trial and error; the final results were not sensitive to a particular choice of this parameter within rather broad limits.

Using a Lagrange multiplier r, the solution of Eq. (11) can be reduced to the minimization of the following cost function:

$$g=\|Ta-p\|^2+r(\|a\|^2-a_0^2). \tag{14}$$

Differentiating g with respect to a, we obtain the following equation:

$$(T^+T+rI)a=T^+p, \tag{15}$$

where I is the unity matrix. Thus, we obtained a "diagonal loading" regularization. However, the value of the regularization parameter r is not prescribed and has yet to be determined using Eq. (13).

Let us express the matrix T^+T in a diagonal form

$$T^+T=UdU^+, \tag{16}$$

where $d=\mathrm{diag}(d_n)$ is a diagonal matrix consisting of real, non-negative eigenvalues d_n of the matrix T^+T and whose unitary matrix of normalized eigenvectors is U. Substituting Eq. (16) into Eq. (15), we have

$$(d+rI)U^+a=U^+T^+p. \tag{17}$$

Denoting $v_n=(U^+T^+p)_n$, we find from Eq. (17)

$$(U^+a)_n=\frac{v_n}{d_n+r}. \tag{18}$$

Now, Eq. (13) gives the following equation for the regularization parameter r:

$$\|a\|^2=\|Ua\|^2=\sum_n\frac{|v_n|^2}{(d_n+r)^2}=a_0^2. \tag{19}$$

This is a single scalar equation with respect to r, and in spite of its nonlinearity it can be easily solved numerically. Substituting the obtained values of r into Eq. (18), we calculated the vector **a**

$$\mathbf{a}=U\left(\frac{v_n}{d_n+r}\right). \tag{20}$$

After the mode amplitudes $a_m(\omega_i)$ were obtained, the complex sound pressure within a chosen frequency bin was calculated as the right-hand side of Eq. (10).

The next step in the modal approach was to find how close the reconstructed sound field was to the measured one. To do this, we introduced the norm of the measured acoustic signal

$$N_P^2=\sum_{b,i}|p_b(\omega_i)|^2. \tag{21}$$

Furthermore, we introduced the norm of the difference between the measured and reconstructed signals

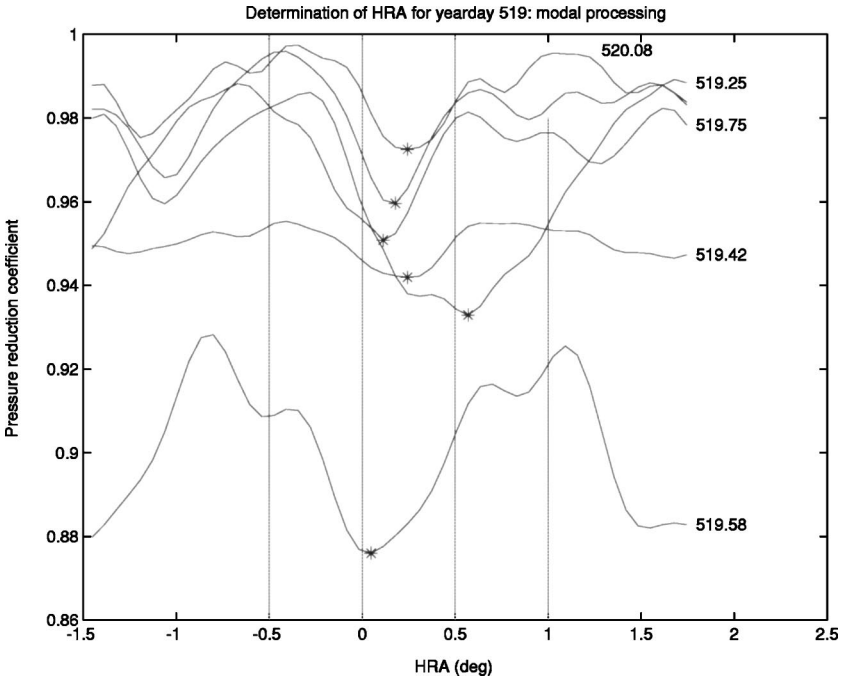


FIG. 9. Pressure reduction coefficients versus the horizontal refraction angles for six transmissions of the 519th day of the experiment. Asterisks mark absolute minima in these coefficients.

$$N_{dP}^2=\sum_{b,i}\left|p_b(\omega_i)-\sum_{m=1}^{M}u_m(z_b,\omega_i)\times\exp(-i\xi_m(\omega_i)x_b-ik_y y_b)a_m(\omega_i)\right|^2. \quad (22)$$

Then, the ratio

$$\kappa_{dP}=\frac{N_{dP}}{N_P}, \quad (23)$$

characterized how well the measured sound field was reconstructed. Small and large values of κ_{dP} correspond to good and poor reconstructions, respectively. The coefficient κ_{dP} is referred to as a ''pressure reduction coefficient.'' The pressure reduction coefficients obtained independently for each narrow frequency bin were then averaged over all 80 bins.

The final step in the modal approach was to calculate the coefficients κ_{dP} for a set of 50 different HRA α, equally spaced within the interval $(-1.5°,2.0°)$. A minimum value of κ_{dP} corresponds to the ''true'' HRA. The difference between minimum and maximum values of κ_{dP} is referred to as a contrast. When the contrast was low (less than 0.01), and hence, the minimum was not well pronounced, the result was rejected and the HRA was not determined.

B. Results

The dependences of the pressure reduction coefficient κ_{dP} versus the HRA α for all six transmissions of day 519 is shown in Fig. 9. The values of the HRA corresponding to absolute minima in these dependences are marked by asterisks.

Corresponding mode amplitudes $|a_m|$ averaged over all frequency bins versus mode number are shown in Fig. 10. The curves corresponding to different transmissions look fairly similar, except for one corresponding to day 519.58. Note that this is exactly the same time when the pressure reduction coefficient is the lowest in Fig. 9. Generally, we observe an increase of amplitudes for higher-order modes. The reason for this increase is not clear; it might be the result of bottom scattering in the vicinity of the VLAs. The curves shown in Figs. 9 and 10 are typical for other days of the experiment.

As was mentioned, the cases when the contrast was less then 0.01 were rejected. The cases with values of a pressure reduction coefficient greater than 0.985 were rejected also. The remaining results of the dependence of the HRA on time are plotted in Fig. 11. Most of the data are concentrated at small values of the HRA; however, there are some cases when the HRA are much different from typical values. These cases were also rejected, and only the values of α within the interval $(-0.3°,0.6°)$ were selected. These values and their fit by a polynomial of the order of 10 is shown in Fig. 5(b). One can see that scattering of α is significant and is of the order of the observed trends. However, the solid line in Fig. 5(b) is similar to that in Fig. 5(a). This supports the conclusion that a seasonal trend in the dependence of the HRA on time exists. It should be noted here that Figs. 5(a) and (b) were obtained by different independent approaches.

Finally, note that the dependence of the HRA on time was obtained in Ref. 17 by another approach. The results obtained are very similar to those in Figs. 5(a) and (b). All these indicate that a seasonal trend in the HRA probably does exist.

VI. CONCLUSIONS

Horizontal refraction of low-frequency acoustic signals that were transmitted for about a year from Kauai, Hawaii and received at the NPAL billboard array near the California coast was studied. Both ray-based and mode-based approaches were used in data processing. The results obtained with these two approaches are similar.

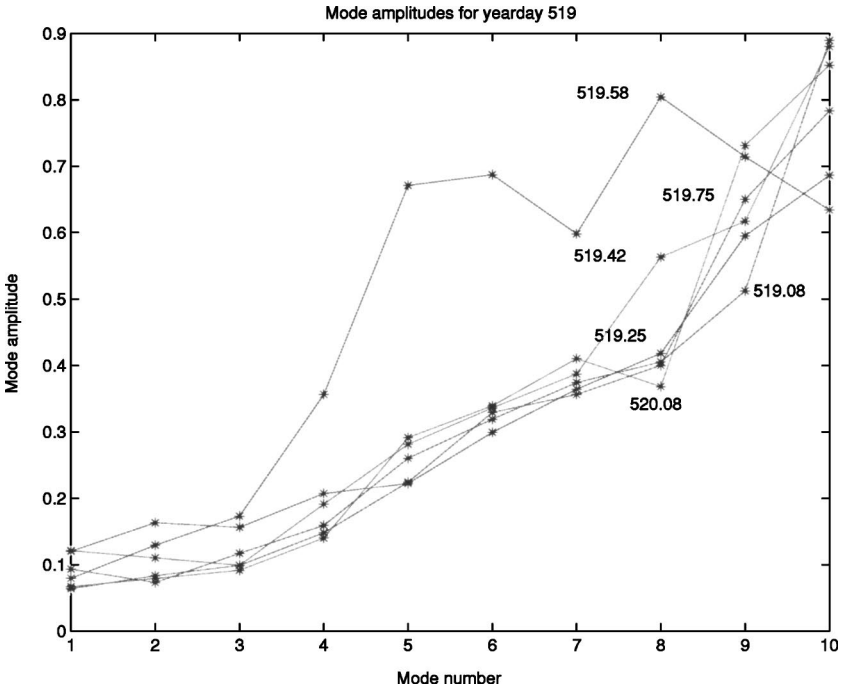


FIG. 10. Mode amplitudes as a function of mode number for the same transmissions as in Fig. 9.

In spite of relatively strong short-period variations in the HRA and rather low values of cross gradients of the sound speed in the Pacific region under investigation, changes in the HRA of the order of 0.4 deg were measured. This happened during the midspring–midsummer period of 1998 and of 1999 when the data revealed weak but systematic bending of the rays towards the south. This suggests that measurements of horizontal refraction may be used for remote sensing of inhomogeneities in the ocean. Appropriate acoustic measurements carried out in more active regions of the ocean and at shorter distances may allow us to estimate temporal variability of horizontal inhomogeneities, including very short time scales. This information is presently difficult to obtain by *in situ* measurements at reasonable cost.

Also, it was found that the upper limit of the standard deviation of the arrival angles due to sound scattering by internal waves was $\overline{\sigma}=0.37$ deg. This value is close to the theoretical estimate $\overline{\sigma}=0.26$ deg of the standard deviation of the HRA due to sound scattering by internal waves. According to this theoretical estimate, the standard deviation of the HRA is proportional to the frequency f and the square root of the sound propagation path R. Therefore, the value of $\overline{\sigma}$ measured at $f=75$ Hz and $R=3900$ m could be used for scaling. This value can help to determine the values of horizontal gradients of sound speed which can be monitored by measuring horizontal refraction at particular distances and frequencies.

The standard deviation $\overline{\sigma}=0.37$ deg corresponds to the

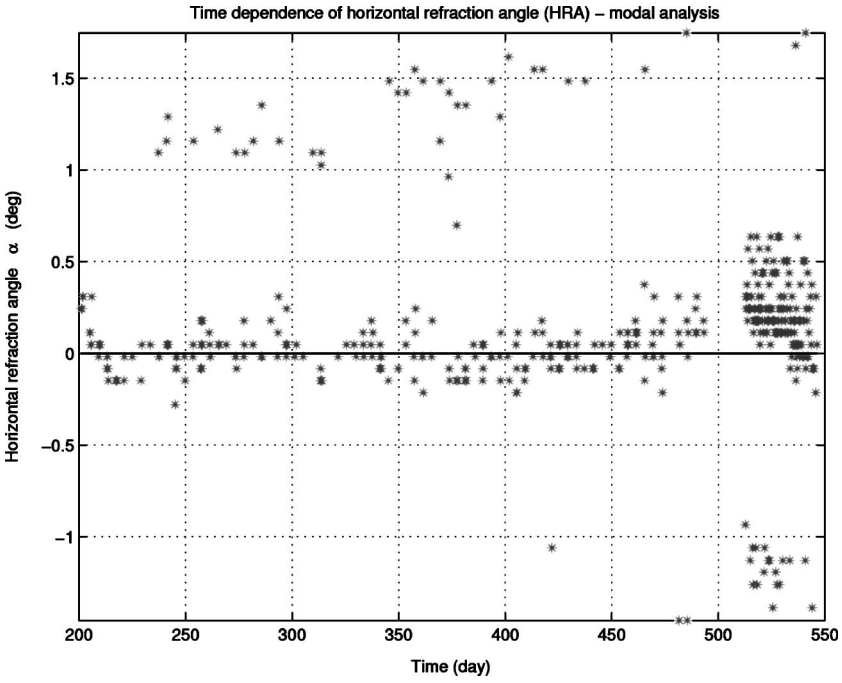


FIG. 11. Dependence of the horizontal refraction angle α on time obtained by the mode-based approach. The values of α are selected according to contrast and maximal pressure reduction coefficient criteria only.

correlation radius $r_c \sim 700$ m of the sound field impinging on the array. This value of r_c is of the order of those obtained from the dependence of the correlation coefficient on the distance between VLAs; see Figs. 5(a) and (b).

ACKNOWLEDGMENT

This work was supported by the ONR Grant N0001402IP-20035.

[1] A. G. Voronovich and E. C. Shang, ''A note on horizontal-refraction-modal tomography,'' J. Acoust. Soc. Am. **98**, 2708–2716 (1995).
[2] A. G. Voronovich and E. C. Shang, ''Numerical simulations with horizontal-refraction-modal tomography. I. Adiabatic propagation,'' J. Acoust. Soc. Am. **101**, 2636–2643 (1997).
[3] A. G. Voronovich and E. C. Shang, ''Horizontal refraction modal tomography of the ocean with mode interactions,'' IEEE J. Ocean. Eng. **24**, 224–231 (1999).
[4] W. H. Munk, ''Horizontal deflection of acoustic paths by mesoscale eddies,'' J. Phys. Oceanogr. **10**, 596–604 (1980).
[5] R. N. Baer, ''Propagation through a three-dimensional eddy including effects on an array,'' J. Acoust. Soc. Am. **69**, 70–75 (1981).
[6] S. Itzikowitz, M. J. Jacobson, and W. L. Siegmann, ''Modeling of long-range acoustic transmission through cyclonic and anticyclonic eddies,'' J. Acoust. Soc. Am. **73**, 1556–1566 (1983).
[7] Yu. P. Lysanov and V. E. Ostashev, ''Horizontal refraction of sound by intrathermocline lens in the ocean,'' Izv., Acad. Sci., USSR, Atmos. Oceanic Phys. **28**, 440–443 (1992).
[8] O. Godin, ''A 2-D description of sound propagation in a horizontally inhomogeneous ocean,'' J. Comput. Acoust. **10**, 123–151 (2002).
[9] O. Godin, A. Voronovich, and V. Zavorotny, ''Coherent scattering of under water sound and its implications for ocean remote sensing,'' in Proceedings of the 10th International Congress on Sound and Vibrations, 2539–2546, Stockholm, Sweden (2003).
[10] R. Oba and S. Finette, ''Acoustic transmission through anisotropic internal wave fields: Transmission loss, cross-range coherence, and horizontal refraction,'' J. Acoust. Soc. Am. **111**, 769–784 (2002).
[11] A. G. Voronovich, V. E. Ostashev, and the NPAL group (J. A. Colosi, B. D. Cornuelle, B. D. Dushaw, M. A. Dzieciuch, B. M. Howe, J. A. Mercer, R. C. Spindel, and P. F. Worcester) ''Experimental studies of horizontal refraction of acoustic signals in the ocean,'' Izv. Atmos. Oceanic Phys. **38**, 716–719 (2002).
[12] A. G. Voronovich, V. E. Ostashev, and the NPAL Group (J. A. Colosi, B. D. Cornuelle, B. D. Dushaw, M. A. Dzieciuch, B. M. Howe, J. A. Mercer, R. C. Spindel, and P. F. Worcester), ''Horizontal refraction of acoustic signals propagating over a long range in the ocean,'' in Proceedings of Acoust. Soc. Am. Meeting, Cancun, Mexico (2002).
[13] The NPAL Group (J. A. Colosi, B. D. Cornuelle, B. D. Dushaw, M. A. Dzieciuch, B. M. Howe, J. A. Mercer, R. C. Spindel, and P. F. Worcester), ''The North Pacific Acoustic Laboratory experiment,'' J. Acoust. Soc. Am. **109**, 2384 (2001).
[14] P. Worcester, ''NPAL98 Billboard Array Experiment Plan,'' at http://atoc.ucsd.edu/npal/Billboard/plan/html (1998).
[15] See Fig. 2.2 in: W. Munk, P. Worcester, and C. Wunsch, *Ocean Acoustic Tomography* (Cambridge University Press, Cambridge, 1995).
[16] M. E. Conkright, J. I. Antonov, O. Baranova, T. P. Boyer, H. E. Garsia, R. Gelfeld, D. Johnson, R. A. Locarnini, P. P. Murphy, T. D. O'Brien, I. Smolyar, and C. Stephens, *World Ocean Database 2001*, edited by S. Levitus (NOAA Atlas NESDIS 42, U.S. Government Printing Office, Washington, D.C., 2002), Vol. 1: Introduction.
[17] M. A. Dzieciuch and P. F. Worcester (personal communication).
[18] M. A. Dzieciuch, P. Worcester, and W. Munk, ''Turning points filters—Analysis of sound propagation on a gyre-scale,'' J. Acoust. Soc. Am. **110**, 135 (2001).
[19] A. G. Voronovich, ''Low-frequency sound propagation through random internal waves,'' in Proceedings of the Fourth European Conference on Underwater Acoustics, edited by A. Alippi and G. B. Canelli, **2**, pp. 751–756 (1998).
[20] K. E. Wage, A. B. Baggeroer, and J. C. Preisig, ''Modal analysis of broadband acoustic Receptions at 3515-km range in the North Pacific using short-time Fourier techniques,'' J. Acoust. Soc. Am. **113**, 801–817 (2003).
[21] J. R. Buck, J. C. Preisig, and K. E. Wage, ''A unified framework for mode filtering and the maximum *a posteriori* mode filter,'' J. Acoust. Soc. Am. **103**, 1813–1824 (1998).
[22] K. E. Wage and the ATOC Group, ''Comparison of broadband mode arrivals at ranges of 3515 km and 5171 km in the North Pacific,'' J. Acoust. Soc. Am. **113**, 2279 (2003).
[23] K. E. Wage and the NPAL Group, ''Analysis of mode coherence and intensity at megameter ranges,'' J. Acoust. Soc. Am. **109**, 2385 (2001).
[24] K. E. Wage, A. B. Baggeroer, and J. C. Preisig, ''Modal structure of low-frequency, broadband receptions at megameter ranges,'' J. Acoust. Soc. Am. **105**, 1116 (1999).

Analysis of multipath acoustic field variability and coherence in the finale of broadband basin-scale transmissions in the North Pacific Ocean

John A. Colosi
Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

Arthur B. Baggeroer
Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Bruce D. Cornuelle, Matthew A. Dzieciuch, Walter H. Munk, and Peter F. Worcester
Scripps Institution of Oceanography, La Jolla, California 92093

Brian D. Dushaw, Bruce M. Howe, James A. Mercer, and Robert C. Spindel
Applied Physics Laboratory, University of Washington, Seattle, Washington 98105

Theodore G. Birdsall and Kurt Metzger
University of Michigan, Ann Arbor, Michigan 48109

Andrew M. G. Forbes
Division of Oceanography, CSIRO, Hobart, Tasmania 7001 Australia

(Received 9 February 2004; revised 10 August 2004; accepted 11 August 2004)

The statistics of low-frequency, long-range acoustic transmissions in the North Pacific Ocean are presented. Broadband signals at center frequencies of 28, 75, and 84 Hz are analyzed at propagation ranges of 3252 to 5171 km, and transmissions were received on 700 and 1400 m long vertical receiver arrays with 35 m hydrophone spacing. In the analysis we focus on the energetic ''finale'' region of the broadband time front arrival pattern, where a multipath interference pattern exists. A Fourier analysis of 1 s regions in the finale provide narrowband data for examination as well. Two-dimensional (depth and time) phase unwrapping is employed to study separately the complex field phase and intensity. Because data sampling occured in 20 or 40 min intervals followed by long gaps, the acoustic fields are analyzed in terms of these 20 and 40 min and multiday observation times. An analysis of phase, intensity, and complex envelope variability as a function of depth and time is presented in terms of mean fields, variances, probability density functions (PDFs), covariance, spectra, and coherence. Observations are compared to a random multipath model of frequency and vertical wave number spectra for phase and log intensity, and the observations are compared to a broadband multipath model of scintillation index and coherence. [DOI: 10.1121/1.1854615]

PACS numbers: 43.60.Cg, 43.30.Ft, 43.20.Fn [AIT] Pages: 1538–1564

I. INTRODUCTION

Low-frequency, long range acoustic transmission data have been collected in the North Pacific Ocean in the mid- and late 1990s as part of the Acoustic Thermometry of Ocean Climate (ATOC) program (see Fig. 1). Ship suspended and bottom mounted sources have been used and long vertical arrays have received the signals at ranges of 3250, 3515, and 5100 km. The sources have mainly been at 75 Hz (20.1 Hz bandwidth at 3 dB), but there was a brief deployment of an HLF-6A source that simultaneously transmitted at 28 and 84 Hz with a bandwidth of about 10 Hz (3 dB) (Worcester *et al.*, 2000; Dzieciuch *et al.*, 2001). Transmitted signals were phase coded pseudorandom sequences (m sequences). These data showed sufficient temporal stability to allow a coherent summation of signal replicas from 10 to 20 minutes to bolster the signal-to-noise ratios (SNR). The resulting signal gains allowed the detection of early ray-like arrivals for use in acoustic thermometry (Worcester *et al.*, 1999; Dushaw *et al.*, 1999; and The ATOC Consortium, 1998). These signals were also exploited by Freitag and Stojanovic (2001) who examined the feasibility of long range communication using 75 Hz, 3250 km transmissions by processing both vertically and temporally.

However, the temporal and vertical variability of these signals has not been analyzed in detail. Colosi, Tappert, and Dzieciuch (2001) presented the results on the effects of *m*-sequence integration on mean pulses shapes and intensity PDFs for a 3252 km path, and Wage *et al.* (2002) presented the results of processing loss (temporal coherence) of resolved low-order modes at 3500-km range by narrowband processing at 85, 75, and 65 Hz. Thus, a quantitative analysis of these data are needed to understand the nature of low-frequency acoustic scattering at these ranges and to guide better signal processing algorithms. In this endeavor the focus is on the energetic finale region, where the arrival can be viewed as an interference pattern of either coupled modes (Wage *et al.*, 2003; Wage *et al.*, 2002; Colosi and Flatté, 1996) or chaotic rays (Simmen, Flatté, and Wang 1997;

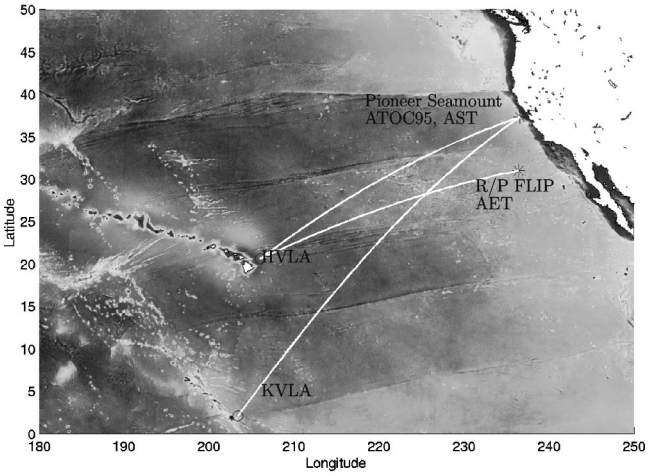


FIG. 1. Acoustic transmission paths studied in this paper. The Acoustic Engineering Test (AET) transmitted from R/P FLIP south–southwest of San Diego to a vertical line array located off the island of Hawaii; the range was 3252 km. A bottom mounted source on Pioneer seamount off the coast of San Francisco transmitted to two VLAs: one off the island of Hawaii (HVLA) and one near Kiritimati (Christmas) Island (KVLA); these ranges were 3515 and 5100 km. During the AST a ship suspended source off Pioneer seamount transmitted to the Hawaii and Kiritimati VLAs.

Beron-Vera *et al.*, 2003). Figure 2 shows the location of the finale and wave front regions in the arrival pattern for a 75 Hz, 3252 km transmission. The finale region is also characterized by intense depth spreading of the acoustical energy, thus leading to an in-filling of the shadow zone below the last arrivals (Duda *et al.*, 1992; Colosi *et al.*, 1994; Colosi and Flatte, 1996). This is the most energetic region of the wave front and therefore there is the benefit of favorable SNRs. Further, an analysis of this region of the wave front is expected to demonstrate the worst-case scenario, since variability in the early wave front region (Fig. 2) has been shown to be weaker than in the finale (Colosi *et al.*, 1999a, 1999b; Colosi, Tappert, and Dzieciuch, 2001; Beron-Vera *et al.*, 2003).

Previous results have shown that the finale region is a multipath interference pattern in which the scintillation index (SI) and variance of log intensity [$\sigma^2_{\ln I}$; See Eq. (7)] are very close to the values of full saturation (Colosi, Tappert, and Dzieciuch, 2001). In full saturation the phase is uniformly distributed between 0 and 2π and the probability density function (PDF) of intensity is close to exponential; thus SI$\sim$1 and $\sigma_{\ln I}\sim$5.6 dB. The results of this paper are consistent with these findings but the present analysis is also able to show how much of this variability occurs at different time scales. Thus, over multiday records of both high and low acoustic frequency data, the near-saturation behavior is found with scintillation values above saturation caused primarily by modulation processes. However, on time scales of 1 to 40 min, this variability goes down considerably. For example, in the 75–84 Hz range SI varies between 0.44 and 0.74, while $\sigma_{\ln I}$ varies between 4.0 and 4.7 dB. At 28 Hz frequency, the values fall even more sharply with SI between 0.29 and 0.70, and $\sigma_{\ln I}$ between 3.1 and 4.4 dB. The case is similar for the phase, which is analyzed using a simple two-dimensional phase unwrapping technique. The previous analysis over multiday records using travel time as a surrogate for phase, found rms phase variations in the finale of order one to two cycles (Colosi *et al.* 1999a). In the present analysis, over the 1 to 40 min time scale rms phase variations are found to be 0.08 to 0.13 cycles in the 75 to 84 Hz region and 0.06 to 0.09 cycles in the 28 Hz region. A direct analysis of the phase over longer time scales is not possible since the phase cannot be tracked over the transmission gaps (typically 4 h). The PDF of phase is shown to be very close to Gaussian.

Frequency spectra of phase and log intensity at 75 and 84 Hz show a red spectrum with no apparent roll-off at low frequency, thus implying that the aforementioned phase and intensity variability would increase over a longer observation time. At 28 Hz acoustic frequency, the spectra have a completely different shape and are rather flat. The difference in spectral shapes at high and low acoustic frequency implies significantly different scattering physics. Comparing these spectra to the random multipath model of Dyson *et al.* (1976) in which the spectra are dominated by fades at high frequency and random walking at low frequency, it is found that the high acoustic frequency phase spectra have some features consistent with the model in the 1 to 10 cph frequency range (random walk), but log-intensity spectra compare very poorly to the model. Applying the Dyson model to the vertical wave number spectra of phase and log intensity the agreement is fairly good, which demonstrates that the interference pattern in depth is more fully randomized than the interference pattern in time. Similar comparisons using the Dyson model, in which spatial data show better agreement than temporal data, have been documented by Porter and Spindel (1977) for moored and drifting receptions of 200 and 400 Hz sound in the North Atlantic.

Having examined phase and intensity variability separately, signal coherence is treated, which involves both intensity and phase simultaneously. In the present work it is shown that the data are consistent with a model in which the following three properties are satisfied: (1) the phase and amplitude are uncorrelated; (2) the phase is Gaussian; and (3) the amplitude decorrelation is important to coherence. Thus, a useful model for coherence is $\langle \psi\psi^*(\Delta)\rangle = \langle AA(\Delta)\rangle \exp[-D(\Delta)/2]$, where $\langle AA(\Delta)\rangle$ is the amplitude covariance function, $D(\Delta)$ is the phase structure function, and Δ could be a temporal or depth lag. This model differs from previous work (Flatté *et al.*, 1979; Flatté and Stoughton, 1988) by including the amplitude term, and not assuming D as quadratic in Δ. The acoustic fields decohere very rapidly in depth, but over 20 and 40 min time scales the time coherence is quite good. While full coherence functions are presented, a good measure of the temporal coherence embodied in a single number are the temporal mean fields. At the higher frequencies of 84 and 75 Hz the temporal mean fields represent 42% to 64% of the total energy, while at the lower 28 Hz the mean fields compose an impressive 60% to 71% of the total energy.

The present analysis also allows a comparison at different propagation ranges, although the propagation paths are geographically separated. Interestingly, phase and intensity variances are somewhat less at the farthest propagation range, and coherence shows no strong dependence on range.

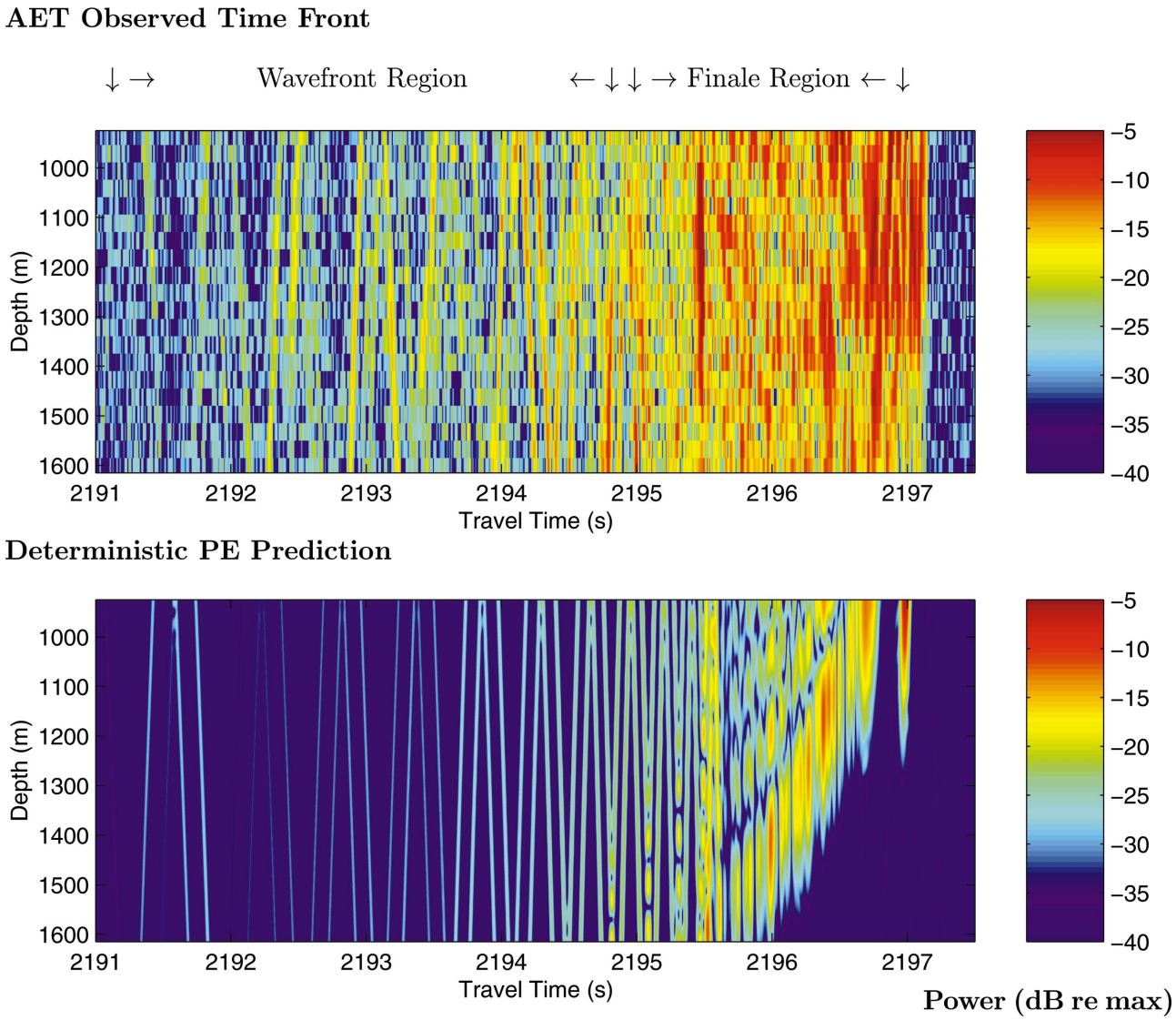


FIG. 2. An example of time fronts from the 3252 km AET (upper) and a numerical model for the AET without internal-wave sound-speed perturbations (lower). The finale region is defined roughly between the travel times of 2195 and 2197.

Finally, all the finale signal statistics were calculated using both full bandwidth and narrow band data, and the acoustic variability is not a significant function of the bandwidth. The narrow band data were obtained by Fourier transforming a 1 s record of the finale and selecting out the carrier band with a resolution of about 1 Hz. A kinematical model of scintillation index and coherence assuming strong multipath interference can explain the weak bandwidth dependence, and points to new observations that could elucidate bandwidth-dependent scattering.

The coherence results presented in this study complement the analysis of Wage *et al.* (2005), who computes temporal coherence functions of low-order mode arrivals from the same datasets discussed in this paper. Agreement between the coherence results of this study and those of Wage would indicate that the cross-modal coherences are very small. While a theoretical description of the modal and full field fluctuations is presently lacking, this will be an area of further active research.

The structure of this paper is as follows. In Sec. II we describe the basic acoustical observations and the processing required to obtain the reduced data used in this analysis. In Secs. III and IV we describe the statistics and spectra of phase and intensity, respectively. In Sec. IV we also address the phase-intensity correlation. Coherence and mean field energy are examined in Sec. V. Section VI has a discussion and more interpretation of the observations where kinematical model of multipath coherence and scintillation are presented. A brief summary is given in Sec. VII.

II. OBSERVATIONS

In this study, acoustical observations obtained from the Acoustic Thermometry of Ocean Climate (ATOC) North Pacific network are utilized (see Fig. 1 and Table I for a summary). The first set of data comes from the November 1994 Acoustical Engineering Test (AET) (see Worcester *et al.*, 1999, and Colosi *et al.*, 1999a for details), in which a source suspended from R/P FLIP several hundreds of miles south–southwest of San Diego transmitted signals with a center frequency of 75 Hz and a bandwidth of 20.1 Hz (3 dB) for 6 days to a 700 m long 20-element vertical line array (VLA)

TABLE I. Experimental data used in this study and the relevant sampling parameters. The parameters N_l, N_z, and N_τ are the number of transmissions, number of hydrophones, and number of time samples per transmission, respectively. The total number of data points is $N_{\text{total}}=N_l N_z N_\tau$. The time between samples is $\delta\tau$ and the depth spacing between hydrophones is δz. The total observation time per transmission, time between the first and last sample, is shown in the last column.

Experiment (frequency)	Range (km)	Start date	End date	N_l	N_z	N_τ	N_{total}	$\delta\tau$ (s)	δz (m)	Obs. time (min)
AET75	3252	Nov 17, 1994	Nov 23, 1994	35	20	40	28000	27.28	35	17.73
AET75		⋯	⋯	1	20	84	1680	27.28	35	37.73
HVLA75	3515	Dec 29, 1995	May 24, 1996	196	40	10	81200	109.12	35	16.37
KVLA75	5171	Dec 29, 1995	Sept 4, 1996	283	40	10	120800	109.12	35	16.37
HVLA28	3502	June 30, 1995	July 8, 1996	29	20	10	5800	109.12	35	16.37
HVLA28		⋯	⋯	5	20	20	2000	109.12	35	34.55
KVLA28	5158	June 30, 1995	July 8, 1996	31	40	10	12400	109.12	35	16.37
KVLA28		⋯	⋯	5	40	20	4000	109.12	35	34.55
HVLA84	3502	June 30, 1995	July 8, 1996	29	20	10	5800	109.12	35	16.37
HVLA84		⋯	⋯	5	20	20	2000	109.12	35	34.55
KVLA84	5158	June 30, 1995	July 8, 1996	31	40	10	12400	109.12	35	16.37
KVLA84		⋯	⋯	5	40	20	4000	109.12	35	34.55

deployed off Hawaii (See Fig. 1). The next group of data comes from a much longer duration experiment conducted in a similar location. In November–December of 1994, two 40 element, 1400 m long vertical arrays were deployed: one off Hawaii (HVLA) and another off Kimitimati (Christmas) Island (KVLA) (Worcester *et al.*, 2000; Dzieciuch *et al.*, 2001). These VLAs received signals transmitted from two different sources. From December 1994 through August 1995, the VLAs received 75 Hz center frequency, broadband (20.1 Hz, at 3 dB) transmissions from a bottom mounted source on Pioneer seamount off the coast of San Francisco, California (see Fig. 1). This 8–9 month series of transmissions is referred to as the ATOC95 experiment. For a 9 day period from 30 June 1995–8 July 1995, the 75 Hz transmissions were interrupted so that a ship suspended HLF-6A source in deep water off Pioneer seamount could simultaneously transmit 28 and 84 Hz signals with a bandwidth of about 10 Hz (3 dB). This dual frequency experiment is referred to as the Alternate Source Test (AST), since the primary objective of this test was to examine the feasibility of using ultra-low-frequency transmissions for acoustic thermometry. In all of these experiments at four hour intervals, the source would transmit 44 replicas of a 27.28 s pseudorandom code with a duration of roughly 1200.320 s (20.005 min). In the AET, 40 replicas were recorded (to avoid end effects in the processing), but in the ATOC95 and AST experiments, replicas were averaged in groups of 4 to give 10 time samples and thus reduce the data volume. In the AET and AST, a few back-to-back 20 min transmissions were made, thus yielding longer continuous time series. Table I lists various parameters of the transmission datasets and sampling parameters. In Table I and throughout this paper, the ATOC95 and AST data are denoted by the VLA name and the center frequency, i.e., HVLA75 and KVLA75 for ATOC95, and HVLA28, HVLA84, KVLA28, and KVLA84 for the AST. The AET data is denoted somewhat differently by AET75.

As mentioned previously in this paper, the focus is on the transmission finale region, where SNR is best. For the HVLA data the SNR is roughly 12 dB (Wage *et al.*, 2002), while the AET data (which were more finely sampled in time by a factor of 4) had roughly 6 dB SNR (Colosi *et al.*, 2001). The SNR at KVLA benefits from a lower noise level and is also about 12 dB (Wage, year). For the acoustic observations the ''raw'' broadband complex demodulates are defined as

$$\psi_r(z,\tau,l,T(z))=A(z,\tau,l,T(z))\exp[i\phi(z,\tau,l,T(z))], \tag{1}$$

where A is the amplitude and ϕ is the phase. The dependent variables are z, the hydrophone depth, τ is geophysical time, $T(z)$ is the travel time relative to the final cutoff and corrected for the mooring shape to give the z dependence, and lastly l denotes a transmission sequence of 40 or 80 m sequences. Thus, the τ parameter varies in increments of 27.28 or $4*27.28=109.12$ s, depending on the dataset, and l denotes the transmission sequence typically separated by 4 h or several days/weeks. Also, narrow band demodulates are considered, which are obtained by Fourier analysis of the record $T\pm 0.5$ s to give a demodulate at the center frequency ± 1 Hz. In the subsequent discussion, the T dependence is dropped, and thus the finale demodulates, 1 s from the final cutoff, are only denoted by the three indexes: z, τ, and l.

Finally, some special cases must be mentioned. For the AST data recorded on the HVLA, the lower half of the 40 element array had failed, and thus HVLA28 and HVLA84 only have the upper 20 phones. Of these upper 20 phones, there was a failure on phone 10, and thus to get a complete gap-free record, interpolation was used to fill in the phone 10 record. Also, special consideration was needed for the 80 period receptions for the AET. The first 40 period reception lasting 1091.200 s was followed by a gap while the system wrote the data to disk. The second 40 period reception started exactly 1200 s after the start of the first 40 period reception, leaving a gap of 108.800 s (3.988 periods) between receptions. Thus, there are two effects of concern: one is a time gap of duration 3.988 m seq's (2.27 min) in the record and the other is a phase shift of the first 40 period record relative to the second 40 period record, since the time gap was not an integer multiple of the m seq duration. A phase shift correction was applied and again interpolation

was used to fill the time gaps to give a continuous time record, yielding 84 time samples for the single long AET data (Table I).

Next, further processing of the extracted finale data are discussed. Two versions of ψ are initially considered: ψ_r, the raw demodulation with a wrapped phase, ϕ_w; and $\psi_u = |\psi_r| \exp[i\,\phi_u]$, the demodulations with a two-dimensional 2-D unwrapped phase, ϕ_u. At each depth the data is normalized to give the unit mean intensity,

$$\langle I \rangle (z_j) = \frac{1}{N_\tau N_l} \sum_{k=1}^{N_\tau} \sum_{l=1}^{N_l} |\psi(z_j, \tau_k, l)|^2 = 1. \tag{2}$$

A smooth unwrapped phase function, ϕ_u, is determined such that the mean-square difference between *gradients* calculated from wrapped and unwrapped phases are minimized. An unweighted least squares method is chosen that results in an equation for ϕ_u, which is the discrete form of Poisson's equation and that can be solved easily using fast cosine transform methods (Zebker and Liu, 1998; Ghilia and Romero, 1994; see Appendix A for details). Using this smoothness criterion means that the unwrapped phase function is not necessarily different from the wrapped phase function by integer multiples of 2π.

Figure 3 shows examples of 40 min duration broadband acoustical data from several experiments at several different center frequencies. Most of the data show a rather grainy intensity pattern, however, the KVLA28 data (but not the HVLA28 data!) shows a much smoother pattern. The middle and right columns of Fig. 3 show the wrapped and unwrapped phases. The 2-D phase unwrapping, being a least square procedure, generates a very smooth phase field, but the accuracy of the phase unwrapping must be tested. To do so, the depth and time lagged coherence functions are computed using the raw demodulates (ψ_r) and the demodulates using the unwrapped phase ψ_u. The coherence is computed according to, $\langle \psi\psi^*(\Delta) \rangle = |\overline{\langle \psi\psi^*(\Delta) \rangle_{z\tau}}|$, where the averages are defined as in Appendix B, Δ could be a time or depth separation, and the absolute value is only taken after averaging over all z, τ, and l. Figure 4 shows the depth and time coherence functions for AET75 and HVLA75 and the agreement between the calculations from the two datasets (ψ_r and ψ_u) is superb. Similar agreement is seen for the other datasets and thus this test demonstrates that the 2-D phase unwrapping can be considered accurate.

Finally, there is the issue of mooring and source motion as well as residual mooring shape errors in the unwrapped phase functions $\phi_u(z,\tau,l)$. It must also be noted that tidal velocities along the acoustic path during a transmission may contribute significantly to the observed phase changes, as noted by Worcester *et al.* (1999). To crudely correct for the mooring/source motion, tidal effects, and possible residual mooring tilt, a least square fit to the phase function is done using a linear model,

$$\phi_m(z,\tau,l) = \langle \phi_u \rangle_{z\tau}(l) + z(\partial_z \phi_u)_m(l) + \tau(\partial_\tau \phi_u)_m(l), \tag{3}$$

to yield a corrected phase function given by, $\phi_c = \phi_u - \phi_m$. Table II lists the mean and rms variation of the fit parameters $(\partial_z \phi_u)_m(l)$ and $(\partial_\tau \phi_u)_m(l)$. Most of the phase trend temporal variability translates to mooring and/or source motion of less than 1 cm/s. The AET75, however, has very strong phase trends that are likely due to the large source motion that occurred during the experiment (see Worcester *et al.*, 1999). The vertical trends are surprisingly large, given that the expected mooring location errors are on the order of 1–2 m. These trends vary between 0.25 and 1 cycle per km, and may primarily be due to the phase variations caused by the vertical interference pattern.

III. PHASE FLUCTUATIONS

The ability to accurately represent the acoustic phase in this multipath interference pattern grants the luxury of examining the statistics of the acoustic field in polar coordinates (i.e., phase and intensity) as opposed to Cartesian coordinates (Dyson *et al.*, 1976).

A. Phase variance

The analysis of phase variability is carried out in both the depth and time directions, and it is important to distinguish between linear (trend) and nonlinear variability. A depth mean temporal phase trend and a time mean depth phase trend were removed, as discussed in the previous section; thus there are still fluctuations in the temporal and depth phase trends that are quantified in this section. The designation into trend and nontrend will be important in the spectral analysis of phase and in the calculation of coherence. Thus, taking the temporal variability as an example, for each depth z and for each transmission l a linear fit to the phase, ϕ_c, is made, yielding estimates of mean phase $\langle \phi_c \rangle_\tau(z,l)$ and trend $\partial_\tau \phi_c(z,l)$. The linear fit is then subtracted from ϕ_c to give a residual phase $\delta\phi_\tau$. A similar operation is performed to produce residual phase, $\delta\phi_z$, in the depth direction. Phase variances for both the detrended ($\delta\phi_\tau$ and $\delta\phi_z$) and original (ϕ_c) data are shown in Table III. The total variance is computed from the demeaned phases ϕ_c such that

$$\sigma^2_{\phi_\tau} = \overline{\langle (\phi_c - \langle \phi_c \rangle_\tau)^2 \rangle_{z\tau}}, \quad \sigma^2_{\phi_z} = \overline{\langle (\phi_c - \langle \phi_c \rangle_z)^2 \rangle_{z\tau}}. \tag{4}$$

The detrended phase variance in the τ and z directions are denoted by $\overline{\langle \delta\phi^2_\tau \rangle_{z\tau}}$, and $\overline{\langle \delta\phi^2_z \rangle_{z\tau}}$, respectively, and the rms phase trend variations in the τ and z directions are denoted by $\overline{\langle (\partial_\tau \phi_c)^2 \rangle_z}^{1/2}$ and $\overline{\langle (\partial_z \phi_c)^2 \rangle_\tau}^{1/2}$. Note that $\overline{\langle \partial_\tau \phi_c \rangle_z} = \langle \partial_z \phi_c \rangle_\tau = 0$ because of the initial phase trend removal from Eq. (3). It must be emphasized that the temporal phase variances shown in Table III are the result of observing the acoustic field over a limited temporal span (approximately 20 and 40 min) and therefore these results cannot be directly compared to estimates of phase variance in which an ensemble over all time scales of the ocean (or internal waves) are considered.

The temporal variability of phase shown in the right-hand columns of Table III is discussed first. What is perhaps most surprising is that the variability is significantly less than a cycle, even at the longer 40 min observation times. The linear changes in phase, on the whole, account for less than 50% of the total phase variance, though there are two exceptions in the poorly sampled 40 min observation time group

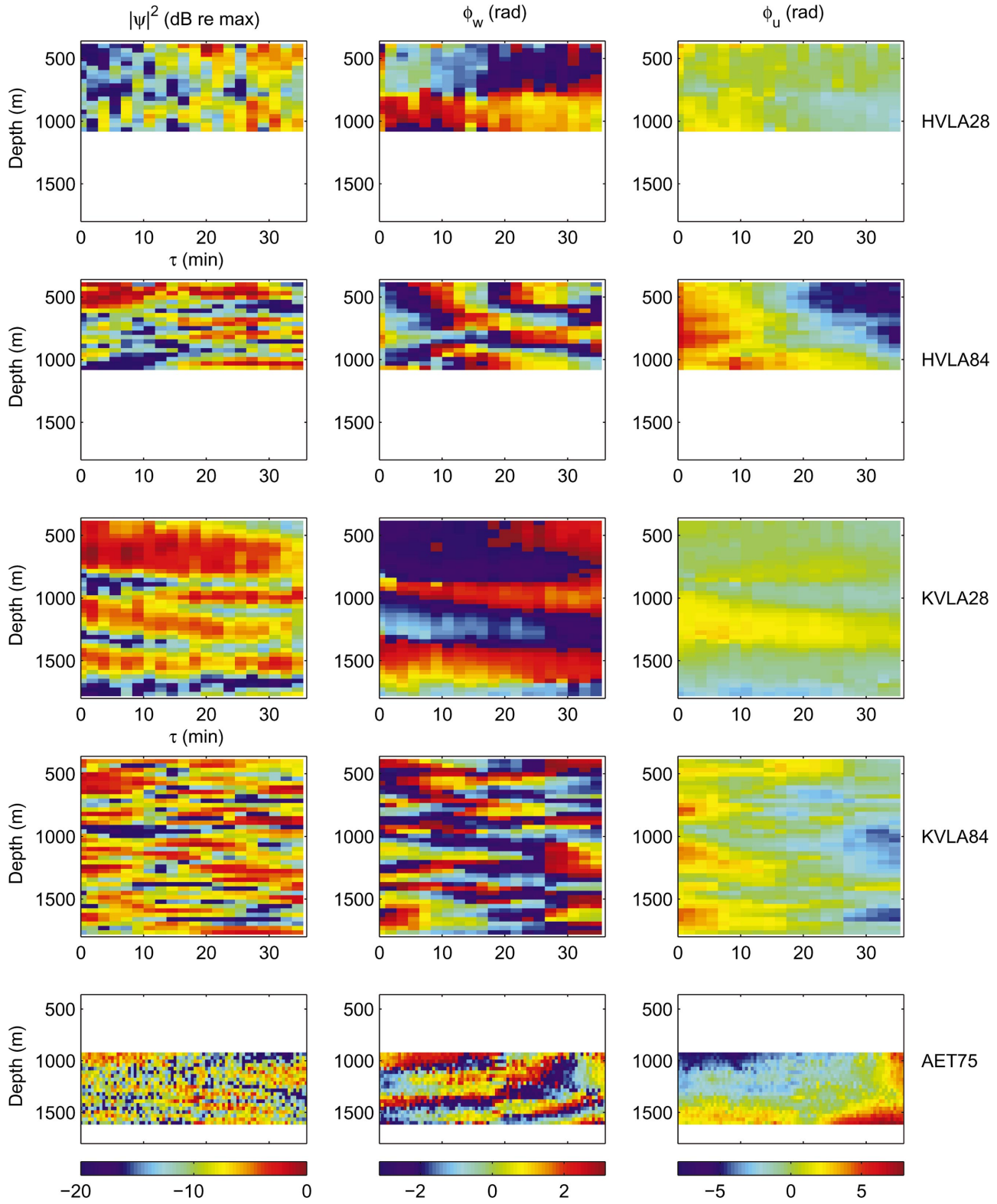


FIG. 3. Examples of finale data for long acoustic transmissions for several of the experiments treated in this paper. The leftmost column gives acoustic intensity (dB, relative to maximum). Center and right columns are the raw wrapped and 2-D unwrapped phases, respectively.

(HVLA84, and KVLA28). At the higher frequencies (84 and 75 Hz), the results are consistent between similar datasets like HVLA84 and HVLA75. Also, the AET data stand out with much larger variances, and this difference may be accountable to the strong, nonlinear, source motion experienced during the AET, as previously mentioned.

In the discussion of temporal phase variability there are three main dependencies that need to be examined: the dependence on propagation range (though the paths are geographically separated), center frequency, and bandwidth. Regarding the dependence on range, geometrical theory for a single path (Flatté *et al.*, 1979), which is clearly an idealiza-

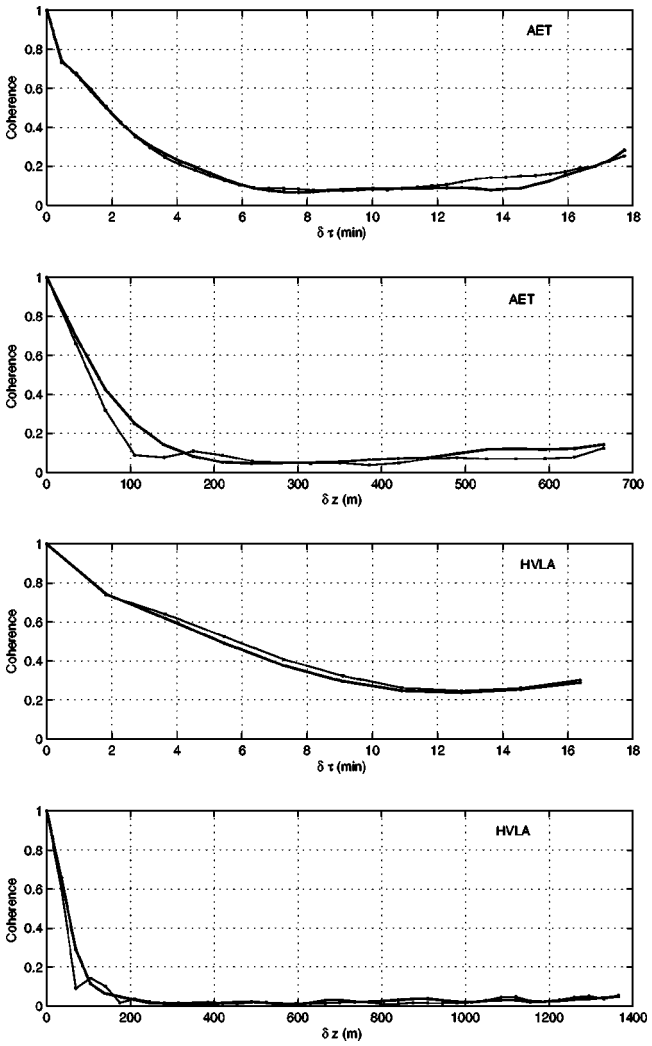


FIG. 4. Broadband depth- and time-lagged coherence from the AET75 and HVLA75 datasets, computed using two different data types. The light curve is computed using the raw demodulates, and the heavy curve is computed using the 2-D unwrapped phase demodulates.

tion given the multipath interference pattern, predicts phase variance to increase linearly. Thus, the predicted increase in variance for the present study are of the magnitude 50%. In both the high- and low-frequency groups (84 and 75 Hz; 28 Hz) these increases are not borne out, and, in fact, in most cases the KVLA data show less variance than the HVLA data. In making this comparison, simultaneous datasets are matched, like HVLA75 and KVLA75, or HVLA28 and KVLA28.

Comparing the simultaneous 84 and 28 Hz data, the lower frequency clearly shows a reduction in variance by a factor of roughly 3, though the 20 min observation time HVLA data shows only a factor of 1.5. Again the idealized, geometrical theory for a single path (Flatté *et al.*, 1979) predicts a reduction in variance by a factor of frequency squared (or nine in this case) that is clearly not observed. Finally, it will be noted from Table III that the phase statistics are not sensitive to whether broadband or narrow band data are used.

Next, the depth variability of the phase is discussed (Table III, left columns). The depth variability of the phase is much larger than the temporal variability since the data is

TABLE II. Statistics of linear phase corrections for mooring motion and tilt derived from the broadband data using Eq. (3). Both mean and rms variations are quoted. Statistics of the corrections for narrowband data (not shown) are very similar.

	Depth variation $\overline{(\partial_z\phi_u)_m} \pm \overline{\delta(\partial_z\phi_u)_m^2}^{1/2}$ (rad/km)	Time variation $\overline{(\partial_\tau\phi_u)_m} \pm \overline{\delta(\partial_\tau\phi_u)_m^2}^{1/2}$ (rad/s)1000
20 min		
Obs. time		
HVLA84	$+0.04\pm7.45$	$+0.45\pm3.28$
HVLA75	$+0.43\pm5.87$	-0.11 ± 1.73
AET75	-0.08 ± 5.49	$+0.19\pm6.40$
HVLA28	-0.53 ± 4.21	$+0.49\pm1.27$
KVLA84	-5.04 ± 5.92	$+0.51\pm2.51$
KVLA75	-0.56 ± 5.78	-0.02 ± 1.58
KVLA28	-1.41 ± 3.75	$+0.08\pm0.93$
40 min		
Obs. time		
HVLA84	-2.41 ± 5.00	-1.98 ± 1.76
AET75	$+9.04\pm\ \cdots$	$+2.40\pm\ \cdots$
HVLA28	-1.07 ± 2.96	-0.53 ± 0.59
KVLA84	-1.24 ± 3.03	-1.93 ± 0.73
KVLA28	-0.75 ± 1.64	-0.55 ± 0.26

spanning a large vertical extent of the interference pattern (see Fig. 3), however again, the magnitude is still less than one cycle. There is very little phase trend above the uniform correction $\overline{(\partial_z\phi_u)_m}$. Regarding the center frequency dependence, the 28 Hz data clearly shows less phase variability than the higher acoustic frequencies, since the interference pattern is a larger scale in the vertical (see Fig. 3). As in the temporal statistics case, there is no consistent pattern with respect to propagation range, and as before, these phase statistics are insensitive to whether broadband or narrow band data are used.

The last two phase statistics to be presented are the mean square phase rates, defined as

$$\nu_t^2 = \langle (d\phi/dt)^2 \rangle = \overline{\langle (d\,\delta\phi_\tau/dt)^2 \rangle_{z\tau}},$$
$$\nu_z^2 = \langle (d\phi/dz)^2 \rangle = \overline{\langle (d\,\delta\phi_z/dz)^2 \rangle_{z\tau}}. \quad (5)$$

The temporal phase rate will be important in the discussion of the frequency spectra of phase and log-intensity as predicted by random multipath theory (Dyson *et al.*, 1976), and the phase rate has been related to the coherence time (Flatté *et al.*, 1979). The phase rate in depth, ν_z^2, is also given in Table III and is related to the vertical wave number spectra of phase and log intensity as well as vertical coherence (Dyson *et al.*, 1976; Flatté *et al.*, 1979). As in previous phase statistics, the AET is anomalous and the geometrical theory scaling of ν_t (i.e., linear in frequency and square root in range) are not borne out by the data. Again, the failure of the single path geometrical theory is not surprising given the complex interference pattern. More discussion on phase rates will come later in the sections on spectra and coherence.

B. Phase PDF

The phase PDF, which is commonly assumed to be Gaussian, is an important quantity in understanding the propagation of waves through random media. In particular,

TABLE III. Broadband and narrow band (denoted with parentheses) phase statistics. Columns 1–4 give statistics along the depth direction and columns 5–8 give statistics along the time (τ) direction. Columns 1 and 5 give the total phase variance after correcting the unwrapped phase ϕ_u for a mooring motion and/or shape-induced time and depth trend [Eq. (3)]. Columns 2 and 6 give variances of phase after removing time-dependent depth trends and depth-dependent temporal trends, respectively. Columns 3 and 7 show the rms phase trends in the depth and time direction, respectively. The last column in each group gives the rms phase rate after removal of the phase trends.

	$\sigma^2_{\phi_z}$ (rad^2)	Depth $\overline{\langle \delta\phi_z^2\rangle_{z\tau}}$ (rad^2)	Variation $\overline{\langle(\partial_z\phi_c)^2\rangle_\tau}^{1/2}$ (rad/km)	ν_z (rad/km)	$\sigma^2_{\phi_\tau}$ (rad^2)	Time $\overline{\langle \delta\phi_\tau^2\rangle_{z\tau}}$ (rad^2)	Variation $\overline{\langle(\partial_\tau\phi_u)^2\rangle_z}^{1/2}$ (rad/s)1000	ν_t (rad/s)1000
20 min obs. time								
HVLA84	1.26(1.10)	1.19(1.02)	1.32(1.41)	23.8(22.8)	0.57(0.65)	0.39(0.36)	1.39(1.72)	7.46(7.05)
HVLA75	2.67(2.49)	2.63(2.46)	0.53(0.45)	26.7(26.8)	0.59(0.53)	0.33(0.28)	1.62(1.62)	7.42(6.63)
AET75	1.21(1.46)	0.89(1.07)	2.82(3.07)	25.4(25.7)	1.14(1.21)	0.79(0.75)	1.88(2.15)	31.0(28.5)
HVLA28	0.49(0.39)	0.39(0.32)	1.53(1.34)	14.1(13.0)	0.37(0.30)	0.28(0.23)	0.92(0.87)	7.07(6.29)
KVLA84	3.10(3.14)	3.05(3.11)	0.53(0.38)	24.8(27.1)	0.73(0.74)	0.45(0.46)	1.68(1.69)	7.41(7.42)
KVLA75	2.72(2.80)	2.68(2.76)	0.49(0.48)	26.3(28.4)	0.54(0.59)	0.24(0.24)	1.74(1.89)	5.87(5.46)
KVLA28	1.41(1.48)	1.39(1.47)	0.37(0.32)	12.5(15.2)	0.22(0.16)	0.15(0.10)	0.88(0.76)	4.99(3.97)
40 min Obs. Time								
HVLA84	1.76(1.66)	1.01(1.13)	4.28(3.59)	22.5(24.0)	1.51(1.27)	0.57(0.58)	1.54(1.32)	6.65(6.00)
AET75	1.54(2.14)	1.02(1.50)	3.56(3.96)	24.9(26.6)	1.91(1.75)	1.39(1.28)	1.15(1.09)	28.0(24.9)
HVLA28	0.52(0.33)	0.37(0.26)	1.95(1.24)	13.4(10.8)	0.48(0.25)	0.35(0.18)	0.58(0.44)	7.01(5.44)
KVLA84	1.58(2.34)	1.53(2.15)	0.59(1.06)	26.3(27.5)	1.20(1.34)	0.77(0.77)	1.04(1.19)	6.78(6.77)
KVLA28	1.23(1.04)	1.15(1.00)	0.68(0.49)	12.5(12.4)	0.37(0.28)	0.15(0.13)	0.74(0.62)	4.18(3.28)

non-Gaussian behavior may be due to non-Gaussian ocean sound speed structure, or strong caustic structures in the wave field. Passing through a caustic results in a $\pi/2$ phase shift, which is the source of the non-Gaussianity. Dzieciuch *et al.* (2004) have suggested that non-Gaussian ocean fine structure may play an important sound scattering role in addition to internal waves, and Williams *et al.* (2001) have found significant non-Gaussianity in high-frequency shallow water transmissions. If the PDF of phase is Gaussian, then the phase structure function will play a significant role in the observable of coherence (Flatté *et al.*, 1979).

Figure 5 shows the PDFs of the observed broadband phase (ϕ_c) for the assorted datasets with a comparison to a normal distribution. The skew and kurtosis of the data are also shown in the figure; A Gaussian PDF has a skew of zero and a kurtosis of 3. The PDFs show Gaussian behavior within the confidence limits of the normal distribution given the number of independent samples. Note that while the PDF is Gaussian within the confidence intervals, the skew and kurtosis can be quite different from Gaussian values, since deviations in the tail of the distribution are critical. If the PDF is, in fact, Gaussian, the standard deviations of the skew and kurtosis minus three are $\sqrt{15/N}$ and $\sqrt{96/N}$, respectively (Press *et al.*, 1992). In practice, it is prudent to suspect non-Gaussianity only if the deviation of skew and kurtosis are several times these standard deviations. Thus the PDF results suggest that strong caustic structures are not dominant in the finale, and that non-Gaussian sound speed structure (and thus non-Gaussian phase perturbations), if present, have averaged, in the sense of the central limit theorem, to give Gaussian phases.

C. Phase spectra

Frequency spectra of the detrended phase ($\delta\phi_\tau$) for the data are shown in Fig. 6. All spectra in this paper were computed by first detrending and then Hanning tapering the data before applying the Fourier transform. The frequency spectra are averages over all hydrophone depths and all transmissions (l). The ten sample records are marginally adequate for a spectral analysis. All the spectra in Fig. 6 are ''red'' with no apparent low-frequency cutoff, and thus phase variance is expected to grow with observation time, as demonstrated in Table III. Further, it should be noted that there is no apparent cutoff near the buoyancy frequency, whose maximum values in this temperate region are between four and eight cycles per hour. Under conditions of multipath interference, a buoyancy frequency cutoff is not expected because the high-frequency content of the spectrum is dominated by fadeouts (Dyson *et al.*, 1976; Porter and Spindel, 1977).

The spectra are compared to the random multipath model of Dyson *et al.* (1976), who derive phase (and log-intensity) spectra for multipath signals by making assumptions about the low- and high-frequency nature of the spectra. For high frequencies (i.e., greater than ν_t), the spectra are assumed to be dominated by fadeouts in which the Cartesian components of the complex field vary linearly with time. At low frequencies (i.e., less than ν_t) the phase is assumed to execute a random walk (and thus have an unbounded variance), while the intensity is assumed to stay bounded. A kinematical model of the phase spectrum consistent with these assumptions is

$$S_\phi(\omega) \propto \frac{1}{\omega^2/\nu_t^2(\alpha+\omega^2/\nu_t^2)^{1/2}}, \tag{6}$$

where α is an empirical constant that determines where the spectrum rolls off from ω^{-3} (fade dominated) to ω^{-2} (random walk dominated). From a numerical experiment, Dyson empirically chooses $\alpha=1.27$. Figure 6 shows model spectra from Eq. (6) with different phase rates of 5, 10, and 30 rad/s.

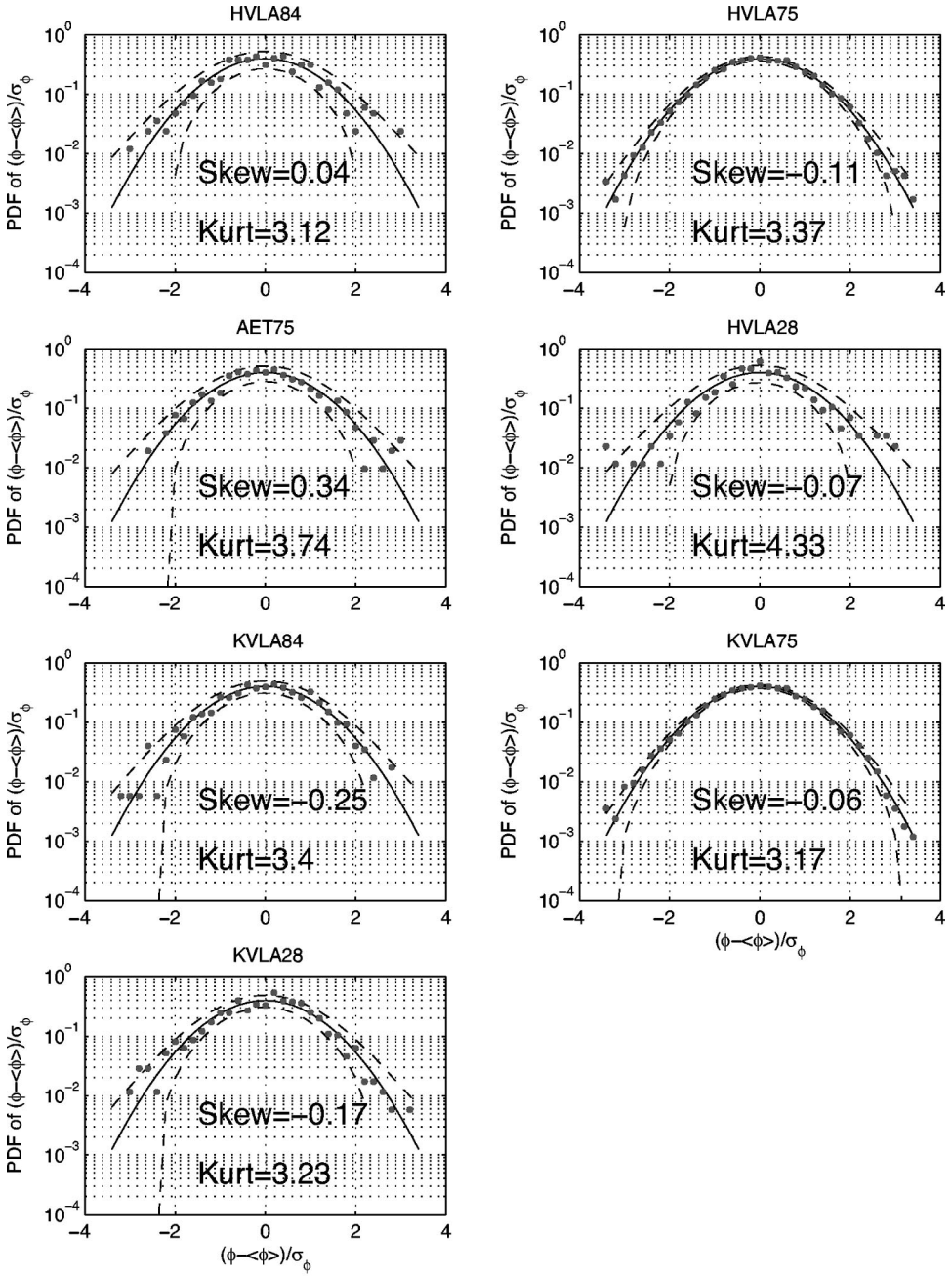


FIG. 5. The PDF of the broadband phase for the assorted experiments. The phase PDF is compared to a normal distribution (solid). The 95% confidence intervals on the normal PDF are plotted in dashed lines. The number of independent samples is obtained by using a vertical correlation length of 4 hydrophones and a temporal correlation time of 10 min. Thus, for HVLA84, HVLA75, AET75, HVLA28, KVLA84, KVLA75, and KVLA28, the number of independent samples is 420, 5880, 525, 435, 870, 8490, and 870, respectively.

The AET75 data, which should be compared to the 30 rad/s model, shows the best agreement with the model at frequencies less than about 10 cph. The high-frequency end of the AET spectrum is too flat and clearly does not fit the model. The other high acoustic frequency (HVLA and KVLA, 84 and 75 Hz) spectra, which should be compared to the model cases with phase rates of 5 and 10 rad/s, show very close to an ω^{-2} shape, but the model places them in the ω^{-3} range. Changing α to 10 (not shown) somewhat improves the comparison, particularly for the KVLA spectra. The low acoustic frequency 28 Hz spectra do not agree with the model spectra at all.

The vertical wave number spectra of detrended phases ($\delta\phi_z$) are shown in Fig. 7. The vertical wave number spectra in this paper are averaged over all times (τ) and all transmissions (l). All the spectra show an approximately k_z^{-3} shape, with a hint of a roll off at the lowest wave number. These spectra can also be compared to the Dyson model [Eq. (6)], where ω is replaced by k_z and ν_t is replaced by ν_z, and the agreement is very good. The model and observed k_z^{-3} shape implies phase variability dominated by fadeouts. Thus, the interference pattern in depth appears more developed (even at 28 Hz) than the interference pattern in time. More of a discussion on this topic will come after the spectra of log intensity are presented in Sec. IV.

IV. INTENSITY FLUCTUATIONS

The analysis now turns to the intensity, where both linear intensity ($I=|\psi_r|^2$) and log-intensity ($\iota=\ln I$) measures are used. In this section, results are presented for intensity moments, intensity PDF, and intensity frequency and wave number spectra. In Sec. IV D correlations between the intensity and the phase are treated.

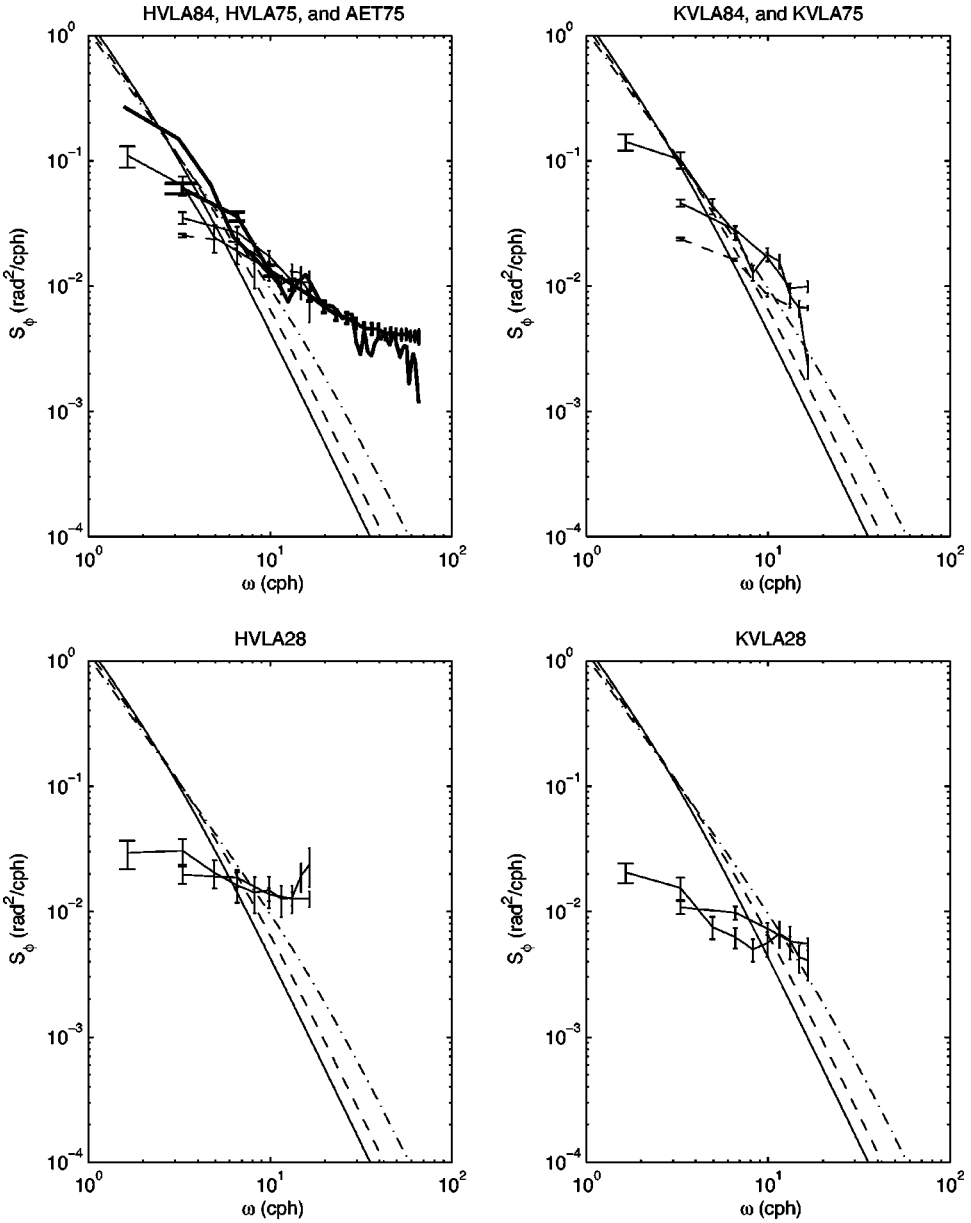


FIG. 6. Frequency spectra of the phases $\delta\phi_\tau$ for the assorted experiments. In the upper left panel the AET75 spectra are plotted with heavy lines, the HVLA84 spectra are plotted with light lines, and the HVLA75 data are plotted with dashed lines. In the upper right panel, the KVLA84 data are plotted with light lines while the KVLA75 data is plotted with dashes. The curves extending out to almost 1 cph are from the long, 40 min data. Error bars show 95% confidence intervals, except for the single transmission 40 min AET data. In all panels, the phase spectra from the Dyson model are displayed (smooth curves) with rms temporal phase rates of 5.0 (solid), 10 (dashed), and 30 (dash-dotted) rad/s. The model spectra use $\alpha=1.27$ and are normalized to give a variance of 1 $(\text{rad})^2$.

A. Intensity moments

For the examination of intensity fluctuations, the focus will be on the scintillation index (SI), and the variance of log intensity [$\iota=\ln(I)$], which are defined as

$$\text{SI}=\langle I^2\rangle/\langle I\rangle^2-1 \quad \text{and} \quad \sigma_\iota^2=\langle \iota^2\rangle-\langle \iota\rangle^2. \tag{7}$$

The SI and σ_ι^2 quantities are complementary because the former is sensitive to the peaks in the intensity distribution while the latter is sensitive to the fadeouts (due to the logarithmic distortion of small intensities). In the case of strong multipath interference, where the real and imaginary parts of ψ are independent Gaussian random variables, the well-known fully saturated results (Dyer, 1970) give SI=1 and $\sigma_\iota=\pi/6^{1/2}\rightarrow(\pi/6^{1/2})(10/\ln 10)=5.57$ dB.

In the phase analysis of Sec. III, variations between transmissions l could not be examined because there was no way of estimating the absolute phase through the time gaps, however, for intensity this problem does not exist. Intensity statistics are compiled over two time scales: One over the total duration of the experiments and the other over the 20 or 40 min observation times. Long-time scale intensity variability is estimated by computing the intensity moments ($\overline{\langle I^2\rangle}$, $\overline{\langle I\rangle}$, $\overline{\langle \iota^2\rangle}$ and $\langle \iota\rangle$) ensemble averaged over time and transmission for each depth. Then using Eqs. (7), SI and σ_ι^2 are computed and averaged over depth. Mathematically this operation is

$$\text{SI}=\langle \overline{\langle I^2\rangle}_\tau/(\overline{\langle I\rangle}_\tau)^2-1\rangle_z, \quad \sigma_\iota^2=\langle \overline{\langle \iota^2\rangle}_\tau-(\overline{\langle \iota\rangle}_\tau)^2\rangle_z. \tag{8}$$

The short time scale calculation goes along similar lines, except SI and σ_ι^2 are computed for each depth and transmission and then averaged over those variables. Mathematically this operation is

$$\text{SI}=\overline{\langle \langle I^2\rangle_\tau/\langle I\rangle_\tau^2-1\rangle_z}, \quad \sigma_\iota^2=\overline{\langle \langle \iota^2\rangle_\tau-\langle \iota\rangle_\tau^2\rangle_z}. \tag{9}$$

Finally, the depth variation of intensity is defined as

$$\text{SI}_z=\overline{\langle \langle I^2\rangle_z/\langle I\rangle_z^2-1\rangle_\tau}, \quad \sigma_{\iota_z}^2=\overline{\langle \langle \iota^2\rangle_z-\langle \iota\rangle_z^2\rangle_\tau}. \tag{10}$$

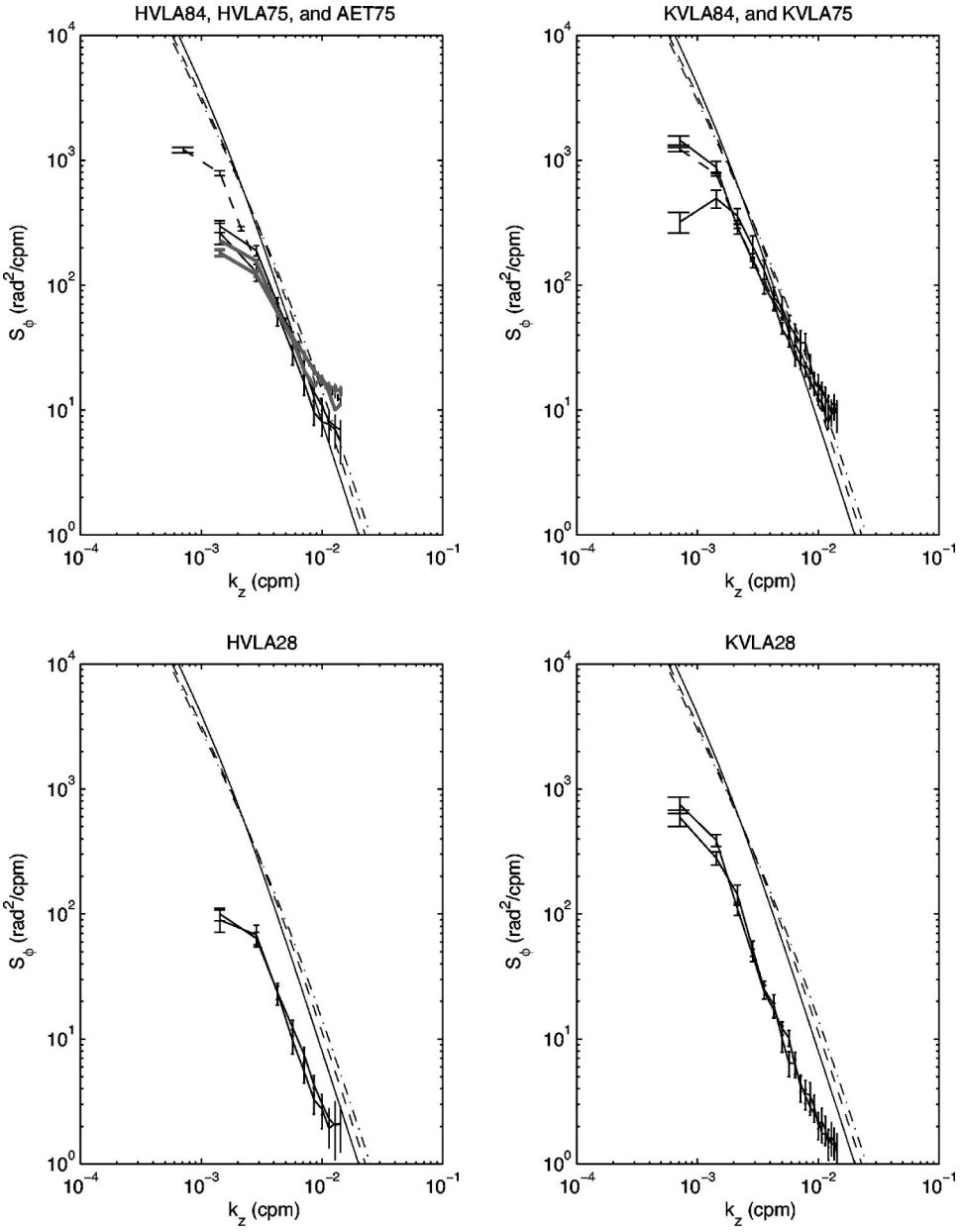


FIG. 7. Vertical wave number spectra of phases $\delta\phi_z$ for the assorted experiments. In the upper left panel the AET75 spectra are plotted with heavy lines, the HVLA84 spectra are plotted with light lines, and the HVLA75 data are plotted with dashed lines. In the upper right panel the KVLA84 data are plotted with light lines while the KVLA75 data is plotted with dashes. Error bars show 95% confidence intervals, except for the single transmission 40 min AET data. In all panels, phase spectra from the Dyson model are displayed (smooth curves) with rms vertical phase rates of 10.0 (solid), 20 (dashed), and 30 (dash-dotted) rad/km. The model spectra use $\alpha=1.27$ and are normalized to give a variance of 2 (rad)^2.

Table IV summarizes the aforementioned intensity statistics for the different experiments. In general, the results show that over the short observation times (20 and 40 min), all the experiments have statistics that are well below the fully saturated values, while the long time scale and depth directions show values quite close to full saturation. With regard to the AET, it is interesting to note that the intensity statistics are not as anomalous in comparison to the other experiments, as were the phase statistics.

The intensity statistics in the time direction (Table IV; Columns 3–6) are discussed first, where the important issues are the dependencies on range (though the paths are separated geographically), acoustic frequency, and bandwidth. Regarding the dependence on range, the two cases of short and long time scale variability are treated separately. For the well-sampled 20 min observation time (rows 1–7), short time scale variability (Columns 3 and 4), it is noted that in all cases both SI and σ_ι are well below fully saturated values; thus the expectation is that these quantities should only grow with range. However, the results show that there is very little difference between the HVLA and KVLA results at 84 and 75 Hz, and at 28 Hz the KVLA actually shows less fluctuation than the HVLA. This result is borne out again in the 40 min observation time data (Columns 3 and 4). Now, regarding the long-time scale variability (Columns 5 and 6) in all cases SI and σ_ι are above fully saturated values. In the next section concerning intensity PDFs it will be shown that the above saturation values are likely due to modulation effects on an already saturated field; thus to the degree that there is range dependence in these values depends critically on the source of modulation. Range dependence in the long time scale variability therefore cannot be easily described. More discussion on this topic will be given in Sec. VI.

Next, the dependency on acoustic frequency is discussed. For the short time scale statistics (Columns 3 and 4), the key comparisons are between the experimental pairs (HVLA84,HVLA28), and (KVLA84,KVLA28) for both the 20 and 40 min observation time. Here, again, since values

TABLE IV. Broadband and narrow band (denoted with parentheses) intensity statistics. The six columns give SI and σ_ι computed in both the depth (columns 1 and 2) and time directions (columns 3–6). Intensity fluctuations in the time direction are computed over both the short time scale [Eq. (9)], and over the total experiment (long) time scale [Eq. (8)]. Long time scale variances for the 40 min observation time could not be computed due to extremely small sample sizes. Short time scale estimates of SI and σ_ι are potentially biased due to a small sample size (N_τ) if the PDF of intensity is near exponential. For $N=10$, 20, 40, and 80, the SI(σ_ι) biases are $-0.17(-0.31)$, $-0.10(-0.16)$, $-0.05(-0.10)$, and $-0.03(-0.07)$, respectively. Sampling errors are a few percent.

	Depth	Fluctuations	Time	Fluctuations		
	SI_z	σ_{ι_z} (dB)	Short time scale SI	σ_ι (dB)	Long time scale SI	σ_ι (dB)
20 min obs. time						
HVLA84	0.80(0.79)	5.3(5.3)	0.46(0.45)	4.0(3.9)	1.35(1.49)	6.0(6.0)
HVLA75	0.97(0.95)	5.6(5.6)	0.56(0.50)	4.4(4.3)	1.24(1.17)	6.0(6.0)
AET75	0.82(0.80)	5.3(5.3)	0.64(0.56)	4.7(4.4)	1.13(1.16)	5.6(5.7)
HVLA28	0.62(0.67)	4.8(4.9)	0.47(0.39)	3.9(3.4)	1.47(1.81)	5.9(6.3)
KVLA84	0.87(0.91)	5.4(5.5)	0.44(0.46)	4.0(4.1)	1.40(1.42)	5.7(5.9)
KVLA75	0.95(0.94)	5.4(5.5)	0.50(0.48)	4.2(4.2)	1.29(1.15)	5.9(6.0)
KVLA28	0.80(0.80)	5.2(5.3)	0.29(0.25)	3.1(2.8)	1.15(1.13)	5.6(5.8)
40 min Obs. Time						
HVLA84	0.87(0.76)	5.1(5.3)	0.60(0.56)	4.7(4.6)	⋯	⋯
AET75	0.67(0.58)	5.0(4.7)	0.74(0.69)	5.1(4.9)	⋯	⋯
HVLA28	0.55(0.54)	4.4(4.2)	0.70(0.44)	4.4(3.7)	⋯	⋯
KVLA84	0.81(0.81)	5.3(5.3)	0.56(0.60)	4.7(4.8)	⋯	⋯
KVLA28	0.70(0.76)	5.4(5.4)	0.42(0.29)	3.8(3.2)	⋯	⋯

are below saturation values the lower frequency is expected to have less fluctuation. Curiously, here the HVLA data show almost no sensitivity to center frequency, while the KVLA data show clearly less fluctuation at the lower frequency. This result is consistent with the previous phase result, which showed that the KVLA phase variance dropped much more rapidly as a function of center frequency than the HVLA (Sec III). For the long time scale variability (Columns 5 and 6), as previously mentioned, modulation processes will be shown to be important, thus as in the range case variation of the SI and σ_ι as a function of frequency cannot be easily described.

Finally, an inspection of Table IV shows, as in the case of phase, that there is very little difference between broadband and narrow band results.

The last topic to be treated in this section is the intensity variability in the depth direction, which cuts across the complex finale interference pattern (see Fig. 2). For the 84 and 75 Hz center frequency cases the SI_z and $\sigma^2_{\iota_z}$ values are very close, but slightly under the full saturation values. The 28 Hz data, on the other hand, show generally smaller fluctuations with the HVLA data showing significantly smaller values compared to the KVLA.

B. Intensity PDF

Figure 8 shows the PDFs of intensity for the data, and a comparison to the exponential distribution. The displayed PDFs are calculated by binning the 20 min observation time intensity data over all transmissions *l*. Thus the SI and σ^2_ι values of these PDFs are given in columns 5 and 6 of Table IV. The displayed PDFs are very similar to previously published PDFs for the AET (Colosi, Tappert, and Dzieciuch, 2001). There are significant deviations from the exponential distribution, primarily at the high intensity end of the PDF, so that SI and σ^2_ι are larger than the true exponential PDF moments. Also shown in Fig. 8 is the modulated exponential (ME) PDF model of Colosi, Tappert, and Dzieciuch (2001), given by

$$P(I)=\frac{\exp(-I/\langle I\rangle)}{\langle I\rangle}\left[1+\frac{b}{2}\left(\frac{I^2}{\langle I\rangle^2}-4\frac{I}{\langle I\rangle}+2\right)\right], \qquad (11)$$

where the modulation parameter *b* is obtained from the scintillation index, $SI=1+2b$. The physical model here is of a saturated process where the mean intensity has a Gaussian modulation. In all cases the ME PDF fits the data very well at both low and high intensities. Thus, the conclusion is that the long time scale behavior of the observations (including the 28 Hz data) are very close to the fully saturated regime and the effect of the multiday or multimonth observation time is to impose a modulation on the saturated fluctuations. An important yet unanswered question is what ocean process is the source of the intensity modulation. Interestingly the fluctuations are not saturated at the 40 min time scale and thus there is an ''unmeasured'' time scale region between the 40 min and multiday time scales over which the fluctuations become saturated. Continuous acoustic measurements over several days or weeks would help to address these two issues, and measurements at shorter ranges would elucidate the approach to saturation. More on these issues comes in Sec. VI.

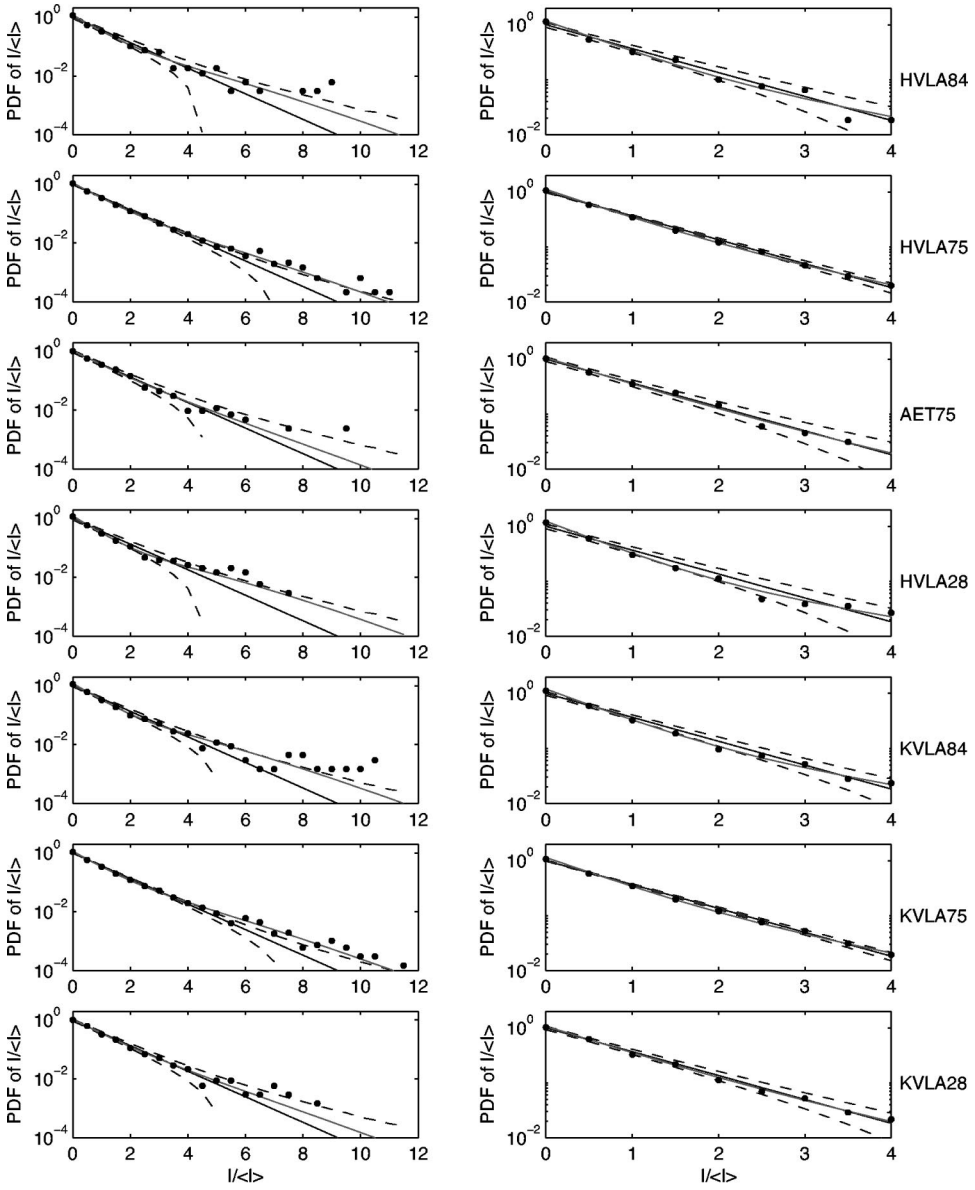


FIG. 8. The PDF of broadband intensity for the assorted experiments. The intensity PDF is compared to a pure exponential distribution (straight–solid). The 95% confidence intervals for the exponential PDF are plotted in dashed. Also plotted are estimates of the Modulated Exponential PDF (curved solid lines), where the modulation parameter $b=0.5\,(\mathrm{SI}-1)$. The right-hand panels show an expanded view of the small intensity region. The number of independent samples is obtained by using a vertical correlation length of 2 hydrophones and a temporal correlation time of 10 min. Thus, for HVLA84, HVLA75, AET75, HVLA28, KVLA84, KVLA75, and KVLA28, the number of independent samples is 840, 11760, 1050, 870, 1740, 16980, and 1740, respectively.

C. Log-intensity spectra

Frequency spectra of log intensity for the various datasets are shown in Fig. 9. As previously stated, all spectra in this paper were computed by first detrending and then Hanning tapering the data before applying the Fourier transform. The frequency spectra are averages over all hydrophone depths and all transmissions (l). These log-intensity spectra, unlike the phase spectra, show a variety of shapes: The HVLA84, HVLA75, and AET75 data show a very flat spectrum at high frequency and sharply increasing values at the lowest frequencies. By contrast, the KVLA84 and KVLA75 data show a slope between ω^{-2} and ω^{-3} across the entire frequency band. The 28 Hz spectra are flat for both the HVLA and the KVLA. These relatively ''flat'' spectra demonstrate why the intensity statistics in Table IV do not grow rapidly with increasing observation time. As noted previously with regard to the phase spectra, there is no high-frequency cutoff near the buoyancy frequency as expected for a multipath interference pattern.

These spectra, like those of phase, are compared to the random multipath model of Dyson *et al.* (1976). Using the assumptions discussed in Sec. III, the log-intensity spectrum of Dyson is

$$S_\iota(\omega)\propto\frac{1}{(\beta+\omega^2/\nu_t^2)^{3/2}}, \tag{12}$$

where β is an empirical constant that determines where the spectrum rolls off from ω^{-3} (fade dominated) to ω^0 (random walk dominated). From a numerical experiment, Dyson empirically chooses $\beta=2.43$. Figure 9 shows model spectra from Eq. (12) with different phase rates of 5, 10, and 30 (rad/s). As in the case of phase, the 28 Hz acoustical data spectra do not match the model at all, however, the higher acoustic frequency data do not show good agreement either [recall that AET75 spectra should be compared to the model with 30 rad/s phase rate (dash–dot), while the other spectra are to be compared to the other two models]. The best comparison is with the KVLA data, where the high-frequency

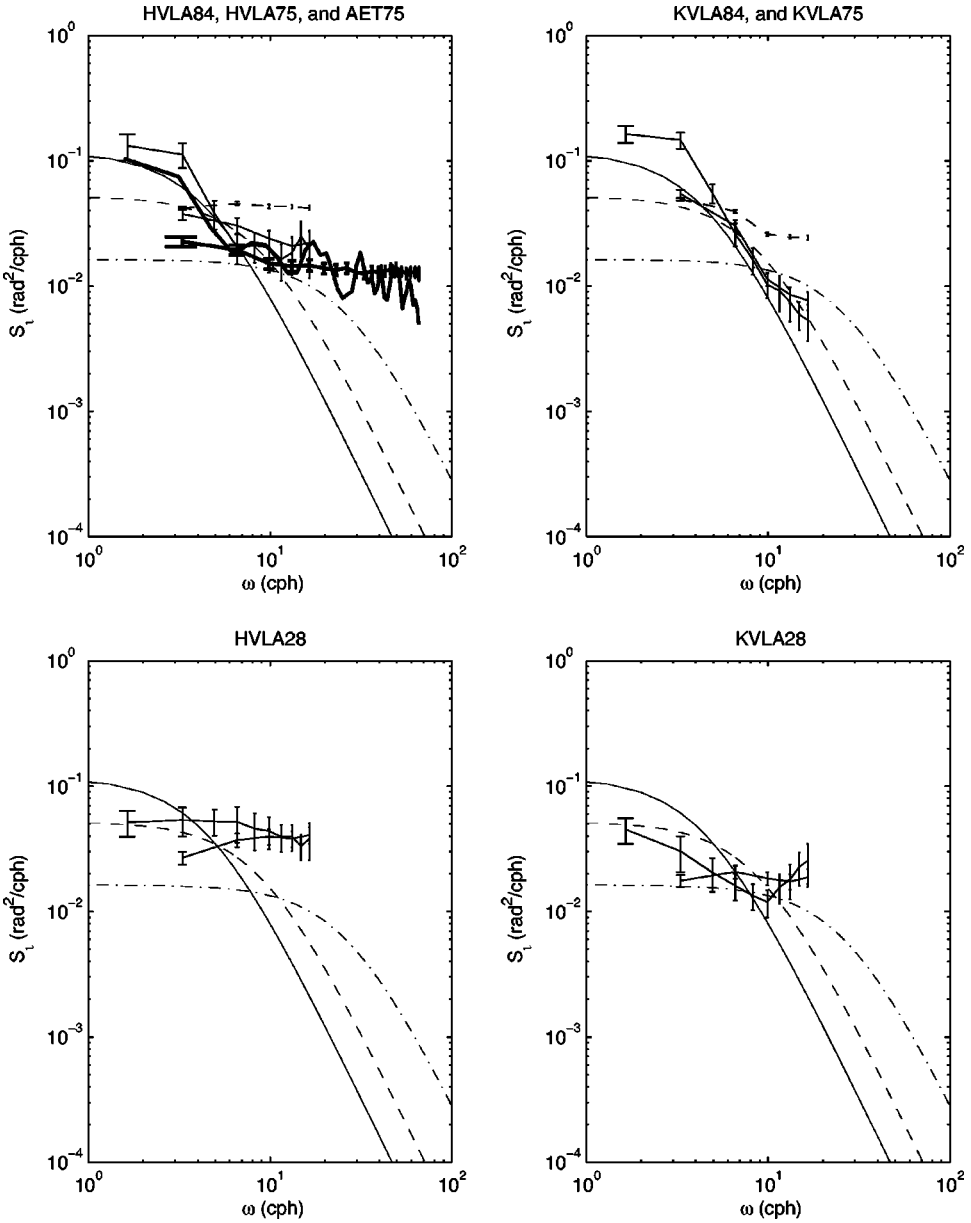


FIG. 9. Frequency spectra of log intensity ι. In the upper left panel the AET75 spectra are plotted with heavy lines, the HVLA84 spectra are plotted with light lines, and the HVLA75 data are plotted with dashed lines. In the upper right panel the KVLA84 data are plotted with light lines while the KVLA75 data is plotted with dashes. The curves extending out to almost 1 cph are from the long, 40 min data. Error bars show 95% confidence intervals, except for the single transmission 40 min AET data. In all panels, the log intensity spectra from the Dyson model are displayed (smooth curves) with temporal phase rates of 5.0 (solid), 10 (dashed), and 30 (dash–dotted) rad/s. The model spectra use $\beta=2.43$ and are normalized to give a variance of $\pi^2/24$.

ω^{-3} shape and the low-frequency rolloff are almost borne out.

Vertical wave number spectra of log intensity are shown in Fig. 10. These vertical spectra are averaged over all times (τ) and all transmissions (l). As discussed in Sec. III, the vertical spectra can also be interpreted in terms of the Dyson model [Eq. (12)] by substituting k_z and ν_z for ω and ν_t. At 28 Hz acoustical frequency, the observations should be compared to the Dyson model with $\nu_z=10$ rad/km (solid) and the agreement is very good. The k_z^{-3} dependence implies a pattern dominated by fadeouts. At the higher acoustic frequencies, which should be compared to the model with $\nu_z=20$ and 30 rad/km, the roll-off between fade-out statistics and (phase) random walking is clearly seen. Again, these results imply a more randomized interference pattern in depth compared to time.

D. Phase log-intensity correlation

Table V shows correlations between ι and ϕ_c, and their rates of change. To a very good approximation, phase and log intensity are uncorrelated. Further, the correlation between absolute values of rates of change are very close to $2/\pi=0.63$, which is the predicted saturation result [Dyson *et al.* (1976)]. The phase/log-intensity correlation will be important when considering the theoretical form of the coherence function that will be discussed in the next section.

V. COHERENCE

In Secs. III and IV, phase and intensity statistics were analyzed and this provides a natural setting to discuss the simplest second moment that involves both the phase and intensity simultaneously, namely the signal coherence, $\langle\psi\psi^*(\Delta)\rangle$. The coherence is an important function since it describes a coherent processing gain,

$$\mathrm{PG}=\left\langle\left|\sum_{n=1}^{N_p}\psi_n\right|^2\right\rangle\Big/\left\langle\sum_{n=1}^{N_p}|\psi_n|^2\right\rangle$$

$$=1+\frac{2}{N_p}\sum_{n=1}^{N_p}(N_p-n)\left|\langle\psi\psi^*(\Delta_n)\rangle\right|, \tag{13}$$

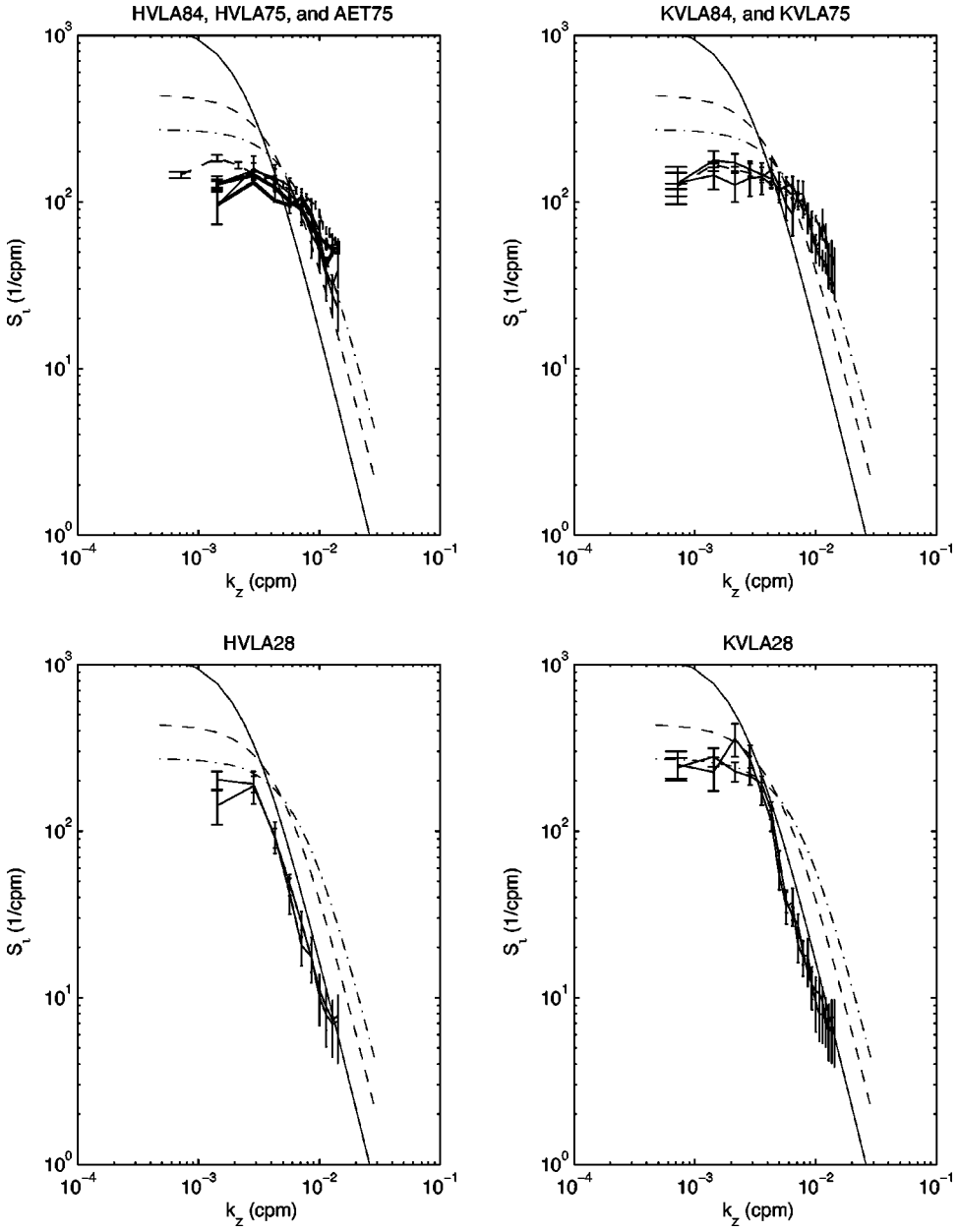


FIG. 10. Vertical wave number spectra of log intensity ι. In the upper left panel the AET75 spectra are plotted with heavy lines, the HVLA84 spectra are plotted with light lines, and the HVLA75 data are plotted with dashed lines. In the upper right panel the KVLA84 data are plotted in solid while the KVLA75 data is plotted with dashed lines. Error bars show 95% confidence intervals, except for the single transmission 40 min AET data. In all panels log-intensity spectra from the Dyson model are displayed (smooth curves) with vertical phase rates of 10 (solid), 20 (dash), and 30 (dash–dotted) rad/km. The model spectra use $\beta=2.43$ and are normalized to give a variance of $\pi^2/6$.

where N_p is the number of fields summed, and Δ_n is the nth lag. The coherence function also can help define ''optimal'' signal processing procedures Further, the contrast between time coherences of resolved normal mode arrivals [Wage *et al.* (2005)] and the full field coherences discussed here will give important information concerning correlations between modes.

In the present analysis of coherence, the wave functions are defined so that the coherence function represents the effects of ''nonlinear'' signal variability. What this means is that the detrended phase functions $\delta\phi_\tau$ and $\delta\phi_z$ are utilized and the following wave functions are defined:

$$\psi_\tau(z,\tau,l)=|\psi_r(z,\tau,l)|\mathrm{e}^{-i\,\delta\phi_\tau(z,\tau,l)},$$

TABLE V. Broadband and narrow band (denoted with parentheses) intensity-phase correlations for the 20 min observation time.

	Time correlation $\overline{\lvert\langle\iota_\tau\delta\phi_\tau\rangle_{z\tau}\rvert}/(\overline{\langle\delta\phi_\tau^2\rangle_{z\tau}\langle\iota_\tau^2\rangle_{z\tau}})^{1/2}$	Depth correlation $\overline{\lvert\langle\iota_z\delta\phi_z\rangle_{z\tau}\rvert}/(\overline{\langle\delta\phi_z^2\rangle_{z\tau}\langle\iota_z^2\rangle_{z\tau}})^{1/2}$	Rate correlation $\overline{\langle\lvert\partial_\tau\iota\rvert\lvert\partial_\tau\delta\phi_\tau\rvert\rangle_{z\tau}}/(\overline{\langle(\partial_\tau\delta\phi_\tau)^2\rangle_\tau\langle(\partial_\tau\iota)^2\rangle_\tau})^{1/2}$
HVLA84	0.0381(0.0439)	0.0186(0.0137)	0.74(0.74)
HVLA75	0.0001(0.0004)	0.0172(0.0023)	0.68(0.69)
AET75	0.0012(0.0084)	0.0029(0.0034)	0.68(0.68)
HVLA28	0.0088(0.0063)	0.0053(0.0077)	0.68(0.68)
KVLA84	0.0054(0.0021)	0.0162(0.0012)	0.74(0.75)
KVLA75	0.0002(0.0015)	0.0014(0.0200)	0.69(0.70)
KVLA28	0.0079(0.0020)	0.0857(0.0847)	0.70(0.71)

$$\psi_z(z,\tau,l) = |\psi_r(z,\tau,l)| e^{-i\,\delta\phi_z(z,\tau,l)}, \tag{14}$$

for the analysis of coherence. The effect of removing the phase trends is most pronounced on the time coherence, since a larger portion of the phase variance in time is associated with the trend (see Table III). The removal of the phase trend in time will account for any depth-dependent mooring motion, and it will also remove some true, low-frequency, ocean-induced variability. Again, the philosophy is to compute coherences that represent the ''nonlinear'' effects of signal variability. With this in mind, the coherence functions are defined as

$$|\langle \psi\psi^*(\delta\tau)\rangle| = |\overline{\langle \psi_\tau \psi_\tau^*(\delta\tau)\rangle_{z\tau}}|, \quad \text{and}$$

$$|\langle \psi\psi^*(\delta z)\rangle| = |\overline{\langle \psi_z \psi_z^*(\delta z)\rangle_{z\tau}}|, \tag{15}$$

where $\delta\tau$ and δz are time and depth lags, and it is understood that the absolute value is taken after ensemble averaging over all appropriately lagged pairs of the τ, z, and l parameters.

An important aspect of the coherence function is the relation to the mean field. In the limit of large Δ where ψ and $\psi^*(\Delta)$ are uncorrelated, $\langle \psi\psi^*(\Delta\to\infty)\rangle = |\langle\psi\rangle|^2$, where the right-hand side of the equation is the mean field energy. In terms of the present calculations the mean field energies are expressed as

$$|\langle\psi\rangle|_\tau^2 = |\overline{\langle\psi_\tau\rangle_{z\tau}}|^2, \quad \text{and} \quad |\langle\psi\rangle|_z^2 = |\overline{\langle\psi_z\rangle_{z\tau}}|^2. \tag{16}$$

Because of the unit intensity normalization [Eq. (2)], Eqs. (16) give the fraction of energy in the mean fields (i.e., a value of $|\langle\psi\rangle|^2 = 0.5$ means half the total energy is in the mean field).

A theoretical form of the coherence function consistent with the observations is sought. From Secs. III and IV it was demonstrated that phase has a Gaussian PDF and that the log-intensity-phase correlations are very small. A model consistent with these findings is [see Appendix C, Eq. (C10)]

$$\langle \psi\psi^*(\Delta)\rangle = \langle AA(\Delta)\rangle \exp[-D(\Delta)/2], \tag{17}$$

where $D(\Delta) = \langle[\phi - \phi(\Delta)]^2\rangle$ is the phase structure function (Flatté *et al.*, 1979), and $\langle AA(\Delta)\rangle$ is the amplitude covariance. A unit mean intensity normalization has been chosen so that the coherence function has value 1 at zero lag. This form differs from the traditional form by retaining the amplitude covariance term (Reynolds *et al.*, 1985; Flatté and Stoughton, 1988). A further approximation that is traditionally made for a single acoustical path is that the phase structure function behaves quadratically near zero lag, which leads to the vertical coherence length ($z_0 \sim \nu_z^{-1}$) and coherence time ($t_0 \sim \nu_t^{-1}$) defined by Flatté and Stoughton (1988). Consistent with geometrical approximations made for this single path, the coherence lengths/times scale as $(\text{frequency})^{-1}$ and $(\text{range})^{-1/2}$.

Thus, turning to the model for coherence, the amplitude covariance, the phase structure function, and the total coherence can be computed separately to examine the relative contributions between phase and amplitude. In all of the foregoing calculations broadband data are used, as essentially identical results are obtained with the narrow band data.

A. Amplitude covariance

Figures 11 and 12 show the time- and depth-lagged covariance functions of amplitude for the various data. Since amplitude is a positive definite quantity, the covariance function will attain the value $\langle a\rangle^2$ when the correlation is zero. Thus, the figures also plot the mean amplitude squared. In full saturation where amplitude has a Rayleigh PDF, $\langle a^2\rangle = 1$ and $\langle a\rangle^2 = \pi/4$, thus the amplitude covariance will decay from 1 to $\pi/4 \sim 0.79$.

In all cases, except KVLA84, the time-lagged amplitude covariance decays rapidly between the zero and first lag, followed by a smoother decorrelation toward $\langle a\rangle^2$. In some cases like the AET75, HVLA75, and HVLA28 data, the rapid decorrelation is dramatic. The rapid decorrelation is not due to noise, since the signal-to-noise ratios, save the AET, are quite good, as discussed in Sec. II. The depth-lagged covariances also show rapid decorrelation, but only at the higher acoustic frequencies of 84 and 75 Hz. The high-frequency cases rapidly decorrelate toward $\langle a\rangle^2$, but the low-frequency cases show more vertical structure (as is evident from Fig. 3).

B. Phase structure function

Figures 13 and 14 show the time- and depth-lagged phase structure functions for the detrended phase. Expressing D in terms of the phase correlation function is helpful, yielding $D(\Delta) = 2(\langle\phi^2\rangle - \langle\phi\phi(\Delta)\rangle)$. Thus, also plotted on these figures is twice the phase variance ($2\overline{\langle\delta\phi_\tau\rangle_{z\tau}}$, and $2\overline{\langle\delta\phi_z\rangle_{z\tau}}$). Regions where D is larger than this value indicate a negative phase correlation, a ubiquitous feature of nearly all the structure functions. Another common feature of the structure functions is a return to a positive correlation at longer time and depth lags. This somewhat counterintuitive result occurs in the time-lagged statistics because our observation times (20 and 40 min) are short compared to the time scales of the important ocean variability (i.e., internal waves); thus the phase records have an energetic low-frequency component (Fig. 6) that manifests itself by an apparent recorrelation. In depth, the effect is due to the vertical interference pattern that produces a roughly k_z^{-3} spectrum, as discussed in Sec. III.

In addition, there is a hint of rapid phase decorrelation from the zeroth to the first temporal lag, as was observed in the amplitude covariance functions. Interestingly there is no such rapid decorrelation in the depth functions. Finally, perhaps because of inadequate sampling or other reasons, the structure functions, save the HVLA28 and KVLA28 depth coherences, do not demonstrate the quadratic behavior near zero lag expected from the single path theory discussed above. Again, this result is not so surprising given the complex interference pattern.

C. Coherence functions

Figures 15 and 16 show the time- and depth-lagged coherence functions for the assorted datasets. Coherence estimates using Eq. (17) with the observed amplitude covariance and phase structure function are also plotted in the figures. In addition, the mean field energies are plotted in the figure and

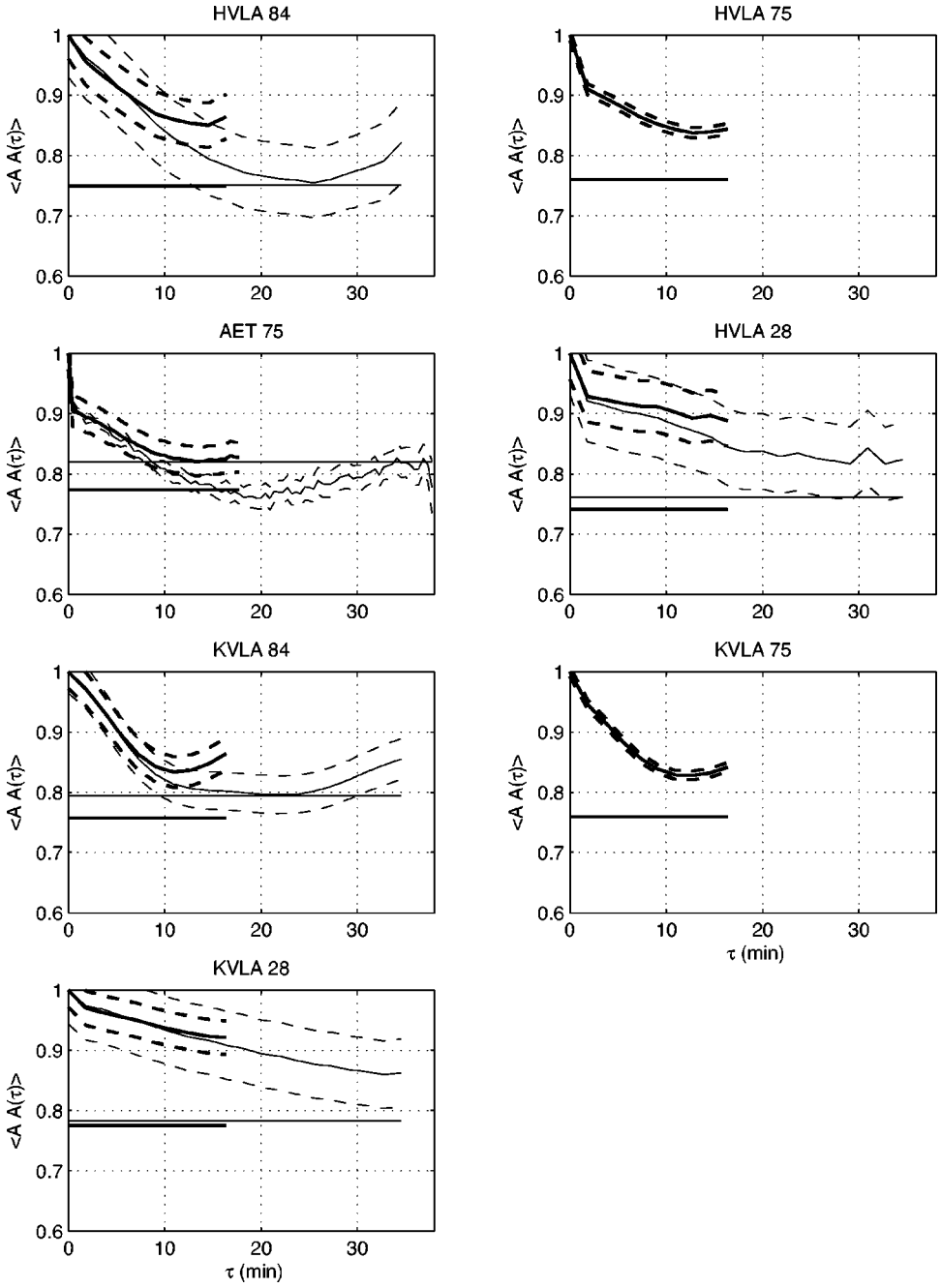


FIG. 11. Time covariance of amplitude for assorted experiments. Light curves denote the longer, 40 min transmissions, while heavy curves are for the 20 min transmissions. Dashed curves are one standard deviation uncertainties in the covariance estimates. Mean amplitude squared, $\langle A \rangle^2$, values are denoted with solid horizontal lines.

recorded in Table VI. In all cases, Eq. (17) does a very good job of describing the coherence function. All coherences, except for the HVLA28 and KVLA28 depth coherences, show an initial rapid drop in coherence due to the initial rapid drops in phase and amplitude covariance as described in the previous sections. After this initial drop there is a smoother decoherence followed by some recovery of coherence at longer lags, attributed to the phase recorrelation caused by low-frequency or low wave number variability.

With regard to temporal coherence (Fig. 15), the functions never decorrelate to zero and the mean field energies are a significant portion of the total energy. It is expected that over longer observation times the coherence functions would approach zero due to increasing phase variability. Thus, the 40 min observation times show less coherence than the 20 min observation times, having sampled more of the ocean's wideband phase variability. The 28 Hz acoustic data is extremely coherent in time, particularly the KVLA data. These results imply that temporal processing gain should be very good, both because of the coherence functions and because of the large mean field energy (as discussed by Worcester *et al.*, 1999).

With regard to the depth coherence (Fig. 16), the results suggest that even in the depth direction there is some significant coherence. The mean field energies here are not consistent with uncorrelated Gaussian random noise (see the Table VI legend). As in the temporal case, the lower, 28 Hz acoustic data shows more coherence, but here it is the HVLA data that is most impressive. The depth coherence is a surprising result given the complex interference pattern seen in the finale (Fig. 2).

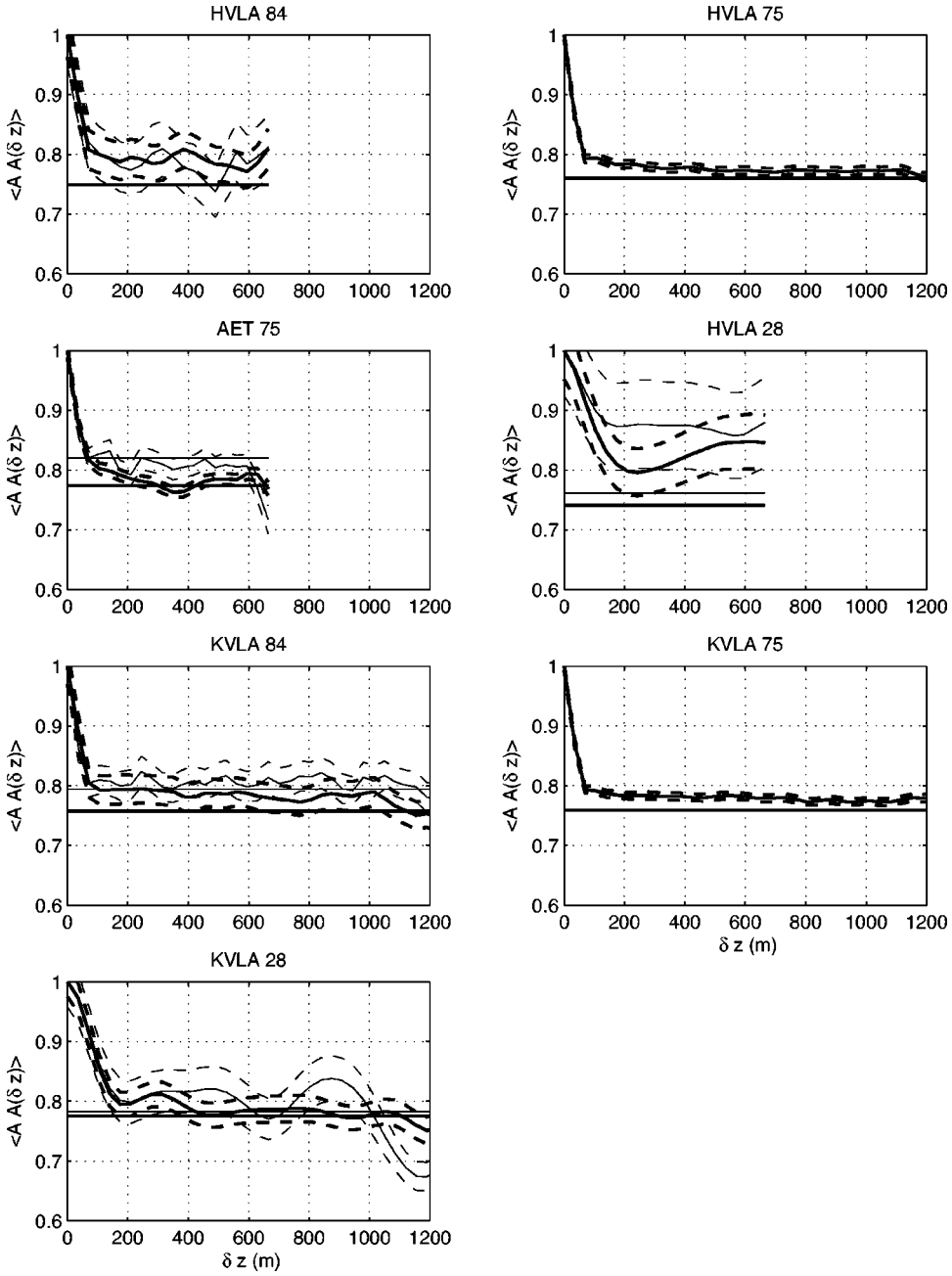


FIG. 12. Depth covariance of amplitude for assorted experiments. Light curves denote the longer, 40 min transmissions, while heavy curves are for the 20 min transmissions. Dashed curves are one standard deviation uncertainties in the covariance estimates. Mean amplitude squared, $\langle A\rangle^2$, values are denoted with solid horizontal lines.

D. Mean field energy

For the discussion of the range, frequency, observation time, and bandwidth dependence of the coherence function, it proves useful to use the mean field energy, which is a single value instead of a function. This approach is also justified since a direct comparison with theoretical predictions of coherence time and coherence depth (Flatté and Stoughton, 1988) depend on a quadratic form to the structure function that is not observed. Along the lines of Eq. (13), the mean field energy is related to the coherence function by

$$|\langle\psi\rangle|^2=\frac{\langle I\rangle}{N_p}+\frac{2}{N_p^2}\sum_{n=1}^{N_p}(N_p-n)|\langle\psi\psi^*(\Delta_n)\rangle|, \tag{18}$$

where, in this case, N_p is equal to N_τ, or N_z depending on whether the mean field is in the time or depth direction. For totally uncorrelated data, $|\langle\psi\rangle|^2=\langle I\rangle/N_p$, while $|\langle\psi\rangle|^2=\langle I\rangle$ for totally correlated data, which follows from the fact that $\Sigma_{n=1}^{N}(N-n)=N(N-1)/2$.

Table VI shows the mean field energies in both the time and depth directions for the assorted data, and the initial focus is on the time direction. Regarding the range dependence, the well-sampled 20 min observation time series (Table VI; Row 2) can be examined first. A feature of note is that the AET75 data, while at the shortest range has the smallest mean field; this is consistent with results in Secs. III and IV, which show the AET to have a much larger fluctuation than the other experiments. Next, the experimental pairs (HVLA84,KVLA84), (HVLA75,KVLA75), and (HVLA28, KVLA28) are compared, and there is essentially no significant range dependence. This conclusion is also supported by the poorly sampled 40 min observation time data.

The issue of frequency dependence is also best treated

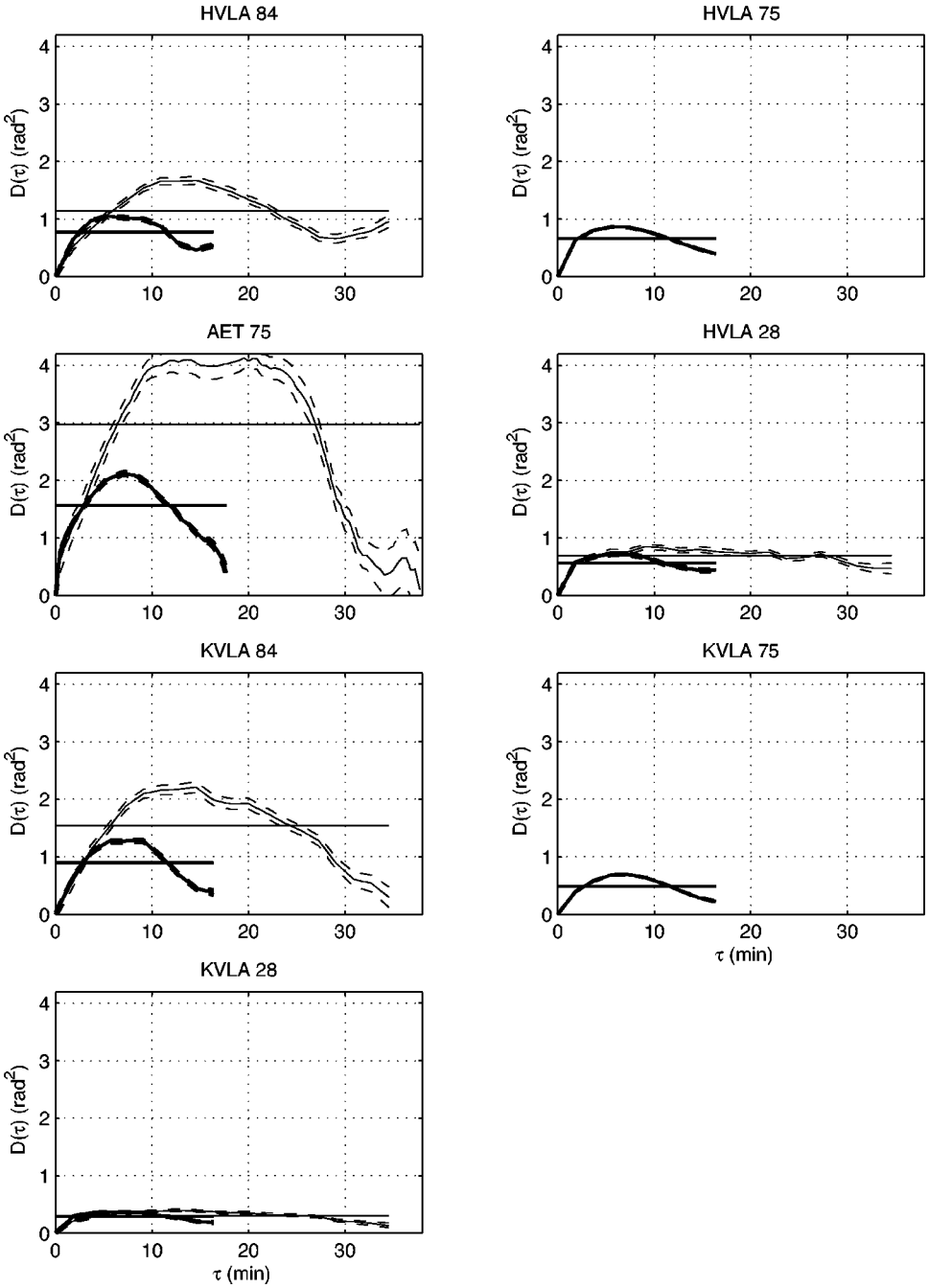


FIG. 13. The phase structure function as a function of time lag. Light curves denote the longer, 40 min transmissions while heavy curves are for the 20 min transmissions. Dashed curves are one standard deviation uncertainties in the structure function estimates. Twice the phase variances are denoted by solid horizontal lines.

by examination of the well-sampled 20 min observation time data. Here the pairs (HVLA84,HVLA28) and (KVLA84, KVLA28) show a 20%–40% increase in the mean field energy from 84 to 28 Hz. The dependence on the observation time is interesting, because the high and low acoustic frequency data show very different behavior. For the 84 Hz data the mean field energy drops by 20% to 40%, while the 28 Hz data shows virtually no dependence on observation time. The AET75 data has only one long transmission and is excluded from the discussion.

Next, the mean fields in the depth direction are discussed, with regard to the aforementioned dependencies. The range dependence of the depth mean fields shows some irregularity, as the HVLA75 and KVLA75 data show almost no decrease, while the other data pairs (HVLA84,KVLA84) and (HVLA28,KVLA28) show a significant drop in the mean field energy with range. We surmise that the 75 Hz data shows no drop because the mean field energy is already small at the HVLA. The frequency dependence of the depth mean fields is as expected with the 28 Hz data, showing significantly larger values.

As in all previous discussions concerning bandwidth, Table VI (both time and depth directions) shows no clear dependence on bandwidth.

VI. DISCUSSION

In this section some of the major results of this paper are discussed and expanded upon. Opportunities for future observational and theoretical work will illustrate the discussion.

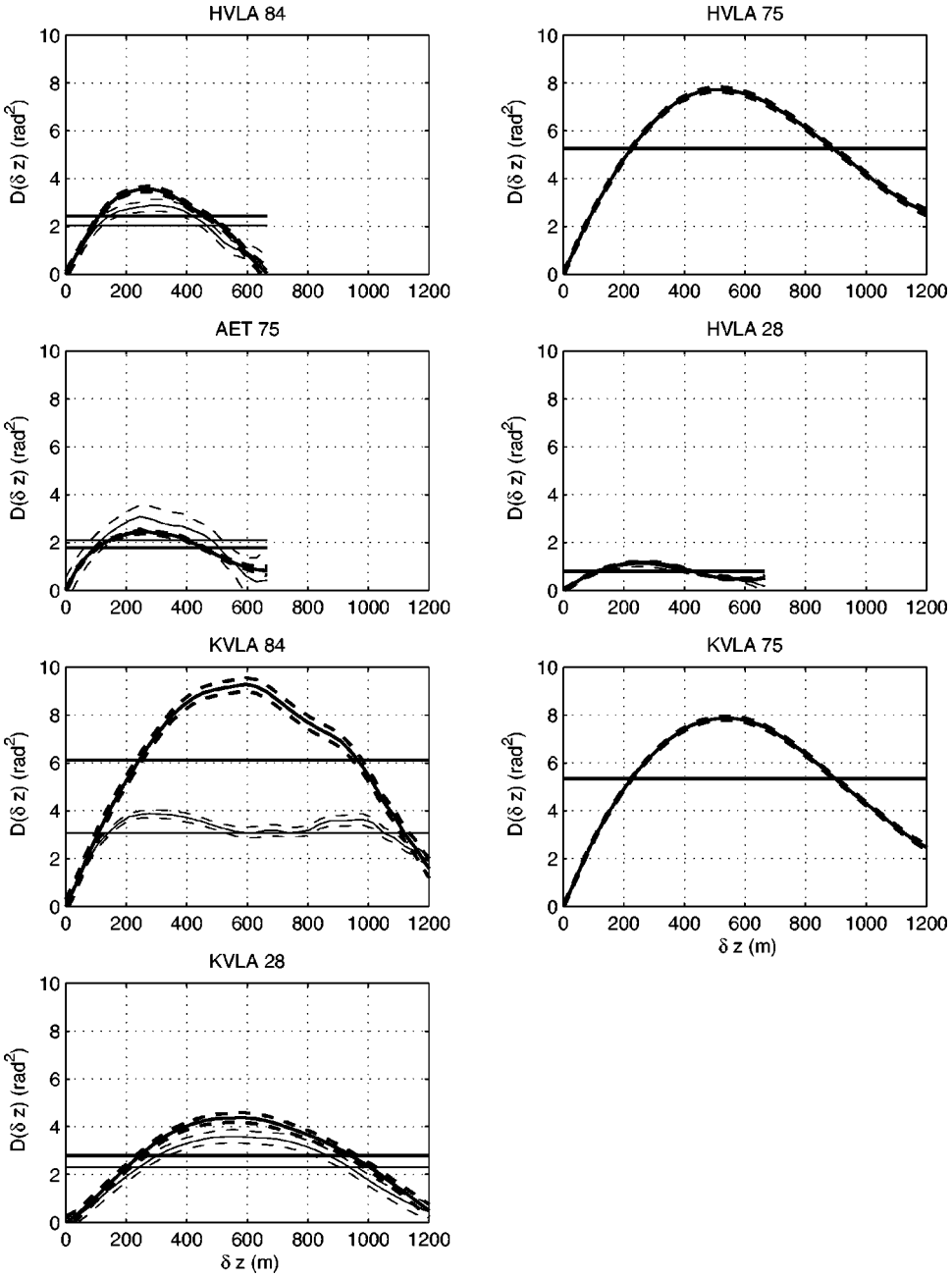


FIG. 14. The phase structure function as a function of depth lag. Light curves denote the longer, 40 min transmissions while heavy curves are for the 20 min transmissions. Dashed curves are one standard deviation uncertainties in the structure function estimates. Twice the phase variances are denoted by solid horizontal lines.

A. Range/geographical dependence of fluctuations

Range scaling of acoustic fluctuation quantities is an important component of ocean acoustic wave propagation study. The acoustic propagation paths used in this analysis (Fig. 1) allows a comparison of a few different propagation ranges, which were generally in the same geographic area (i.e., the Eastern North Pacific Ocean). To observe range scaling, one would ideally transmit sound over a single propagation path in which signals were received at various intermediate points along the path. Propagation through climatologically different regions of the ocean or across fronts should be minimized; thus east–west propagation is generally desired. In this study, however, none of these idealized criteria are met and ambiguities exist. For example, the shortest propagation path (3252 km), the AET often showed larger fluctuations than any of the longer paths at a similar frequency (i.e., greater phase and intensity variances, and lower coherence). With regard to the HVLA and KVLA paths (3500 and 5100 km) in all observables it was the longer, KVLA, path that showed equal or smaller fluctuations compared to the HVLA. On a note of consistency, however, it is found that the 84 Hz AST and 75 Hz ATOC95 transmissions on the same propagation paths show very similar fluctuations. Thus, the conclusion is that the HVLA and KVLA path data are different because they pass through significantly different oceanographic regions, namely the KVLA path samples the perhaps more benign equatorial and tropical regions.

Regarding differences between AET and HVLA 84 and 75 Hz results, which have similar propagation paths, there are several hypotheses. First, the differences may be due to the large source motion that occurred during the AET, however HVLA AST results were also from a ship suspended source that experienced motion. Further, and as previously

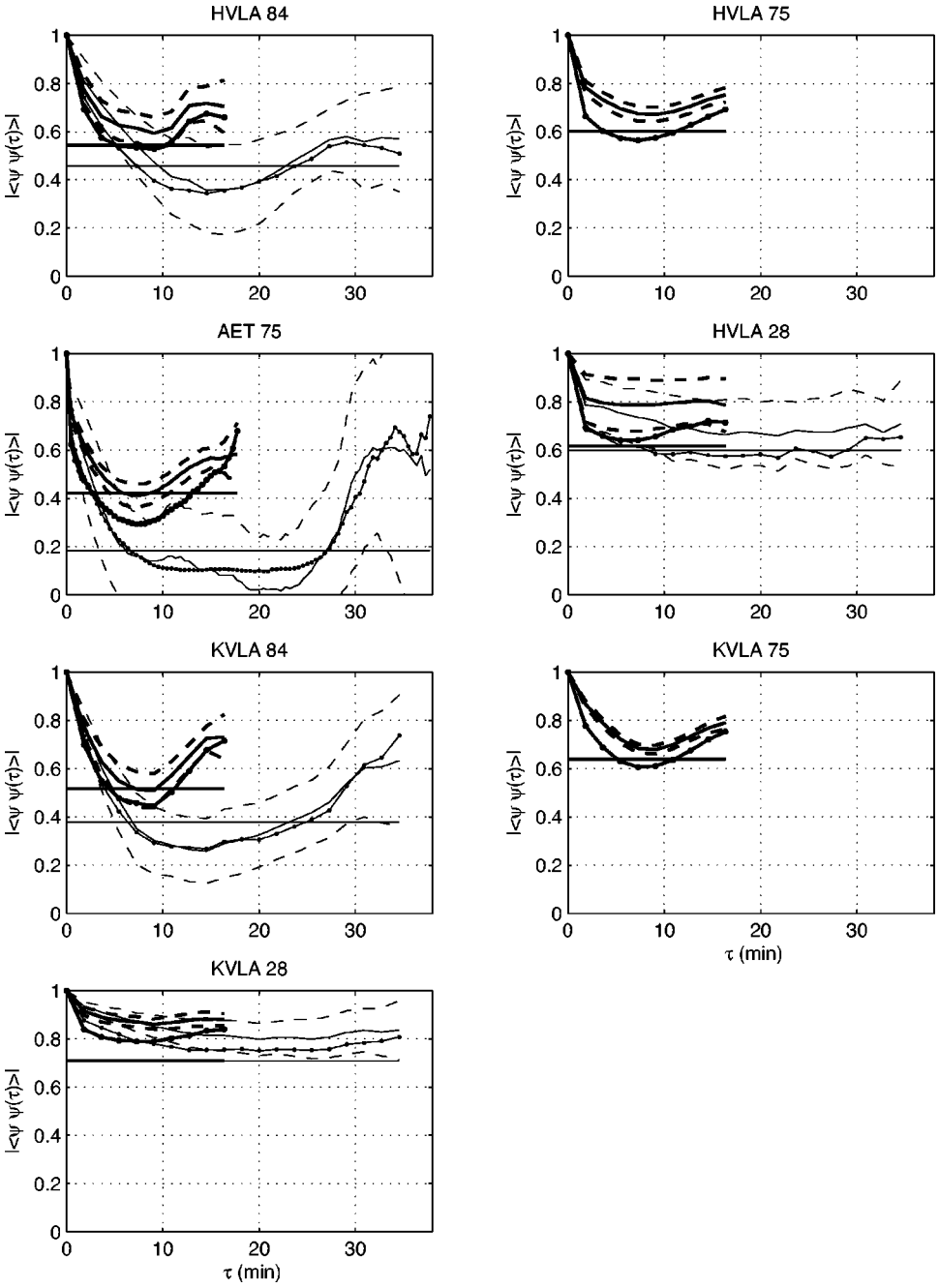


FIG. 15. Time coherence for the assorted experiments. Light curves denote the longer, 40 min transmissions, while heavy curves are for the 20 min transmissions. Dashed curves are one standard deviation uncertainties in the coherence estimates. Dotted curves denote the coherence computed using $\langle aa(\tau)\rangle \exp[-D(\tau)/2]$. The square of the mean field estimates, $|\overline{\langle \psi_\tau \rangle_{z\tau}}|^2$, are denoted by solid horizontal lines.

mentioned, the AST results are consistent with the ATOC95 results in which the source was mounted on the seabed. The differences may also be due to seasonal variability of acoustic propagation; AET was conducted in the fall, while HVLA data were collected in the winter and spring. But, again, AST results collected in the late spring/summer were consistent with ATOC95 HVLA data collected in the winter and spring. Finally there may, in fact, be a sensitivity to the geographic separation of the paths, though there are less obvious reasons to suspect this than with the HVLA, KVLA comparison.

B. Center frequency dependence of fluctuations

Another fundamental question in wave propagation studies is the scaling of fluctuations with acoustic center frequency, and thus we go from comparing different transmission paths to comparing center frequency differences along a given path. This is an area where there is very little theoretical knowledge, beyond results that are primarily valid at high frequencies and short range (Flatté *et al.*, 1979; Flatté 1983; Henyey and Macaskill, 1996). Further, theoretical results focus on the fluctuations near a deterministic geometrical acoustics ray path, not the interference pattern of many such rays. Briefly, phase is expected to be sensitive to the largest scales of the sound speed fluctuations, while intensity is sensitive to smaller scales near the first Fresnel zone. Thus at high enough frequencies, weak fluctuation theory predicts that phase variance scales like the square of frequency and log-intensity variance and SI scale linearly with frequency. The Rytov approximation provides some guidance here. The weighting function on the sound-speed fluctuation spectrum for phase variance is unity for large scales, and, on average, 1/2 for smaller scales. The weighting function for log ampli-

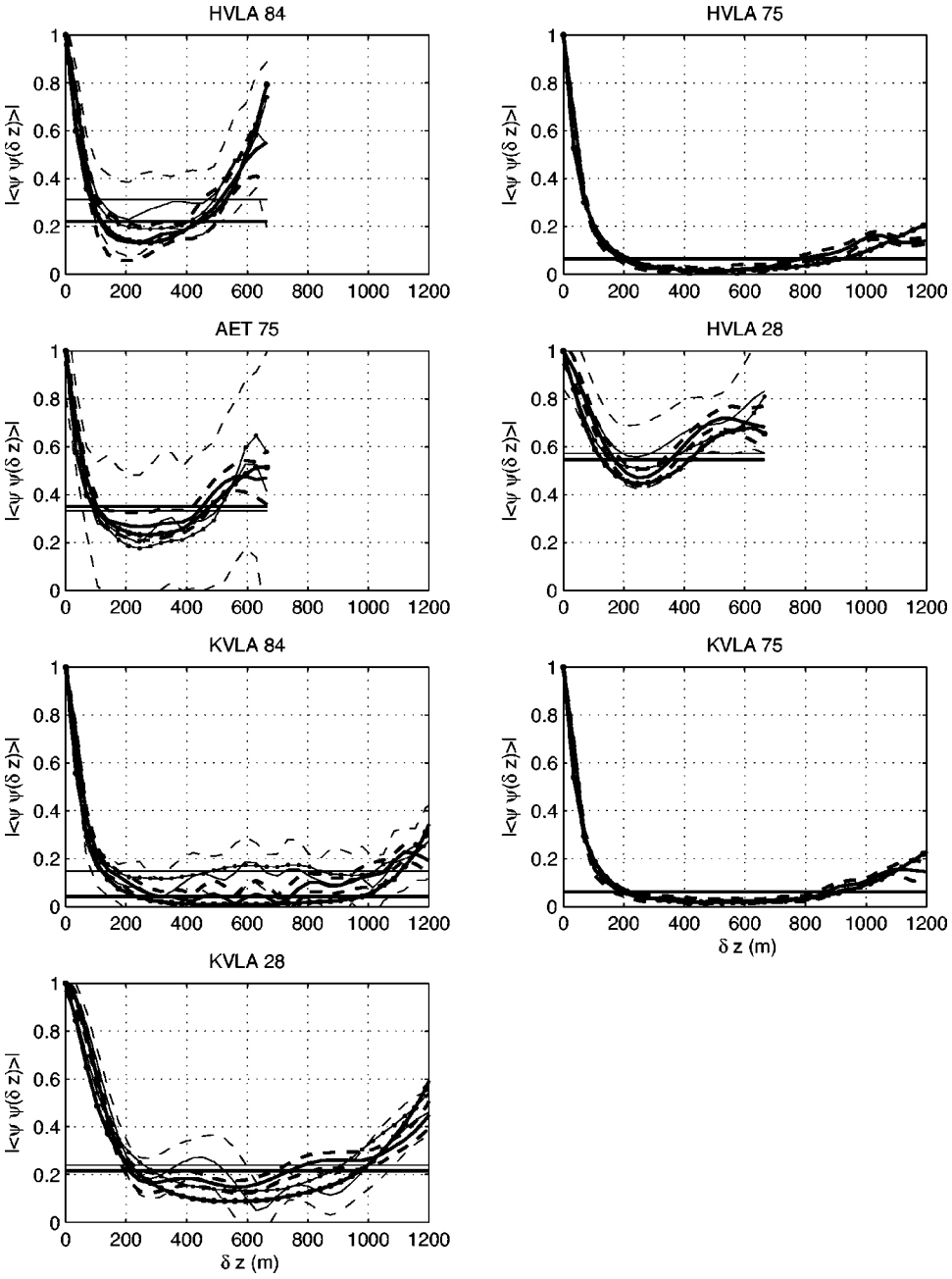


FIG. 16. Depth coherence for the assorted experiments. Light curves denote the longer, 40 min transmissions, while heavy curves are for the 20 min transmissions. Dashed curves are one standard deviation uncertainties in the coherence estimates. Dotted curves denote the coherence computed using $\langle aa(\delta z)\rangle \exp[-D(\delta z)/2]$. The square of the mean field estimates, $|\overline{\langle \psi_z \rangle_{z\tau}}|^2$, are denoted by solid horizontal lines.

tude is unity for scales near the first Fresnel radius and is, on average, 1/2 for smaller scales. The scintillation index is given by $\mathrm{SI}=\exp(\sigma_l^2)-1\sim\sigma_l^2$. As scattering strength increases intensity variability becomes dominated by interference effects whose frequency scaling is poorly understood. Finally, when scattering is very strong, both phase and intensity are dominated by interference effects and intensity statistics are fully saturated, while phase executes a random multipath (Dyson *et al.*, 1976).

Here, comparisons have been made between 84 and 28 Hz transmissions during the 9 days of the AST. In virtually all cases, the lower-frequency data show less fluctuation, but

TABLE VI. Broadband and narrow band (denoted with parentheses) mean field energies. All of the mean field energies exceed the expected value from samples of uncorrelated Gaussian noise. (For $N=10$, 20, 40, and 80 the Gaussian noise values are 0.10, 0.05, 0.03, and 0.01).

	HVLA84	HVLA75	AET75	HVLA28	KVLA84	KVLA75	KVLA28
20 min obs. time							
Depth mean $\lvert\overline{\langle\psi_z\rangle_{z\tau}}\rvert^2$	0.22(0.28)	0.06(0.07)	0.35(0.28)	0.55(0.56)	0.04(0.03)	0.06(0.06)	0.22(0.18)
Time mean $\lvert\langle\psi_\tau\rangle_{z\tau}\rvert^2$	0.54(0.54)	0.60(0.63)	0.42(0.43)	0.62(0.61)	0.52(0.52)	0.64(0.64)	0.71(0.72)
40 min obs. time							
Depth mean $\lvert\overline{\langle\psi_z\rangle_{z\tau}}\rvert^2$	0.31(0.22)	⋯	0.33(0.17)	0.57(0.63)	0.15(0.08)	⋯	0.24(0.34)
Time mean $\lvert\langle\psi_\tau\rangle_{z\tau}\rvert^2$	0.46(0.44)	⋯	0.18(0.29)	0.60(0.71)	0.38(0.39)	⋯	0.71(0.73)

also the frequency spectra at the two frequencies are completely different. The differences in spectra suggest very different scattering physics and thus the difficulty in interpretation of the magnitude of fluctuation reduction is expected. The fully geometric case in which phase variance and mean square phase rate (ν_t^2) scale as the square of the frequency (i.e., a factor of 9 here), and the spectra are independent of frequency are simply not demonstrated by the data. The interpretation of these results represents a major unsolved problem in basin-scale acoustic propagation.

Before leaving this topic, however, it is important to mention AST results for resolved wave front arrivals observed on two horizontal arrays at the 150 and 700 km range (Worcester *et al.*, 2000). At the 150 km range, large phase (i.e., travel time) fluctuations at both 28 and 84 Hz are reported, but there was a reduction of SI from 0.64 to 0.26; nearly a factor of 3! At the 700 km range, on the other hand, the phase variance at 28 and 84 Hz were 1.9 and 29.0 (rad^2) (a factor of 15), while SI values were 0.11 and 0.65 (a factor of 6). Again, the scalings with frequency (and range) are quite variable, although it must be noted that these two paths traverse rather different oceanographic regions.

C. Time and space interference patterns

The analysis of this paper shows that interference effects are very important, but they are quite different in the time and depth directions, as resolved by our limited vertical and temporal sampling. In the depth direction both high and low acoustic frequencies are very nicely described by the Dyson model, which shows the pattern dominated by fadeouts at high wave number. High acoustic frequencies also show some random walk behavior at low wave number. The key parameter here is the mean square phase rate in depth ν_z. In the temporal direction the interference pattern does not follow the Dyson model and thus is not consistent with a fully randomized multipath pattern, over the 20 and 40 min observation times. This result is consistent with the correlation and coherence functions from Figs. 11–16, which show that the vertical aperture of the observations involved many correlation lengths, while the temporal observation time barely involves one correlation time. The critical question is whether longer observation times would show a significantly more developed temporal interference pattern; if not this would imply some degree of determinsitic multipathing. Similar differences between spatial and temporal interference patterns were observed by Porter and Spindel (1977) for moored and drifting receptions at 200 and 400 Hz in the North Atlantic. This is an area that will require further theoretical or modeling work.

D. Bandwidth dependence of fluctuations

The bandwidth dependence of acoustic variability is also an area in ocean acoustic propagation that is poorly understood (Flatté *et al.*, 1979; Dashen *et al.*, 1985; Colosi *et al.*, 1999b), but has great relevance to signal statistics and stability. In this section a discussion of the topics of (1) the approach to saturation, an intensity statistic, and (2) coherence, a strongly phase dominated statistic, are presented. In this discussion a kinematical model of multipath wave-packet interference [see Colosi and Baggeroer (2003)] is utilized, which is based on a wave function of the form

$$\psi = \sum_{k=1}^{N} a_k \mathrm{e}^{-Q^2\phi_k^2/2} \mathrm{e}^{-i\phi_k} \mathrm{e}^{-i\theta_k}, \tag{19}$$

where N is the number of interfering wave packets, a_k are the amplitudes of the waves, $Q = \Delta\omega/\omega_c$ is the fractional bandwidth of the Gaussian wave packets, and ϕ_k and θ_k are the travel-time-induced and nontravel-time induced (for example, caustic phases) phase shifts. In this model the separate multipaths are uncorrelated, and for a given multipath phase and amplitude, are uncorrelated (see Appendix C).

1. Approach to saturation

Using the multipath model of Eq. (19), Colosi and Baggeroer (2004) were able to analytically compute the scintillation index near saturation, where they found

$$\mathrm{SI} \simeq 1 + \frac{1}{N}(\Gamma - 2), \quad \text{narrow band: } Q = 0,\ \Phi \gg 1, \tag{20}$$

$$\mathrm{SI} \simeq 1 + \frac{1}{N}(Q\Phi\Gamma - 2), \quad \text{broadband: } Q\Phi \gg 1; \tag{21}$$

where $\Gamma = \langle a^4 \rangle / \langle a^2 \rangle^2$ is a ratio of the multipath amplitude moments (see Appendix C). Note that the asymptotics are different for the narrow band and broadband case, and that these results are not a function of the nontravel-time-induced phase variance $\langle \theta_k^2 \rangle = \Theta^2$. An important parameter gauging the effect of bandwidth is the time–bandwidth product $Q\Phi$, where Φ is the rms phase fluctuation caused by travel time fluctuations. The large and small Γ cases are noteworthy. If Γ is small enough so that $Q\Phi\Gamma \ll 2$, then SI will approach saturation (i.e., SI=1) from below 1, and will not be significantly affected by bandwidth. If, on the other hand $\Gamma \gg 2$, SI will approach saturation from above 1, and the correction to SI−1 will be larger for the broadband case by a factor of $Q\Phi$; that is, a broadband signal will approach saturation more slowly. These results also have implications for the multipath phase. In saturation, the phase is uniformly distributed between 0 and 2π; thus in the large Γ case if narrow band signals are approaching saturation more rapidly than the broadband case, then narrow band phase variance should be larger than broadband variance.

It remains to compute the time–bandwidth product $Q\Phi$ for the present broadband experimental data. An equivalent Gaussian pulse that matches the AET and ATOC95 75 Hz pulses (Colosi *et al.*, 1999a; K. Metzger, 2004) has a fractional bandwidth of roughly $Q = 12/75 = 0.16$. For the AST, the fractional bandwidth at both 28 and 84 Hz is smaller by a factor of 2 or 3 (Worcester *et al.*, 2000). As previously mentioned, results from the AET using travel time fluctuation of the final arrival as a surrogate for phase, found rms phase fluctuations in the finale of order one to two cycles (over a six day period). Thus, for the 75 Hz transmissions $Q\Phi$ is of order 1 to 2, and for the 84 Hz data it is estimated that $Q\Phi$ is of order 0.3 to 1 (i.e., Φ the same as 75 Hz, but

Q reduced). Assuming the multipaths behave geometrically in phase at 28 Hz, $Q\Phi$ is reduced by an additional factor of 3 over the 84 Hz case, yielding values between 0.1 and 0.3. Clearly this is a regime where an asymptotic analysis for large or small $Q\Phi$ is not entirely appropriate. Thus, the narrow band result [Eq. (20)] applies directly to our narrow band data, while the broadband result [Eq. (21)] marginally applies to our 75 Hz data.

Now this kinematical picture is applied to the SI observations, where the long-time scale intensity variability must be considered since sampling has occurred over the entire ''broadband'' of ocean internal-wave variability (See Table IV, Column 5). As was pointed out previously, both broadband and narrow band estimates of SI are larger than 1, but the excess scintillation above 1 has been attributed to modulation (Sec. IV). Thus, the observations do not reveal whether SI has approached saturation from above or below 1. Therefore, the discrimination between broadband and narrow band cases for long or short observation times cannot be made since the statistics are so close to saturation. This also explains why the phase statistics are not a function of bandwidth since both narrow band and broadband data are already close to saturation.

Observations at shorter ranges are needed to resolve this question, and will give important information concerning the amplitude and phase statistics of the interfering multipaths. For example, the narrow band approach to saturation will constrain the magnitude of Γ, while the broadband approach to saturation will constrain Φ^2.

2. Multipath coherence

Next, broadband effects on coherence are discussed. The wave packet multipath model [Eq. (19)] can be used to derive an analytic expression for coherence (see Appendix C), so that

$$\frac{\langle \psi_1 \psi_2^*(\Delta)\rangle}{\langle I\rangle} = \frac{1}{\langle I\rangle}\left(\frac{\langle aa(\Delta)\rangle G_Q(\rho)}{((1+Q^2\Phi^2)^2 - Q^4\Phi^4\rho^2)^{1/2}} + \frac{(N-1)\langle a\rangle^2 G_Q(0)}{1+Q^2\Phi^2}\right), \tag{22}$$

$$G_Q(\rho;Q,\Phi) = \exp\left[-\frac{D(\rho)}{2+Q^2 D(\rho)}\right],$$

$$G_0(\rho;Q=0,\Phi) = \exp[-\Phi^2(1-\rho)], \tag{23}$$

where $\rho(\Delta) = \langle \phi\phi(\Delta)\rangle/\Phi^2$ is the autocorrelation coefficient of the lagged phase, $D(\rho) = 2\Phi^2(1-\rho)$ is the phase structure function, and $\langle aa(\Delta)\rangle$ is the amplitude correlation function. The mean intensity is, simply,

$$\langle I\rangle = \frac{\langle a^2\rangle}{(1+2Q^2\Phi^2)^{1/2}} + \frac{(N-1)\langle a\rangle^2 G_Q(0)}{1+Q^2\Phi^2}. \tag{24}$$

For simplicity, the nontravel-time-induced phase shifts are not considered.

Some asymptotic analysis of Eq. (22) is useful. Considering the limit of no phase correlation $\rho=0$ (i.e., large lags), it is found that

$$\frac{\langle \psi_1 \psi_2^*(\Delta)\rangle}{\langle I\rangle} = \frac{N}{N-1}\,\frac{G_Q(0)}{\gamma F(Q\Phi) + G_Q(0)}, \tag{25}$$

$$F(Q\Phi) = \frac{1+Q^2\Phi^2}{(1+2Q^2\Phi^2)^{1/2}}, \quad \gamma = \frac{1}{N-1}\,\frac{\langle a^2\rangle}{\langle a\rangle^2}. \tag{26}$$

The important parameter γ (which is closely related to Γ from the scintillation analysis) measures the relative contributions from the two terms in the coherence, the first term being the contribution from the decorrelation of the individual multipaths, and the second term coming from the energy of the independent multipaths. For $\gamma \gg G_Q(0)$, $G_0(0)$, both the broadband and narrow band cases give a small asymptotic coherence of order $G_0(0)/\gamma$ and $G_Q(0)/(F\gamma)$; that is, the coherence will be dominated by the decorrelation of the multipaths and thus will decay significantly. The bandwidth dependence of this limit depends on the relative size of $G_0(0)$ and $G_Q(0)/F$. In the limit of large $Q\Phi$ the results are $\exp(-\Phi^2)/\gamma$ and $\exp(-Q^{-2})/(\gamma Q\Phi)$ for the narrow band and broadband cases: the decay goes from exponential in Φ for the narrow band case to power law in Φ for the broadband case. On the other hand, for $\gamma \ll G_0(0)$ (narrow band), and $F\gamma \ll G_Q(0)$ (broadband), the coherence will be of order $1-\gamma/G_0(0)$ and $1-F\gamma/G_Q(0)$ for the narrow band and broadband cases, respectively. Thus, the narrow band case will decorrelate less by a factor of $F(Q\Phi)$, which is equal to $Q\Phi$ in the limit of large $Q\Phi$.

At the other end of the coherence function $\rho = 1-\epsilon$, $\epsilon \ll 1$ the asymptotics yields

$$\frac{\langle \psi_1 \psi_2^*(\Delta)\rangle}{\langle I\rangle} = 1 - \epsilon\,\frac{F\gamma(1+\Phi^2)}{F\gamma + G_Q(0)}. \tag{27}$$

In this calculation the amplitude covariance is assumed to have the form $\langle aa(\Delta)\rangle = \langle a^2\rangle(1-\epsilon)$. For $\gamma \gg G_Q(0)$, $G_0(0)$, both the broadband and narrow band cases give a coherence of order $1-\epsilon(1+\Phi^2)$, independent of bandwidth. In the other case, for $\gamma \ll G_0(0)$ (narrow band), and $F\gamma \ll G_Q(0)$ (broadband), the coherence will be of order $1-\epsilon\gamma(1+\Phi^2)$ and $1-\epsilon\gamma(1+\Phi^2)F$, respectively. Here the broadband case will decay more rapidly by a factor of $F(Q\Phi)$.

Now this kinematical model is applied to the observations. The coherence analysis of this paper has been limited to short time scales of 20 and 40 min observation times, thus $\Phi \simeq 1$ rad, and for the 75 Hz source, $Q\Phi \simeq 0.2$. Hence, with these values the kinematical model gives $F(Q\Phi) \sim F(0) = 1$ and $G_Q(0) \sim G_0(0)$, and so broadband statistics are virtually indistinguishable from narrow band statistics, as observed. Clearly, longer observation times will give larger $Q\Phi$ as Φ grows. The interesting question is will the bandwidth dependence of the coherence be manifest at these longer observation times; that is, kinematically, will the large or small γ cases be observed. This data would give important information about the statistics of individual multipath amplitudes a_k and number N.

VII. SUMMARY

A comprehensive analysis of the space–time scales of observations of broadband basin-scale acoustic transmission data in the finales of the receptions at center frequencies of 28, 75, and 84 Hz has been presented. Over time scales of 20 to 40 mins, rms phase variations are extremely small and on the order of 0.1 cycles for all frequencies. Phase variations are shown to approximately fit a Gaussian PDF. Over the same time scales, the intensity variability is relatively weak, with 75 and 84 Hz SI values between 0.4 and 0.7, and 28 Hz SI values between 0.3 and 0.7. Over long, multiday time scales, the intensity variability is consistent with fully saturated statistics (i.e., SI$\simeq$1) and modulation effects, so that a modulated exponential PDF of intensity fits the observed PDFs very well. Phase variability cannot be studied at long time scales because of long gaps in the transmissions. Frequency and vertical wave number spectra of phase and log intensity reveal the basic time–space scales of the observations. Frequency spectra at 84 and 75 Hz are completely different from 28 Hz, implying significantly different scattering physics, while vertical wave number spectra at all acoustic frequencies show results consistent with a fully developed interference pattern. Frequency spectra in all cases are ''red,'' implying that longer observation times are needed to fully resolve the variability.

With an understanding of the important space–time scales of phase and intensity we are able to present a comprehensive analysis of coherence. A theoretical coherence model that is consistent with the observations utilizes the facts that phase and amplitude are uncorrelated, phase is Gaussian, phase correlation is not quadratic in the lag, and amplitude decorrelation plays a key role. Consistent with previous results of phase and amplitude stability over 20 and 40 min time scales, the observed coherence is high. Using the mean field as an indicator of coherence it is found that at 84 and 75 Hz acoustic frequencies the mean field accounts for 40%–60% of the total energy, while at 28 Hz the proportion grows even larger, to 60%–70% of the energy.

For all of the observables in this study both a broadband and a narrow band analysis has been carried out, and we find essentially no difference between broadband and narrow band results. Broadband kinematical models of SI and coherence show that our observations are not in regimes in which a significant bandwidth sensitivity could be observed.

Finally, the observations indicate a surprising degree of sensitivity to the geographic separations of the transmission paths. It is found that the longest path, KVLA, is as stable or moreso than the shorter path HVLA. Further, the HVLA path is more stable than the AET path, which has a comparable range.

The present analysis poses many new questions that will be the focus of future theoretical and observational work.

ACKNOWLEDGMENTS

The existence of the data presented in this paper represents a monumental scientific and political achievement, that could not have been done without the heroic efforts of many people. In the view of the first author (Colosi) key individuals were Arthur Baggeroer, Chris Clark, Adam Frankel, Andrew Forbes, Bruce Howe, Jim Mercer, Kurt Metzger, Walter Munk, Bob Spindel, and Peter Worcester, and for their efforts the author is truly grateful. Further, the first author would also like to extend special appreciation to the Alternate Source Test (AST) group of Bruce Howe, Jim Mercer, Kurt Metzger, and Peter Worcester, whose skillful work provided a very unique dataset. The contents of this paper have been shaped and refined over many useful conversations with the NPAL group, Jim Lynch, and Tim Duda. Much inspiration has been derived from conversations with, and the work of Stan Flatté and Fred Tappert. This is Woods Hole Oceanographic Institution contribution No. 11073.

APPENDIX A: 2-D PHASE UNWRAPPING

Following Ghilia and Romero (1994), we define wrapped and unwrapped phase functions ϕ_w and ϕ_u for a 2-D lattice $(x,y)=(i,j)$. From the knowledge of the wrapped phases ϕ_w, the following differences are computed:

$$\Delta^x(i,j)=\phi_w(i+1,j)-\phi_w(i,j) \quad \text{and}$$
$$\Delta^y(i,j)=\phi_w(i,j+1)-\phi_w(i,j) \tag{A1}$$

and all edge values of Δ^x and Δ^y are set to zero. Next, the objective function of the difference between wrapped and unwrapped phase gradients is formed to obtain

$$J^2=\sum_{i=1}^{N_x-1}\sum_{j=1}^{N_y}(\phi_u(i+1,j)-\phi_u(i,j)-\Delta^x(i,j))^2 + \sum_{i=1}^{N_x}\sum_{j=1}^{N_y-1}(\phi_u(i,j+1)-\phi_u(i,j)-\Delta^y(i,j))^2. \tag{A2}$$

The least square solution ϕ_u that minimizes J^2 is given by

$$(\phi_u(i+1,j)-2\phi_u(i,j)+\phi_u(i-1,j))+(\phi_u(i,j+1) -2\phi_u(i,j)+\phi_u(i,j-1))=\Gamma(i,j), \tag{A3}$$

$$\Gamma(i,j)=(\Delta^x(i,j)-\Delta^x(i-1,j))+(\Delta^y(i,j)-\Delta^y(i,j-1)). \tag{A4}$$

Equation (A3) is seen to be the discrete form of Poisson's equation. Setting all gradients of ϕ equal to zero on the edges (Neumann boundary conditions), the numerical solution of Eq. (A3) can be done very efficiently using fast cosine transforms.

APPENDIX B: NOTATION FOR STATISTICAL AVERAGES

A notation for statistical averages needs to be defined since some complications arise due to the three-dimensionality of the data. For a general observable $F(z,\tau,l)$ the means over z and τ are defined using angle brackets so that

$$\langle F\rangle_z=\frac{1}{N_z}\sum_{j=1}^{N_z}F(z_j,\tau,l),$$

$$\langle F\rangle_\tau = \frac{1}{N_\tau}\sum_{j=1}^{N_\tau} F(z,\tau_j,l),$$

$$\langle F\rangle_{z\tau} = \frac{1}{N_\tau N_z}\sum_{k=1}^{N_\tau}\sum_{j=1}^{N_z} F(z_j,\tau_k,l). \tag{B1}$$

Averages over transmission number l, however, are denoted with an overbar, giving

$$\overline{F} = \frac{1}{N_l}\sum_{l=1}^{N_l} F(z,\tau,l). \tag{B2}$$

APPENDIX C: MULTIPATH COHERENCE THEORY

Consider a multipath–broadband wave function of the form (see Colosi and Baggeroer, 2004),

$$\psi = \sum_{k=1}^{N} a_k e^{-Q^2\phi_k^2/2} e^{i\phi_k}, \tag{C1}$$

where Q=bandwidth/center frequency (for the assumed Gaussian wave packet, Q is the ratio of the standard deviation of the frequency and the center frequency). The multipaths are assumed uncorrelated and for an individual multipath the amplitude and phase are uncorrelated, thus

$$\langle a_k a_j\rangle = \langle a\rangle^2 \quad (k\neq j), \quad \langle a_k a_k\rangle = \langle aa(\Delta)\rangle,$$

$$\langle \phi_k\phi_j\rangle = \Phi^2\delta_{k,j}, \quad \text{and } \langle a_k\phi_j\rangle = 0. \tag{C2}$$

The PDF of the amplitude a is unconstrained but a Gaussian joint probability function for the phase is assumed, which is given by

$$P(\phi_1,\phi_2;\rho) = \frac{1}{2\pi\Phi^2(1-\rho^2)^{1/2}} \times \exp\left[-\frac{\phi_1^2 - 2\rho\phi_1\phi_2 + \phi_2^2}{2\Phi^2(1-\rho^2)}\right], \tag{C3}$$

where $\rho(\Delta) = \langle\phi_1\phi_2(\Delta)\rangle/\Phi^2$ is the autocorrelation coefficient of the lagged phase, Φ^2 is the phase variance, and Δ could be a temporal or spatial lag.

The coherence is given by

$$\langle\psi_1\psi_2^*(\Delta)\rangle = \sum_{k=1}^{N}\sum_{j=1}^{N}\langle a_k a_j(\Delta)\rangle \times \langle e^{-Q^2(\phi_k^2+\phi_j^2(\Delta))/2} e^{i(\phi_k-\phi_j(\Delta))}\rangle. \tag{C4}$$

There are N terms for $j=k$ that have the form

$$\langle aa(\Delta)\rangle \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} d\phi_1\, d\phi_2\, P(\phi_1,\phi_2;\rho(\Delta)) \times e^{-Q^2(\phi_1^2+\phi_2^2)/2} e^{i(\phi_1-\phi_2)}, \tag{C5}$$

and there are $N(N-1)$ terms for $j\neq k$ that have the form

$$\left(\langle a\rangle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} d\phi_1\, d\phi_2\, P(\phi_1,\phi_2;\rho(\Delta)=0) e^{-Q^2\phi_1^2/2} e^{i\phi_1}\right)^2. \tag{C6}$$

The joint probability function P is assumed Gaussian [Eq. (C3)], which makes all the integrals Gaussian. Thus they are easily solved by standard methods to give

$$\frac{\langle\psi_1\psi_2^*(\Delta)\rangle}{\langle I\rangle} = \frac{1}{\langle I\rangle}\left(\frac{N\langle aa(\Delta)\rangle G(\rho)}{((1+Q^2\Phi^2)^2 - Q^4\Phi^4\rho^2)^{1/2}} + \frac{N(N-1)\langle a\rangle^2 G(0)}{1+Q^2\Phi^2}\right), \tag{C7}$$

$$G(\rho;Q,\Phi) = \exp\left[-\frac{D(\rho)}{2+Q^2 D(\rho)}\right], \tag{C8}$$

where $D(\rho) = 2\Phi^2(1-\rho)$ is the phase structure function, and the mean intensity is

$$\langle I\rangle = \frac{N\langle a^2\rangle}{(1+2Q^2\Phi^2)^{1/2}} + \frac{N(N-1)\langle a\rangle^2 G(0)}{1+Q^2\Phi^2}. \tag{C9}$$

For only one path ($N=1$) the analytics can be compared to standard results; thus

$$\frac{\langle\psi_1\psi_2^*(\Delta)\rangle}{\langle I\rangle} = \frac{\langle aa(\Delta)\rangle}{\langle a^2\rangle} G(\rho)\left[1 + \frac{Q^4\Phi^4(1-\rho^2)}{1+2Q^2\Phi^2}\right]^{-1/2} \tag{C10}$$

and in the limit of zero bandwidth ($Q=0$) Eq. (17) is recovered.

Beron-Vera, F. J., Brown, M. G., Colosi, J. A., Tomsovic, S., Virovlyansky, A., Wolfson, M. A., and Zaslavsky, G. (**2003**). "Ray dynamics in a long-range acoustic propagation experiment," J. Acoust. Soc. Am. **114**, 1226–1242.

Colosi, J. A., Flatté, S. M., and Bracher, C. (**1994**). "Internal-wave effects on 1000-km oceanic acoustic pulse propagation: Simulation and comparison to experiment," J. Acoust. Soc. Am. **96**, 452–468.

Colosi, J. A., and Flatté, S. M. (**1996**). "Mode coupling by internal waves for multimegameter acoustic propagation in the ocean," J. Acoust. Soc. Am. **100**, 3607–3620.

Colosi, J. A., Scheer, E. K., Flatté, S. M., Cornuelle, B. D., Dzieciuch, M. A., Munk, W. H., Worcester, P. F., Howe, B. M., Mercer, J. A., Spindel, R. C., Metzger, K., Birdsall, T. G., and Baggeroer, A. B. (**1999a**). "Comparisons of measured and predicted acoustic fluctuations for a 3250-km propagation experiment in the eastern North Pacific Ocean," J. Acoust. Soc. Am. **105**, 3202–3218.

Colosi, J. A., and the ATOC Group (**1999b**). "A review of recent results on ocean acoustic wave propagation in random media: Basin scales," IEEE J. Ocean. Eng. **24**, 138–155.

Colosi, J. A., Tappert, F. D., and Dzieciuch, M. A. (**2001**). "Further analysis of intensity fluctuations from a 3252-km acoustic propagation experiment in the eastern North Pacific Ocean," J. Acoust. Soc. Am. **110**, 163–169.

Colosi, J. A., and Baggeroer, A. B. (**2004**). "On the kinematics of broadband multipath scintillation and the approach to saturation," J. Acoust. Soc. Am. **116**(3), pp. 3515-3522.

Dashen, R., Flatté, S. M., and Reynolds, S. (**1985**). "Path-integral treatment of acoustic mutual coherence functions for rays in a sound channel," J. Acoust. Soc. Am. **77**, 1716–1722.

Dushaw, B. D., Howe, B. M., Mercer, J. A., and Spindel, R. C. (**1999**). "Multi-megameter range acoustic data obtained by bottom mounted hydrophone arrays for measurement of ocean temperature," IEEE J. Ocean. Eng. **24**, 203–215.

Duda, T. F., Flatté, S. M., Colosi, J. A., Cornuelle, B. D., Hildebrand, J. A., Hodgkiss, W. S., Worcester, P. F., Howe, B. M., Mercer, J. A., and Spindel, R. C. (**1992**). "Measured wavefront fluctuations in 1000-km pulse propagation in the Pacific Ocean," J. Acoust. Soc. Am. **92**, 939–955.

Dyer, I. (**1970**). "Statistics of sound propagation in the ocean," J. Acoust. Soc. Am. **48**, 337–345.

Dyson, F., Munk, W., and Zetler, B. (**1976**). ''Interpretation of multipath scintillations Eleuthera to Bermuda in terms of internal waves and tides,'' J. Acoust. Soc. Am. **59**, 1121–1133.

Dzieciuch, M. A., Worcester, P. F., and Munk, W. H. (**2001**). ''Turning point filters: Analysis of sound propagation on a gyre scale,'' J. Acoust. Soc. Am. **110**, 135–149.

Dzieciuch, M. A., *et al.* (**2004**). ''Sound propagation through a spicy ocean, the SOFAR overture,'' J. Acoust. Soc. Am. **116**(3) 1447-1462.

Flatté, S. M., Dashen, R., Munk, W., Watson, K., and Zachariasen, F. (**1979**). *Sound Transmission Through a Fluctuating Ocean* (Cambridge University Press, Cambridge).

Flatté, S. M. (**1983**). ''Wave propagation through random media: Contributions from underwater acoustics,'' Proc. IEEE **71**, 1267–1294.

Flatté, S. M., and Stoughton, R. B. (**1988**). ''Predictions of internal-wave effects on ocean acoustic coherence, travel time variance, and intensity moments for long-range propagation,'' J. Acoust. Soc. Am. **92**, 939–955.

Freitag, L., and Stojanovic, M. (**2001**). ''Basin-scale acoustic communication: A feasibility study using tomography M-sequences,'' *Proc. Oceans 2001*, Honolulu, HI, pp. 2256–2261.

Ghilia, D. C., and Romero, L. A. (**1994**). ''Robust two-dimensional weighted and unweighted phase unwrapping that uses fast transforms and iterative methods,'' J. Opt. Soc. Am. A **11**, 107–117.

Henyey, F., and Macaskill, C. (**1996**). ''Sound through the internal wave field,'' in *Stochastic Modeling in Physical Oceanography* (Birkhouser Boston), pp. 141–184.

Metzger, K. (**2004**) (private communication).

Porter, R. P., and Spindel, R. C. (**1977**). ''Low-frequency acoustic fluctuations and internal gravity waves in the ocean,'' J. Acoust. Soc. Am. **61**, 943–958.

Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (**1992**). *Numerical Recipes in* FORTRAN: *The Art of Scientific Computing*, 2nd ed. (Cambridge University Press, Cambridge).

Reynolds, S. A., Flatté, S. M., Dashen, R., Buehler, B., and Maciejewski, P. (**1985**). ''AFAR measurements of acoustic mutual coherence functions of time and frequency,'' J. Acoust. Soc. Am. **77**, 1723–1731.

Simmen, J., Flatté, S. M., and Wang, G. Y. (**1997**). ''Wavefront folding, chaos, and diffraction for sound propagation through ocean internal waves,'' J. Acoust. Soc. Am. **102**, 239–255.

Wage, K. E., Baggeroer, A. B., and Preisig, J. C. (**2002**). ''Modal analysis of broadband acoustic receptions at 3515-km range in the North Pacific using short-time Fourier techniques,'' J. Acoust. Soc. Am. **113**, 801–817.

Wage, K. E., Baggeroer, A. B., Birdsall, T. G., Dzieciuch, M. A., Howe, B. M., Mercer, J. A., Metzger, K., Munk, W. H., Spindel, R. C., and Worcester, P. F. (**2003**). ''A comparative study of mode arrivals at megameter ranges for 28, 75, and 84 Hz sources,'' *Proc. Oceans 2003*, San Diego, CA, pp. 1001–1008.

Wage, K. E., M. A. Dzieciuch, P. F. Worcester, B. M. Howe, and J. A. Mercer (**2005**). ''Mode coherence at megameter ranges in the North Pacific Ocean,'' J. Acoust. Soc. Am. **117**, 1565–1581 (2005)

Williams, K. L., Henyey, F. S., Rouseff, D., Reynolds, S. A., and Ewart, T. E. (**2001**). ''Internal wave effects on high-frequency acoustic propagation to horizontal arrays—Experiment and implications to imaging,'' IEEE J. Ocean. Eng. **26**, 102–112.

Wage, K. E. (**2004**) (private communication).

Worcester, P. F., Cornuelle, B. D., Dzieciuch, M. A., Munk, W. H., Howe, B. M., Mercer, J. A., Spindel, R. C., Colosi, J. A., Metzger, K., Birdsall, T. G., and Baggeroer, A. B. (**1999**). ''A test of basin-scale acoustic thermometry using a large aperture vertical array at 3252-km range in the eastern North Pacific Ocean,'' J. Acoust. Soc. Am. **105**, 3185–3201. Worcester, P. F., Howe, B. M., Mercer, J. A., Dzieciuch, M. A., and the Alternate Source Test (AST) Group (**2000**). ''A comparison of long-range acoustic propagation at ultra-low (28 Hz) and very-low (84 Hz) frequencies,'' in *Proc. of the US-Russia Workshop on Experimental Underwater Acoustics*, edited by V. I. Talanov, Nizhny Novgorod, Institute of Applied Physics, Russian Academy of Sciences, pp. 93–104.

ATOC Consortium (**1998**). ''Ocean climate change: Comparison of acoustic tomography, satellite altimetry, and modeling,'' Science **281**, 1327–1332.

Zebker, H. A., and Liu, Y. (**1998**). ''Phase unwrapping algorithms for radar interferometry; residue cut, least squares, and synthesis algorithms,'' J. Opt. Soc. Am. A **15**, 586–598.

Mode coherence at megameter ranges in the North Pacific Ocean

Kathleen E. Wage
Department of Electrical and Computer Engineering, George Mason University, Fairfax, Virginia 22030

Matthew A. Dzieciuch and Peter F. Worcester
Scripps Institution of Oceanography, La Jolla, California 92093

Bruce M. Howe and James A. Mercer
Applied Physics Laboratory, University of Washington, Seattle, Washington 98195

(Received 27 April 2004; revised 22 July 2004; accepted 23 July 2004)

This article analyzes the coherence of low-mode signals at ranges of 3515 and 5171 km using data from the Acoustic Thermometry of Ocean Climate (ATOC) and Alternate Source Test (AST) experiments. Vertical line arrays at Hawaii and Kiritimati received M-sequences transmitted from two sources: the 75-Hz bottom-mounted ATOC source on Pioneer Seamount and the near-axial dual-frequency (28/84 Hz) AST source deployed nearby. This study demonstrates that the characteristics of the mode signals at 5171-km range are quite similar to those at 3515-km range. At 75 Hz the mode time spreads are on the order of 1.5 s, implying a coherence bandwidth of 0.67 Hz. The time spread of the 28-Hz signals is somewhat lower, but these signals show significantly less frequency-selective fading than the 75-Hz signals, suggesting that at the lower frequency the multipaths are temporally resolvable. Coherence times for mode 1 at 75 Hz are on the order of 8 min for the 3515-km range and 6 min for 5171-km range. At 28 Hz mode 1 is much more stable, with a magnitude-squared coherence of greater than 0.6 for the 20-min transmission period. [DOI: 10.1121/1.1854551]

PACS numbers: 43.30.Qd, 43.30.Re, 43.30.Bp [AIT] Pages: 1565–1581

I. INTRODUCTION

Broadband receptions at megameter ranges have a complicated structure, which consists of a series of multipath arrivals. Tomographers typically separate these arrivals into two categories for processing: the early steep-angle arrivals, which are analyzed using ray models, and the energetic finale arrivals, which are analyzed using normal-mode models. One of the difficulties in using any of these arrivals for tomography is the fluctuations caused by internal waves. The internal wave effects on the rays have been studied extensively using the path-integral theoretical framework, summarized in the monograph by Flatté *et al.*[1] By comparison, the energetic late arrivals associated with the low-order modes have received much less attention. Some theoretical aspects of mode propagation in a fluctuating ocean have been developed in the literature. Dozier and Tappert's seminal work[2,3] describes a theoretical analysis of mode intensity statistics. They used numerical simulations to verify their predictions. In several papers[4,5] Sazontov and others studied the impact of internal waves on mode coherence. Their theoretical work focuses on how mode coherence affects signal processing techniques such as matched filtering[6] and beamforming.[7–9] In terms of experimental results, Colosi *et al.*[10] analyzed the reception finale at 1000-km range using the SLICE89 data set. They showed that the broadening of the transmission finale is due to internal wave effects. Later, Colosi and Flatté[11] studied mode coupling for megameter range propagation using parabolic equation (PE) simulations.

The Acoustic Thermometry of Ocean Climate (ATOC) experiment provided a unique opportunity to study long-range propagation in the North Pacific. In a series of papers[12–14] ATOC researchers used the resolved ray arrivals at ranges of greater than 3 Mm for tomographic inversions. Others considered the problem of source localization at megameter ranges using ATOC transmissions.[15] Several investigators have studied different aspects of the reception finale. Dzieciuch *et al.*[16] developed the turning-point filter to analyze the transition between the ray-like and mode-like arrivals in basin-scale transmissions. Colosi *et al.*[17,18] used data from the ATOC Engineering Test (AET) to analyze intensity fluctuations in the late arrivals. In this issue, Colosi[19] considers the statistics of the acoustic field in the finale region of the ATOC receptions. Wage *et al.*[20] analyzed the mode arrivals in the ATOC vertical array receptions at 3515-km range using short-time Fourier transform (STFT) processing. The STFT analysis showed that the lowest ten modes have complicated multipath arrival patterns characterized by frequency-selective fading and a high degree of temporal variability. Based on the ATOC measurements at 3515 km, Wage *et al.*[20] concluded that the lowest ten modes have roughly equal power and are incoherent, which agrees with Dozier and Tappert's theoretical predictions. Coherence times of the peak mode arrivals were estimated to be on the order of 5.5 min.

This paper extends the previous mode analysis by making use of the extensive data set compiled by the ATOC project. In addition to the 3515-km range receptions discussed above, the ATOC experiment also recorded receptions at a range of 5171 km using a second vertical line array (VLA). The two VLAs recorded 9 months' worth of recep-

tions from the ATOC 75-Hz source located on Pioneer Seamount off the California coast. In conjunction with ATOC, another experiment called the Alternate Source Test (AST) transmitted signals at two carrier frequencies simultaneously over a period of 9 days.[21] The ATOC VLAs recorded these 28-Hz and 84-Hz AST transmissions, facilitating a study of mode fluctuations at different carrier frequencies. Since the AST source was deployed close to the ATOC source, the transmission ranges are comparable. One key difference between the two sources is that the ATOC source was bottom-mounted, while the AST source was deployed in the water column in deeper water. The two sources excited the low-order modes differently.

The ATOC and AST experiments provide a wealth of data with which to study the low-order mode arrivals for different ranges, carrier frequencies, and source excitations. Using these data sets, this paper characterizes the statistics of the low-mode arrivals. Specifically, the analysis has three objectives: (1) to investigate how well the conclusions[20] from the 3515-km range receptions apply to the receptions at 5171 km; (2) to examine the dependence of the mode statistics on source frequency by comparing the ATOC and AST transmissions; (3) to assess whether there are measurable differences in the low-mode arrival patterns due to using bottom-mounted vs suspended sources. Along with offering some new insights into the nature of mode fluctuations at megameter ranges, this paper presents a set of experimentally determined statistics that may be useful in designing better signal processing for the finale and for evaluating stochastic models of mode scattering.

The outline of this article is as follows. Section II describes the relevant details of the ATOC and AST experiments and presents examples of receptions at each range and carrier frequency. This section also discusses the observed differences between receptions from the bottom-mounted ATOC source and the suspended AST source. Simulations indicate that these differences are due to bathymetric interactions near the seamount source. Section III describes the mode-filtering techniques used for the ATOC and AST data sets and illustrates them with examples of the mode 1 estimates for a set of sample receptions. The final part of Sec. III discusses the source and receiver motion corrections that must be applied to the data prior to mode coherence calculations. Section IV presents the mode statistics estimated from the data and compares them to simulation results. Section V concludes the paper with a summary of results and a discussion of how they might be applied in future research.

II. EXPERIMENT DESCRIPTION

In 1995–1996 low-frequency sources located off the California coast transmitted broadband pulses that were recorded by a network of receivers in the North Pacific, including two VLAs and a number of U.S. Navy Sound Surveillance System (SOSUS) arrays.[22] This paper focuses on the VLA data sets since those are the most appropriate for analyzing the axial mode arrivals. Figure 1 shows the geodesic paths to the two 40-element VLAs, which were located near Hawaii and Kiritimati. The following sections review the relevant details of the ATOC and AST transmissions and describe the receptions used in the mode analysis.

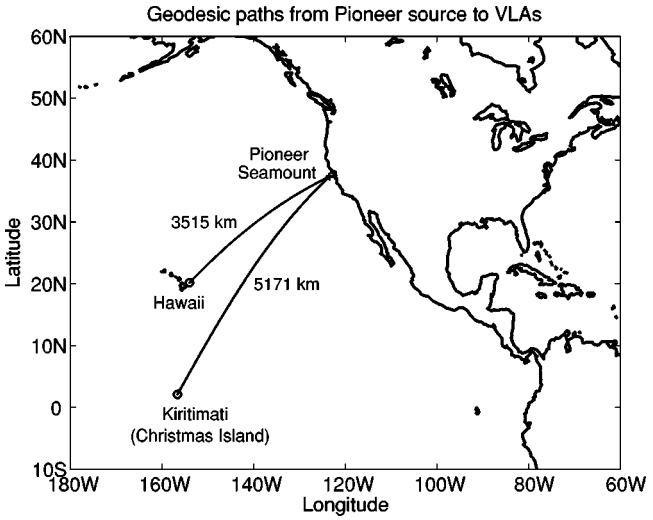


FIG. 1. Geodesic paths from the Pioneer seamount source to the ATOC vertical line arrays at Hawaii and Kiritimati.

A. Transmissions

The ATOC source was located on Pioneer Seamount (indicated in Fig. 1) at a depth of approximately 939 m. Ranges from the ATOC source to the Hawaii and Kiritimati VLAs were 3515 and 5171 km, respectively. The Pioneer source transmitted phase-encoded pseudorandom sequences at a center frequency of 75 Hz, with a 3-dB bandwidth of 37.5 Hz. With 2 cycles/digit, the 1023-digit M-sequence provided a time resolution of 26.67 ms. Each transmission consisted of 44 repetitions of the 27.28-s M-sequence. The ATOC Marine Mammal Research Program determined the transmission schedule. During specified periods, the source transmitted for 20 min every 4 h.

Transmissions from Pioneer Seamount were temporarily discontinued during the AST, which took place in late June and early July of 1996. A source was deployed from the M/V INDEPENDENCE to a depth of 652 m at a position 7 nm southwest of the ATOC Pioneer source. Ranges from the AST source to the Hawaii and Kiritimati VLAs were 3502 and 5157 km, respectively. This suspended source transmitted pseudorandom sequences at two carrier frequencies simultaneously: 28.0425 Hz (fundamental) and 84.1276 Hz (third harmonic). Each had a 3-dB bandwidth of approximately 10 Hz. With 3 and 9 cycles/digit, respectively, these signals provided a travel-time resolution of 106.98 ms. Similar to the ATOC Pioneer transmissions, each M-sequence was 27.28 s long. Most transmissions were for 20 min, occurring at intervals of 4 h, but there were five longer (40-min) transmissions over the course of the experiment. See Worcester *et al.*[21] for additional details about the AST transmissions.

B. Receptions

The VLAs at Hawaii and Kiritimati each consisted of 40 hydrophones with 35-m spacing. To conserve disk space over the long deployment, the hardware computed four-period av-

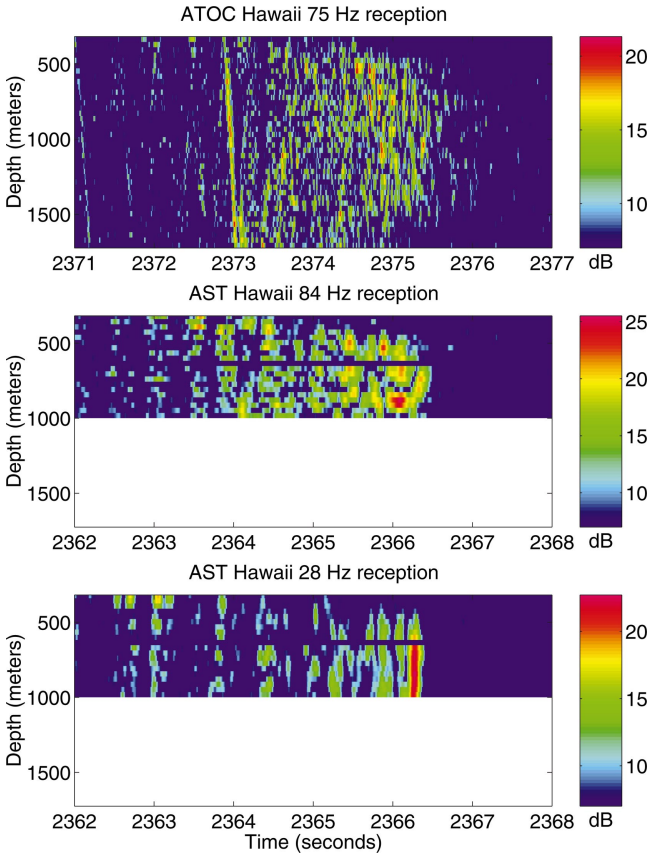


FIG. 2. Receptions at the Hawaii array during the ATOC and AST experiments. Each plot represents one four-period average recorded by the receivers. The top plot shows a reception of a 75-Hz ATOC source transmission. The bottom two plots are examples of AST receptions at the 84- and 28-Hz center frequencies. Although the plots have different starting times, the time-width of all plots is 6 s. Note that the bottom half of the Hawaii array failed prior to the AST experiment; thus, the AST signals were received by only the top half of the VLA. In addition, the tenth hydrophone from the surface on the Hawaii VLA failed prior to the AST experiment.

erages of the M-sequences broadcast by the sources. The VLAs recorded ten of these four-period averages (spanning 18.2 min) for each regular ATOC/AST transmission and recorded 20 four-period averages for the long transmissions from the AST source. The position of the VLAs was tracked using a long-baseline acoustic navigation system. Navigation data were recorded immediately before and after each transmission, and the mooring motion for each sensor was derived from these measurements. For additional information on the ATOC receiver hardware, see the paper by Howe *et al.*[22]

After the arrays were recovered, the time series for each of the four-period averages was complex demodulated and matched filtered. Preprocessing for all the receptions discussed in this paper consists of demodulation and pulse compression, followed by application of the mooring corrections (assuming a nominal wave number). Figure 2 shows representative examples of receptions at the Hawaii array during the ATOC and AST experiments. The top plot shows the received pressure field as a function of time and depth for one of the four-period averages of an ATOC 75-Hz source transmission. The bottom two plots show examples of four-period averages for the 84- and 28-Hz signals from the AST source at the same array. Since the bottom half of the Hawaii VLA failed after approximately 5 months, the AST transmissions were recorded only on the top half of this array. In addition, hydrophone 10 at the Hawaii VLA appears to have failed prior to AST. Note that the starting time for the ATOC reception plot is different from that of the AST plots, due to the fact that the path to Hawaii from the ATOC source is slightly longer than for the AST source. All three of the plots in Fig. 2 display a 6-s time interval.

The time series in Fig. 2 are characteristic of receptions at megameter ranges. The familiar ''accordion'' pattern visible in the early part (2371.5 to 2373 s) of the 75-Hz reception shown in Fig. 2 is due to the constructive interference of the higher-order modes, which travel faster than the lower modes in deep water. The diffuse finale region (2373.5 to 2375.5 s) consists of the lower-order mode arrivals. The AST receptions at 84 and 28 Hz exhibit similar features, with a few notable differences. First, the difference in the time resolution (bandwidth) of the ATOC and AST sources, 26.67 vs 106.98 ms, is evident in the plots. Second, comparing the three plots in Fig. 2 indicates that the AST receptions have a sharper final cutoff, which appears to be tapered in depth, i.e., the shallower hydrophones have earlier cutoffs than the deeper hydrophones. The following section postulates an explanation for this feature. Third, the 28-Hz AST reception is dominated by a strong arrival with spatial characteristics similar to mode 1, whereas the finale regions of the 75- and 84-Hz receptions are characterized by much more complicated interference patterns.

Figure 3 shows representative examples of receptions at the Kiritimati array during the ATOC and AST experiments. For this VLA the tenth hydrophone from the surface appears to have calibration problems, at least for the 75- and 84-Hz receptions. Unlike at the Hawaii array, the 75-Hz reception at Kiritimati is very similar to the 84-Hz reception. Both the 75-Hz data and the 84-Hz data show a sudden cutoff, with little tapering in depth, i.e., the shallow and deep sensors cut off almost simultaneously. Similar to Hawaii, the 28-Hz reception shown in Fig. 3 has a strong arrival with spatial characteristics similar to mode 1.

The ATOC and AST data sets are extensive, including receptions over a 9-month period from late December 1995 until the VLAs were recovered in September 1996. Table I summarizes important information about the receptions in the data sets for each of the experiments. The receptions from the Pioneer source at the Hawaii and Kiritimati arrays are denoted by HVLA75 and KVLA75, respectively. For this analysis the HVLA and KVLA data sets were divided into groups of receptions, clustered over 2- to 4-day periods. During these periods the Pioneer source nominally transmitted for approximately 20 min every 4 h. The Kiritimati data set includes more groups than Hawaii because this study ignores HVLA75 receptions after yearday 509, which is when the lower half of the Hawaii VLA failed, making mode filtering much more difficult.

The AST experiment took place over 9 days from 30 June 1996 to 8 July 1996. During this time period 34 good receptions were recorded on each of the arrays. Of those, five were for the longer 40-min transmissions; the remaining

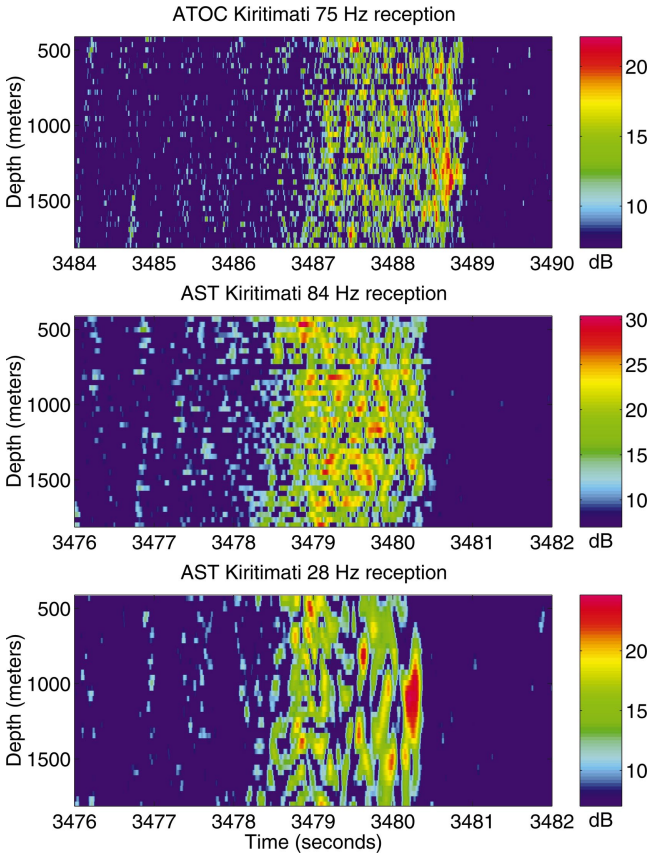


FIG. 3. Receptions at the Kiritimati array during the ATOC and AST experiments. Each plot represents one four-period average recorded by the receivers. The top plot shows a reception of a 75-Hz ATOC source transmission. The bottom two plots are examples of AST receptions at Kiritimati for the 84- and 28-Hz center frequencies. Note that the top plot has a different starting time than the bottom two plots; the timewidth of all three plots is 6 s.

were recordings of standard 20-min transmissions. The 28-Hz carrier frequency receptions at each of the arrays are denoted by HVLA28 and KVLA28, and the 84-Hz transmissions by HVLA84 and KVLA84, respectively.

C. Comparison of receptions from bottom-mounted and suspended sources

As noted above, the AST receptions at Hawaii exhibit a much sharper cutoff as compared to the ATOC 75-Hz receptions. This effect is even more pronounced when the signal is averaged over the entire recording period ($\approx$18.2 min or 40 M-sequences). Figure 4 shows the 40-period average for an ATOC reception and an AST reception. The ATOC reception from the Pioneer source clearly shows between 0.5 to 1.0 s of weaker arrivals following the most energetic axial arrivals, whereas the AST reception has a very sharp cutoff. Since the source frequencies and ranges of these two receptions are quite similar, the primary difference is the bathymetry at the source. Heaney[23] attributes the weaker arrivals in the ATOC finale to bathymetric interaction near the seamount source. The following simulations support this theory and predict the effects that should be evident in the low order mode estimates for the ATOC and AST HVLA receptions.

Figure 5 shows the actual bathymetry along the path from Pioneer to the Hawaii VLA. At its steepest, the slope of the Pioneer-Hawaii path is approximately 20 deg. To examine the effects of the source bathymetry on long-range propagation, consider two broadband simulations implemented using the RAM PE code.[24] The first simulation places the source on the seamount, using the sloping bathymetry shown in Fig. 5. The second simulation keeps the source at the same depth, but removes the seamount, i.e., the bathymetry is as shown by the dashed line in Fig. 5. In both simulations the background sound speed is taken from temperature and salinity profiles for winter, derived from the *World Ocean Atlas*.[25,26] The background environment in both cases is perturbed by internal wave fluctuations at $\frac{1}{2}$ Garrett-Munk strength, generated using the method of Colosi and Brown.[27] The center frequency for these simulations is 75 Hz.

The upper set of plots in Fig. 6 shows the results for the simulation using the measured Pioneer seamount bathymetry. The top plot in this set is the synthesized pressure time series at Hawaii. Note that the finale region is not clearly defined by a sharp cutoff, similar to the ATOC reception discussed above. Below the pressure time series is the modal time series obtained by projecting the field (finely sampled in depth) onto the mode functions at the receiver. In the seamount simulation, modes 1–5 have strong arrivals around 2375 s, followed by almost 1 s of weaker arrivals. The bottom set of plots in Fig. 6 shows the results of the simulation with the seamount removed. In contrast with the seamount simulation, the time series for the suspended source exhibits a very sharp cutoff. Modes 1–5 again have strong arrivals around 2375 s, but they are not followed by any weaker arrivals. Note that the finale region of the suspended source simulation resembles the finale of the AST reception shown in Fig. 2. While more accurate modeling of the near-source bathymetric interaction could be done using a propagation code that incorporates shear effects (such as the study by Sperry *et al.*[28]), these simulations provide strong evidence that the more diffuse finale region seen in the ATOC Hawaii receptions is due to bottom interaction near the Pioneer seamount source.

III. MODE PROCESSING

This section describes the mode-filtering methods used to analyze the ATOC and AST data sets. Mode estimation is a well-known problem that numerous authors have addressed.[29–32] The mode processor designed for the ATOC/AST receptions uses a standard pseudoinverse solution for the spatial filter, coupled with a Fourier decomposition to facilitate broadband processing. Section III A summarizes the parameters for each of the mode filters (designed for the six data sets) and discusses other implementation details. Section III B presents the results of mode processing for each of the sample receptions given in Sec. II. These examples illustrate some of the essential characteristics of the mode arrivals at the three source frequencies and two ranges. Finally, Sec. III C describes how the source motion and mooring mo-

TABLE I. Summary of the ATOC and AST data sets. Yeardays are referenced to 1995.

Data set	Group	Yeardays	Month	No. receptions		No. hydrophones
				20 min	40 min	
HVLA75	1	362–366	Dec.	20		40
	2	370–372	Jan.	10		40
	3	376–377	Jan.	10		40
	4	394–396	Jan.	16		40
	5	407–409	Feb.	9		40
	6	427–428	Mar.	11		40
	7	432–436	Mar.	20		40
	8	465–467	Apr.	14		40
	9	470–473	Apr.	14		40
	10	477–481	Apr.	24		40
	11	485–489	Apr.–May	14		40
	12	499–502	May	13		40
	13	506–509	May	13		40
KVLA75	1	362–366	Dec.	20		39
	2	370–372	Jan.	11		39
	3	376–377	Jan.	6		39
	4	394–396	Jan.	9		39
	5	407–409	Feb.	11		39
	6	427–428	Mar.	11		39
	7	432–436	Mar.	0		39
	8	465–467	Apr.	2		39
	9	470–473	Apr.	14		39
	10	477–481	Apr.	19		39
	11	485–489	Apr.–May	15		39
	12	499–502	May	18		39
	13	506–509	May	16		39
	14	534–537	Jun.	13		39
	15	540–542	Jun.	8		39
	16	566–568	Jul.	11		39
	17	581–584	Aug.	19		39
	18	590–593	Aug.	23		39
	19	597–598	Aug.	11		39
	20	604–606	Aug.	12		39
	21	612–612	Sep.	5		39
HVLA28		547–556	Jun.–Jul.	29	5	19
KVLA28		547–556	Jun.–Jul.	29	5	39
HVLA84		547–556	Jun.–Jul.	29	5	19
KVLA84		547–556	Jun.–Jul.	29	5	39

tion corrections were applied. These corrections are crucial for the estimation of mode coherence statistics, as discussed in Sec. IV.

A. Broadband mode processors for ATOC and AST signals

Mode processing assumes the pressure measured by an array consists of a weighted sum of modes plus noise, i.e., for a vertical array

$$\mathbf{p}(r,\Omega)=\mathbf{\Phi}(r,\Omega)\mathbf{a}(r,\Omega)+\mathbf{n}(\Omega), \tag{1}$$

where $\mathbf{p}$ is the pressure vector at frequency Ω and range r, $\mathbf{\Phi}$ is the matrix of sampled mode shapes, $\mathbf{a}$ is the vector of narrow-band mode amplitudes, and $\mathbf{n}$ is the vector of observation noise. For a broadband source, the measured vector time series $\mathbf{\Psi}(r,\tau)$ can be written as the Fourier synthesis of the narrow-band components defined in Eq. (1)

$$\mathbf{\Psi}(r,\tau)=\int_{\Omega}(\mathbf{\Phi}(r,\Omega)\mathbf{a}(r,\Omega)+\mathbf{n}(\Omega))e^{j\Omega\tau}d\Omega. \tag{2}$$

The goal of mode filtering is to determine estimates of the Fourier mode amplitudes, $\mathbf{a}(r,\Omega)$, or the corresponding mode time series $\boldsymbol{\alpha}(\tau)$

$$\boldsymbol{\alpha}(r,\tau)=\int_{\Omega}\mathbf{a}(r,\Omega)e^{j\Omega\tau}d\Omega. \tag{3}$$

In general the mode shapes vary as a function of frequency, meaning that mode filtering requires a Fourier decomposition prior to the spatial processing to avoid problems with mode shape mismatch.[33] The simplest solution[34–36] to the broadband mode filtering problem is to take a fast Fourier transform (FFT) of the entire reception, do narrow-band spatial processing in each FFT bin, and inverse transform to obtain the estimated modal time series. An alternative approach[20] is to use short-time Fourier transform (STFT)

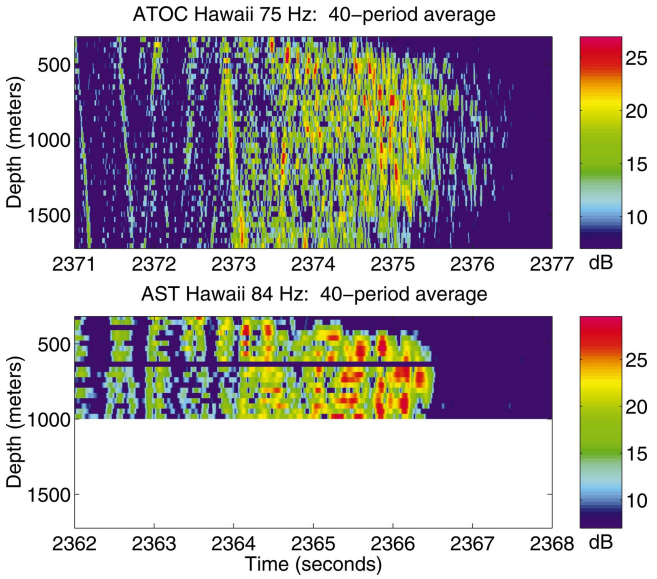


FIG. 4. Forty-period averages for receptions at the Hawaii VLA. The top plot shows an ATOC 75-Hz reception and the bottom plot shows an AST 84-Hz reception.

processing, where the signals are bandpass filtered and mode filtering is done for each band. The result is time-varying spectra for each of the modes. Each approach has its advantages. The FFT processor produces a single broadband time series for each mode that is useful for analyzing the multipath arrival patterns. Each individual FFT bin is comparable to the output of a narrow-band PE run, which facilitates comparisons between experiment and predication without requiring computationally intensive broadband simulations. The main advantage of the short-time Fourier approach is that it permits a study of the characteristics of individual arrivals within a multipath pattern (rather than combining all the paths together), provided they are temporally separable by the STFT processor. The ATOC and AST analysis relies on both techniques.

Once the signal is separated into bins, mode estimation reduces to a classical linear inverse problem. The pseudoinverse (PI) filter[31] is a standard solution where the estimated mode amplitudes are defined to be the product of pseudoinverse of the sampled modeshape matrix and the measured pressure vector. A more general form[37] allows for diagonal loading of the pseudoinverse to accommodate ill-

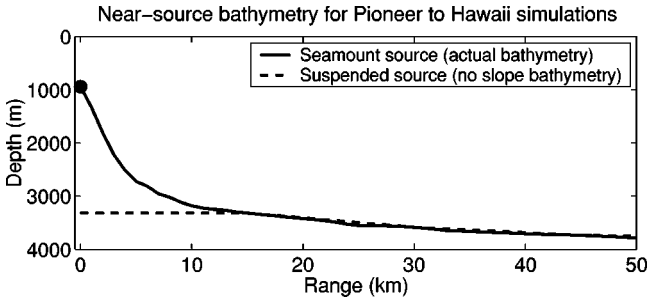


FIG. 5. Bathymetry for the PE simulations. The solid line is the measured bathymetry along the track from Pioneer seamount to the Hawaii VLA. The dashed line is the bathymetry used in the suspended source simulations (the seamount is removed). A large dot indicates the source location.

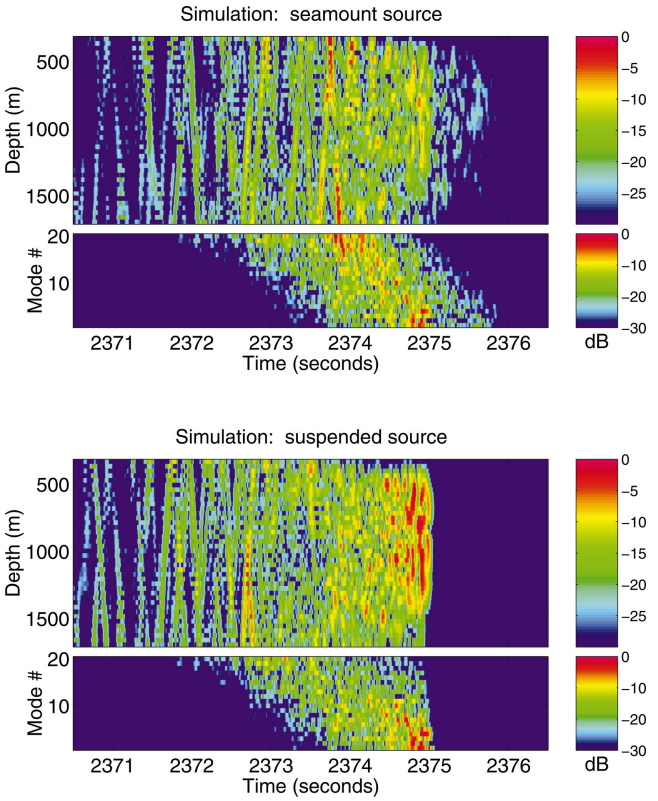


FIG. 6. Broadband PE simulations for winter environment perturbed by $\frac{1}{2}$ Garrett–Munk-strength internal waves. The seamount simulation uses the measured bathymetry for the Pioneer seamount ATOC source. The mid-water column simulation uses a source at the same depth as Pioneer seamount, but the seamount is not included in the bathymetry. Top plot for each simulation is the pressure time series: $20\log_{10}|p(t)|$; the bottom plot is the associated time series for the first 20 modes.

conditioning due to inadequate spatial sampling of the modes, i.e.,

$$\hat{\mathbf{a}} = (\mathbf{\Phi}^T\mathbf{\Phi} + \gamma\mathbf{I})^{-1}\mathbf{\Phi}^T\mathbf{p}, \tag{4}$$

where $\hat{\mathbf{a}}$ is the vector of estimated mode amplitudes and $\mathbf{I}$ is the identity matrix. The parameter γ controls the amount of diagonal loading.

The spatial resolution of the PI mode filter is determined by how well the array samples the modes. Figure 7 shows plots of the mode shapes[38] at the Hawaii and Kiritimati arrays, respectively. Each plot shows the shapes at two frequencies: 75 Hz (ATOC center frequency) and 28 Hz (lower AST center frequency). The plots also indicate the hydrophone depths for each array. Note that the arrays span more of the first 10 modes at 75 Hz than they do at 28 Hz.

Although the continuous modeshape functions are orthogonal, the sampled modeshapes may not be. The sampled mode shape correlation matrix, defined by $\mathbf{\Phi}^T\mathbf{\Phi}$, indicates how orthogonal the sampled modes are and predicts how well the array can separate them. Figure 8 shows the sampled mode shape correlation matrices at 28- and 75 Hz for the two array locations. Each of the 40-element VLAs was designed to estimate the first 10 modes at 75 Hz. The mode shape correlation matrix for HVLA75 confirms that the sampled mode shapes up to mode 10 are effectively orthogonal; above mode 10, neighboring modes are correlated. This is in sharp contrast to the 28-Hz results for the Hawaii array,

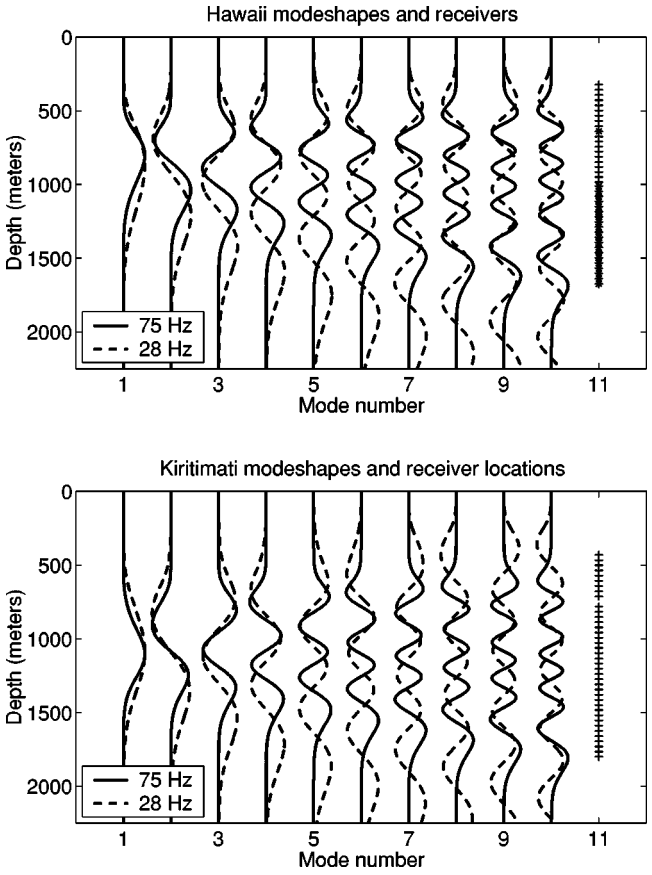


FIG. 7. Mode shapes at 75- and 28 Hz for the Hawaii and Kiritimati VLAs. The hydrophone depths are shown at the far right. The "+" symbols indicate phones that remained operational throughout the experiment, and the "*" symbols indicate phones that failed prior to the AST part of the experiment.

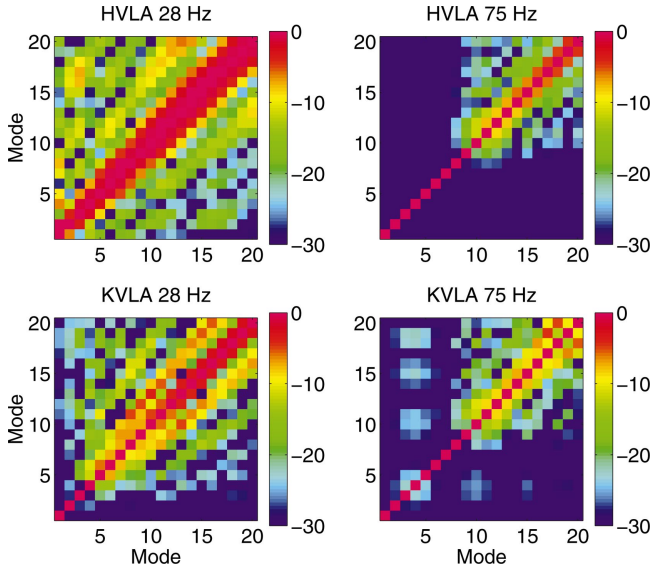


FIG. 8. Sampled mode shape correlation matrices for the Hawaii and Kiritimati VLAs. Color scale is in dB. The Hawaii VLA 75-Hz results are for the full 40-element array, whereas the 28-Hz results are for the 19 elements that were operational during the AST experiment. The Kiritimati VLA results are for the 39-element VLA (i.e., a 40-element VLA with one bad phone).

which indicate a high degree of correlation between neighboring modes. The correlation is caused by inadequate sampling, due to the fact that only 19 hydrophones were operational during the AST experiment. For the Kiritimati array, the 75-Hz results are similar to those at Hawaii, with modes up to 10 being largely uncorrelated. The small amount of crosstalk between modes 3, 4, and 5 is due to the nonuniform sampling caused by the failure of hydrophone 10. At 28 Hz the figure shows that modes up to 4 are approximately orthogonal, as sampled by the 39-element Kiritimati VLA.

As Fig. 8 indicates, modes 1–10 of the ATOC 75-Hz signal are effectively orthogonal as sampled by both of the arrays; thus, they can be estimated with a standard PI filter with no diagonal loading. Some diagonal loading is required to process the Kiritimati 28-Hz data set since the VLA does not sample these lower-frequency modes as well. In addition, diagonal loading is required to stabilize the mode filters designed for the degraded Hawaii VLA in order to compensate for the failure of 21 of the 40 hydrophones. Table II summarizes the parameters for the PI mode filters designed for the ATOC and AST data sets. The rationale for including modes higher than 10 (which the VLAs were designed for) in the filters for the KVLA28 and HVLA84 data sets is that this forces more of the crosstalk from modes not included in the estimate into the higher-order modes. See the article by Wage *et al.*[20] for a more thorough discussion of mode filter design in the context of the ATOC experiment. The fourth column of Table II lists which modes are expected to be reliably estimated by these filters, i.e., which modes have low crosstalk from other modes. The estimates of the number of "good" modes are somewhat conservative. For instance, the table indicates that only mode 1 can be estimated for the Hawaii AST receptions (HVLA28 and HVLA84). This is due to the large amount of nearest-neighbor crosstalk for the degraded (19-phone) Hawaii VLA. Modes higher than 1 could be estimated in groups, e.g., mode 2 and 3 together, but this study focuses only on the mode 1 signals, which are better separated.

B. Mode 1 processing examples

As an example of the processing techniques described above, consider the mode 1 estimates derived from the six sample receptions discussed in Sec. II. Figures 9 and 10 show the estimates for the Hawaii and Kiritimati receptions at the three carrier frequencies. In each plot the short-time Fourier mode spectrum is shown above the time series derived from the FFT processor. Short-time Fourier mode estimates consist of a time series for each of the frequency bins contained in the processor. The STFT processor uses the design parameters selected for the previous analysis[20] of the ATOC Hawaii data. Specifically, the processor uses a 0.4-s

TABLE II. Mode filter parameters for ATOC and AST analyses.

Data set	Modes in PI	Diagonal loading	"Good" modes
HVLA75	1–10	None	1–10
KVLA75	1–10	None	1–10
HVLA28	1–5	5%	1
KVLA28	1–20	10%	1–4
HVLA84	1–20	10%	1
KVLA84	1–10	None	1–10

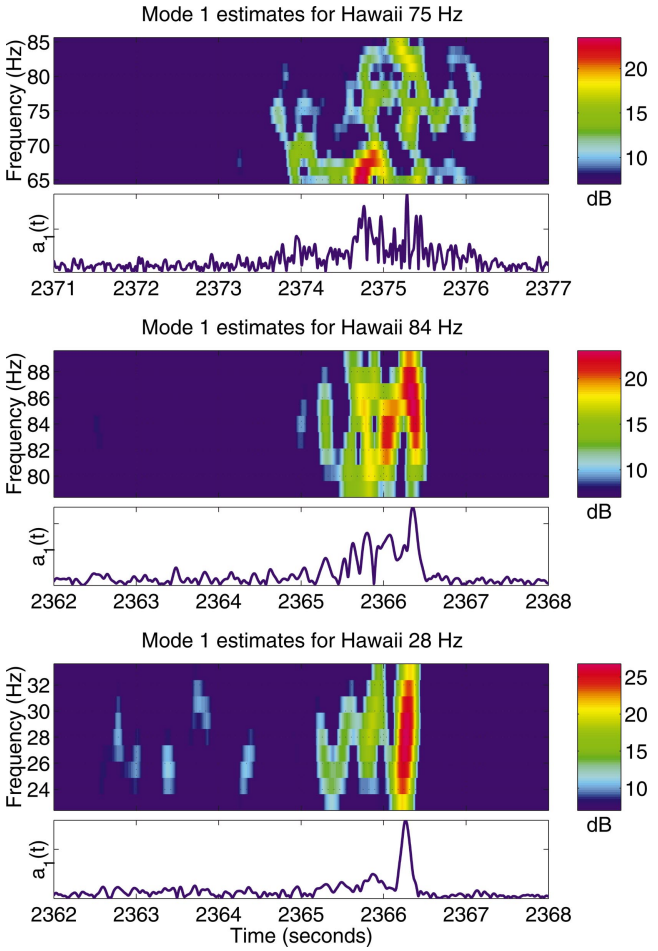


FIG. 9. Mode 1 estimates for the Hawaii receptions shown in Fig. 2. The color images are the short-time Fourier estimates, and the line plots are the time-series output of the FFT processor. The 0-dB level in the color plots corresponds to the predicted noise floor.

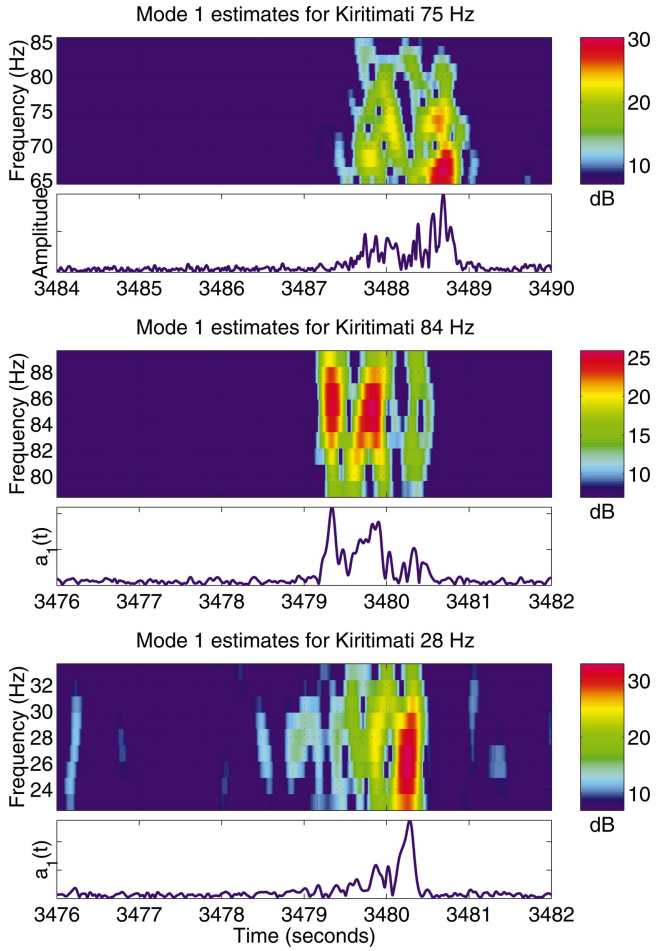


FIG. 10. Mode 1 estimates for the Kiritimati receptions shown in Fig. 3. The color images are the short-time Fourier estimates, and the line plots are the time-series output of the FFT processor. The 0-dB level in the color plots corresponds to the predicted noise floor.

Hanning window, which provides a frequency resolution of approximately ± 2.5 Hz. With this window, the processor can begin to temporally resolve arrivals that are 0.2 s apart. Note that for the plots in Figs. 9 and 10, the bin spacing is 1.25 Hz; thus, neighboring frequency bins are highly correlated. Oversampling by a factor of 2 merely improves the appearance of the plots, but does not change the underlying frequency resolution of the STFT processor. The color scale for the STFT plots is set so that the 0-dB level corresponds to the predicted noise floor.

Figure 9 illustrates several general characteristics of the mode estimates for the ATOC and AST data sets. As noted in a previous article,[20] the HVLA75 results indicate that there is no dominant arrival in mode 1 at 3515-km range. Instead there is a complicated multipath arrival pattern, attributed to internal wave-induced scattering. The short-time Fourier estimates show significant frequency-selective fading, which occurs when two signals with different phase characteristics arrive at the same time (within the 0.2-s resolution of the processor). The constructive and destructive interference of multipath arrivals can produce large peaks and deep fades in the arrivals. As expected from the discussion of receptions for the bottom-mounted source, the mode 1 estimate does not have a sharp cutoff; rather, it trails off with approximately 0.5 s of weaker arrivals.

Figure 10 demonstrates that the mode estimates at 5171-km range are comparable to those at 3515-km range. The KVLA75 mode 1 estimate exhibits frequency-selective fading that is quite similar to the HVLA75 estimate. The arrivals are spread over approximately 1.5 s, which is almost identical to the HVLA75 result, assuming that the 0.5 s of weak arrivals at the end of the HVLA75 estimate are ignored.

Figures 9 and 10 show that the 84-Hz estimates are similar to the 75-Hz estimates, though they have sharper cutoffs, as expected for the suspended AST source. It is important to note that the STFT bins for the ATOC arrivals span a 20-Hz bandwidth from 65 to 85 Hz, whereas the STFT bins for the AST arrivals span half that bandwidth (79 to 89 Hz). The frequency-selective fading in the 84-Hz results is comparable to that observed at 75 Hz, i.e., arrivals consistent over bandwidths of approximately 5 Hz.

The 28-Hz mode 1 estimates shown in Figs. 9 and 10 contrast sharply with the higher-frequency results. At 28 Hz, mode 1 is dominated by a single large arrival. The HVLA28 plot in Fig. 9 shows a strong unfaded final arrival, i.e., an arrival that is consistent across the entire source band. The

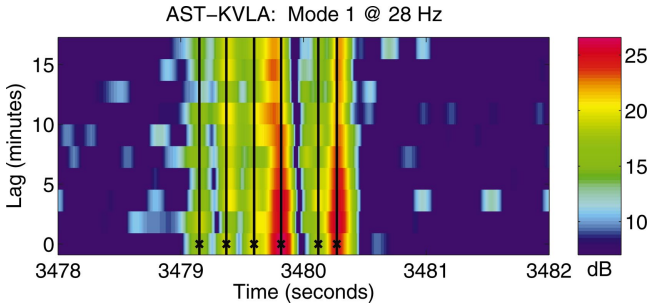


FIG. 11. Plot of short-time Fourier mode estimates in the 28-Hz bin for ten successive 4-period averages in a single reception at the Kiritimati VLA. The *x* axis represents arrival time within a single four-period average, and the *y* axis represents lag time over the approximately 20-min source transmission. The black ×'s mark the peak arrival times in the first four-period average.

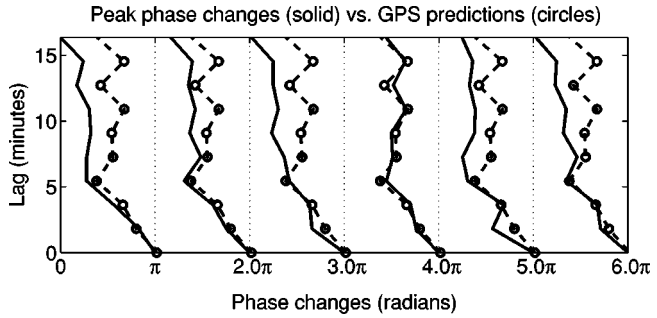


FIG. 12. Relative phase as a function of lag time for the six peak arrivals shown in Fig. 11. Each peak is offset by π radians along the *x* axis, e.g., the curve starting at 2π corresponds to the second peak from the left in Fig. 11. Solid lines correspond to the measured peak phases. The dashed lines marked with circles correspond to the phase changes predicted by the ship GPS record.

lack of frequency-selective fading indicates that this arrival is temporally separable from the earlier multipaths (within the resolution of the STFT processor). The KVLA28 plot in Fig. 10 also shows a large final mode 1 arrival, but it fades a bit toward the upper edge of the band.

C. Source and VLA motion compensation

The results in Figs. 9 and 10 are for single receptions, i.e., one 4-period average of the received M-sequence. In order to study the temporal variations across multiple receptions, the mode processing algorithm must account for any source and receiver motion. This section describes how motion compensation was implemented for the ATOC and AST data sets.

Since the ATOC source was bottom-mounted, those receptions require no source correction. The AST source was deployed from the M/V INDEPENDENCE; thus, the AST receptions must be corrected for fluctuations due to source motion. During the experiment the ship was dynamically positioned using P-code GPS and maintained its station to within about 20 m.[21] Since the acoustic tracking of the suspended source failed, the only available information about source motion is the ship GPS record. To see how well the observed fluctuations match the predicted fluctuations due to source motion, consider an example for the Kiritimati VLA. Figure 11 shows the short-time Fourier estimates in the 28-Hz bin for successive 4-period averages within a single 20-min source transmission. The black ×'s on the plot mark the location of peaks in the first 4-period average that are at least 12 dB above the estimated noise floor. The black lines track those peak arrival times through the estimates at successive lag times. The peak arrivals appear fairly consistent for this reception, though the amplitudes do vary across the 20 min. Figure 12 shows how the phases of the six peak arrivals vary as a function of lag time. Phases are relative to the phase of the first 4-period average for each peak (which has been subtracted out). In this reception, all of the peaks show a sharp negative phase trend, with an overall phase shift of approximately $\pi/2$ radians within the first 5 min.

Given a change in range, it is possible to predict the phase changes in mode 1 as a function of the lag time *t*, i.e.,

$$\Delta\theta(t) = \overline{k_1}\Delta r(t), \tag{5}$$

where $\overline{k_1}$ is the range-averaged wave number for mode 1, and $\Delta r(t)$ is the change in range along the geodesic path. The source GPS data provide an estimate of the range changes as a function of time. To account for the 4-period averaging that is inherent in the ATOC processing, the GPS prediction of the phase changes is defined as

$$\Delta\theta_{\mathrm{GPS}}(t) = \arg\left(\sum_{n=1}^{4} e^{j\overline{k_1}\Delta r_n(t)}\right), \tag{6}$$

where $\Delta r_n(t)$ corresponds to the range associated with the *n*th M-sequence in the 4-period average. Figure 12 shows the GPS phase prediction for the KVLA28 reception described above. The plot shows six identical copies of the GPS prediction as dashed lines with circle markers alongside the phase data (solid lines) for each of the six peaks. Over the first 5 min where the steepest phase transition occurs, the GPS prediction and the observed data are in close agreement. After that, the fourth peak appears to track the GPS prediction, but the other peaks show an additional phase trend that is not in the prediction.

The above example is typical of results for other receptions. Since the GPS data appear to predict a significant component of the observed phase fluctuations, they are used to apply a source motion correction to the AST receptions during mode processing. There will be a residual error due to the fact that the true source motion is not exactly represented by the GPS for the source ship. Some of this error may be removed by the procedure used to correct for the VLA mooring motion, as discussed below.

As indicated in Sec. II, long-baseline navigation systems tracked each VLA during the experiment. Navigation measurements taken immediately before and after each reception yield good estimates of hydrophone depths and their positions relative to the anchor reference point at the beginning of the reception. The velocity of the mooring is expected to be low, on the order of 1 cm/s maximum.[39] In previous work,[20] Wage *et al.* used an approximate mooring velocity estimated from the navigation data to correct for motion of the VLA during the 20-min transmission. While the results appeared reasonable, it is not clear that all of the effects of mooring motion were removed by that correction procedure. Motivated by Colosi's use of linear phase corrections to re-

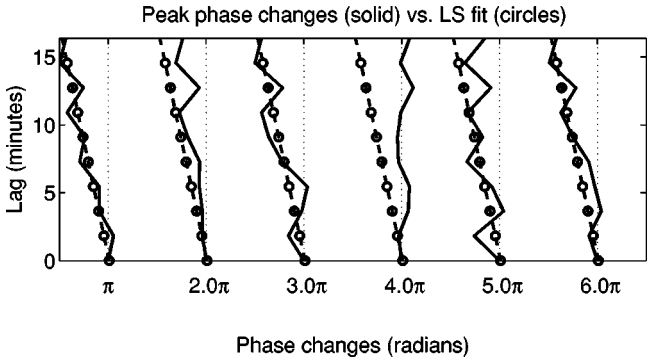


FIG. 13. Relative phase as a function of lag time for the six peak arrivals in Fig. 11 after application of the source motion correction. Each peak is offset by π radians along the x axis, e.g., the curve starting at 2π corresponds to the second peak from the left in Fig. 11. Solid lines correspond to the measured peak phases. Dashed lines marked by circles correspond to a linear fit to the estimated phase trend for the reception.

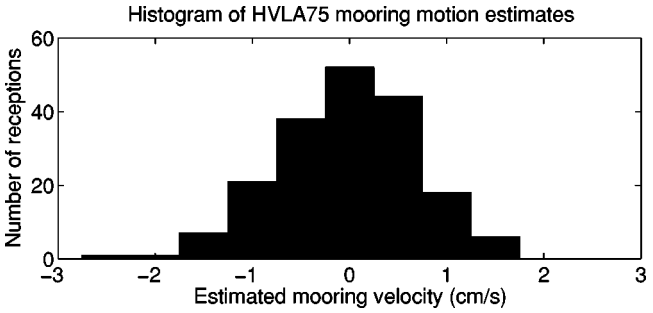


FIG. 14. Histogram of mooring velocity estimates for the 188 good ATOC receptions at the Hawaii VLA.

move residual mooring motion in his analysis[19] of the pressure field, this paper considers a similar approach for the mode analysis.

To understand how the VLA motion was estimated, consider Fig. 13, which shows the relative phases of the peaks in KVLA28 example reception (discussed above) after the source correction has been applied. Five out of the six peaks have phases that decrease approximately linearly as a function of lag time. The dashed line marked with circles plotted next to each peak represents a linear least-squares fit to the phase associated with the third peak. The third peak was chosen because it has the most consistent amplitude across all the lag times. The linear approximation appears to fit all the peaks reasonably well, with the exception of peak 4. Note that peak 4 shows the most fading in amplitude across the 20-min transmission (see Fig. 11). Assuming that the linear trend in phase corresponds to mooring motion, the slope is directly related to the mooring velocity. For example, the velocity inferred from Fig. 13 is -1 cm/s, which is reasonable given prior expectations about the ATOC moorings.

This example suggests a simple procedure for implementing mooring corrections. First, estimate the linear phase trend of the peak mode 1 arrival at the center frequency that has the most consistent amplitude across all lags. Use this to infer the velocity of the mooring. Then, given the mooring velocity and the initial VLA location, determine the position of the array elements at the lag times corresponding to each 4-period average. Finally, use the estimated array positions (hydrophone depths and ranges) to implement the mode processing for each 4-period average. Applying this procedure to all of the 28-Hz AST data yields good results. During the AST experiment, the estimated mooring velocities for the Hawaii VLA are between -2 and $+2$ cm/s, while the velocities of the Kiritimati VLA are between -1.5 and $+1.5$ cm/s. Since the 28- and 84-Hz transmissions were simultaneous, the mooring velocities obtained from the 28-Hz data can be used to compensate the 84-Hz receptions.

A similar approach can be applied to determine the mooring corrections for the ATOC receptions, though the amount of fading at 75 Hz makes it more difficult to find arrivals that are consistent across the extent of a 20-min transmission. The algorithm searches for peak arrivals that fade by less than 3 dB in amplitude over 20 min to use for the velocity estimation. The search begins with mode 1 in the 75-Hz bin. If no peaks in mode 1 meet the criteria, the program looks for peaks in modes 2 and 3. If no suitable peaks are found at 75 Hz, the search expands to the 72.5-, 70-, and 77.5-Hz bins. This procedure produced reasonable VLA motion estimates for both the Hawaii and Kiritimati arrays. Figure 14 shows a histogram of the mooring velocity estimates for the 188 good receptions in the HVLA75 data set. The estimates are concentrated between $+1$ and -1 cm/s.

While this approach to the VLA motion corrections is *ad hoc* and does not satisfy any optimality criteria, it does produce good results. In particular, as will be shown in the next section, this motion compensation algorithm leads to longer coherence times for the modes than were observed in previous work[40] with the same data sets.

IV. MODE STATISTICS

As the examples in the previous section illustrate, the low-mode signals are quite complicated at megameter ranges away from a broadband source. The frequency-selective fading observed in the mode arrivals is typical for propagation through time-varying multipath channels. In the communications literature, fading multipath channels are characterized by their statistics. The texts by Van Trees[41] and Proakis[42] provide useful introductions to statistical models for such channels. The purpose of this section is to compute the statistics of the ATOC and AST mode arrivals that are relevant to the future development of a channel model. Specifically, the following sections present estimates of the multipath time spread, coherence bandwidth, cross-mode coherence, and temporal coherence. Experimental results are compared to PE simulations.

A. Time spread

At megameter ranges, the low-mode signals consist of multiple arrivals spread over time. To quantify the multipath time spread, consider the average power as a function of arrival time, i.e.,

$$K_m(\tau) = \mathcal{E}[\alpha_m(\tau) * \alpha_m(\tau)], \tag{7}$$

where $\mathcal{E}$ is the expectation operator, $\alpha_m(\tau)$ is the time series for the mth mode, and $*$ indicates complex conjugation. Proakis[42] calls this quantity the multipath intensity profile of

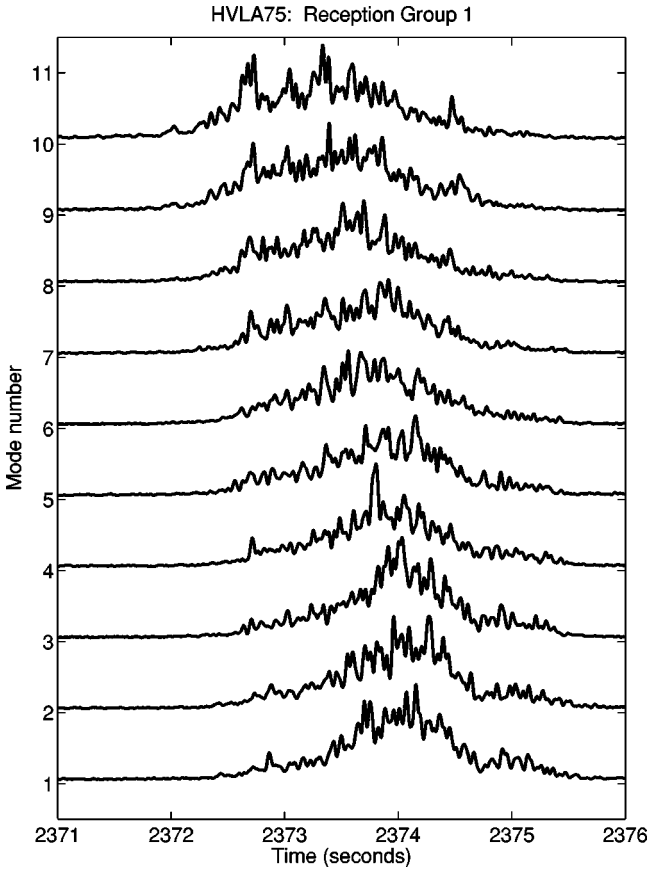


FIG. 15. Multipath intensity profiles for modes 1–10 at Hawaii computed using data from the first group of ATOC receptions.

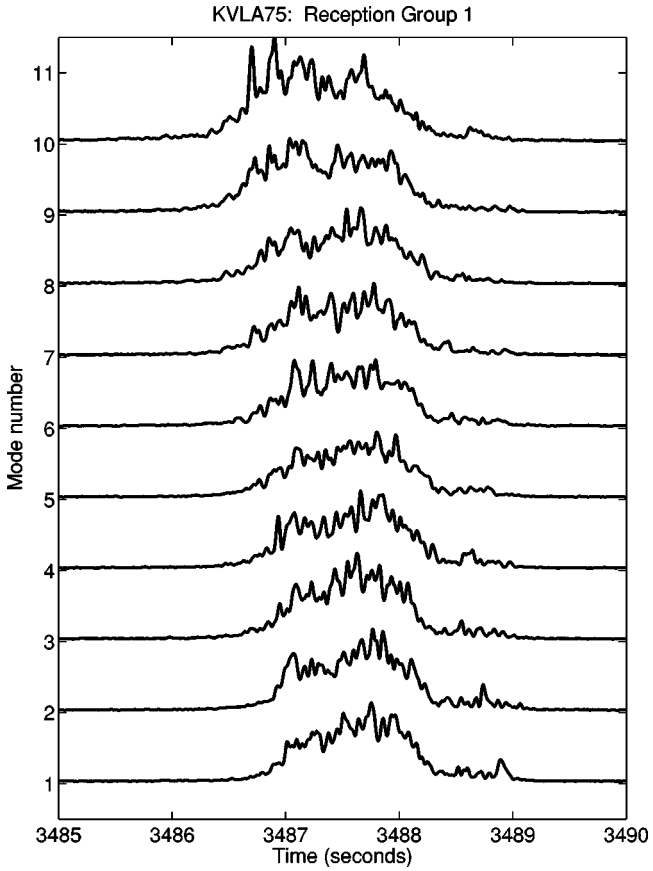


FIG. 16. Multipath intensity profiles for modes 1–10 at Kiritimati computed using data from the first group of ATOC receptions.

the channel. Assuming that the channel is relatively stationary over a period of days, $K_m(\tau)$ can be estimated by averaging over the FFT-processed mode estimates for a reception group (as defined in Table I). Figure 15 shows the multipath intensity profiles for the first 10 modes of reception group 1 in the HVLA75 data set. This plot highlights several characteristics of the low-mode arrivals at 3515-km range. First, the arrivals in modes 1–10 are spread over several seconds, with the higher modes having slightly larger time spreads than the lower modes within that group of 10. Second, the average arrival time is a function of mode number, e.g., the mean arrival time for mode 10 is smaller than that for mode 1. Third, the plot shows that the main arrivals in each mode are followed by the approximately 0.5 s of weaker arrivals that Sec. II attributes to bathymetric interaction near the Pioneer source.

To compare the 3515-km results with those at 5171 km, Fig. 16 shows the multipath intensity profile for the first 10 modes in reception group 1 of the KVLA75 data set. The plot shows that, similar to the Hawaii results, the arrivals in modes 1 through 10 are spread over several seconds. At Kiritimati the first 10 modes are more tightly grouped in arrival time than at Hawaii, i.e., there is less dispersion between the modes. Worcester *et al.*[21] note that the Pioneer-Kiritimati path is more oceanographically complex, particularly near the equator. Additional mode-to-mode scattering would tend to erase the mode-dependent trends in arrival time. Note that in this reception group, there appear to be a set of weaker arrivals trailing the main arrivals, particularly in modes 1 and 2. Interestingly, this is not a consistent feature in the KVLA75 data set. Many of the other reception groups show a much sharper cutoff. It is not clear whether these late arrivals can be attributed to bathymetric interaction near the source.

Figures 17 and 18 show how the multipath intensity profiles for mode 1 vary over the course of the ATOC experiment. A particularly strong shift in arrival time is evident between the first and second reception groups in each plot. These shifts are likely due to mesoscale activity since they are not in agreement with the predicted seasonal trends.[12,20]

Figure 19 shows the multipath intensity profiles for mode 1 measured at Hawaii and Kiritimati during the AST experiment. The plots contrast the results for the two AST source frequencies. In the Hawaii data there is a significant difference between the profiles for the 28-Hz source and the 84-Hz source. At 28 Hz, one arrival appears to dominate, whereas at 84 Hz the arrivals are much more spread out. At Kiritimati, the overall spread is roughly the same for the 28- and 84-Hz profiles, but the 28-Hz result indicates some distinct individual arrivals rather than the blur of arrivals seen in the 84-Hz profile.

The leading and falling edges for each mode can be defined as the minimum and maximum times, respectively, that the multipath intensity profiles exceed a threshold. To see how to pick the threshold, consider the example shown in Fig. 20. The plot shows the profile for mode 1 in group 13 of

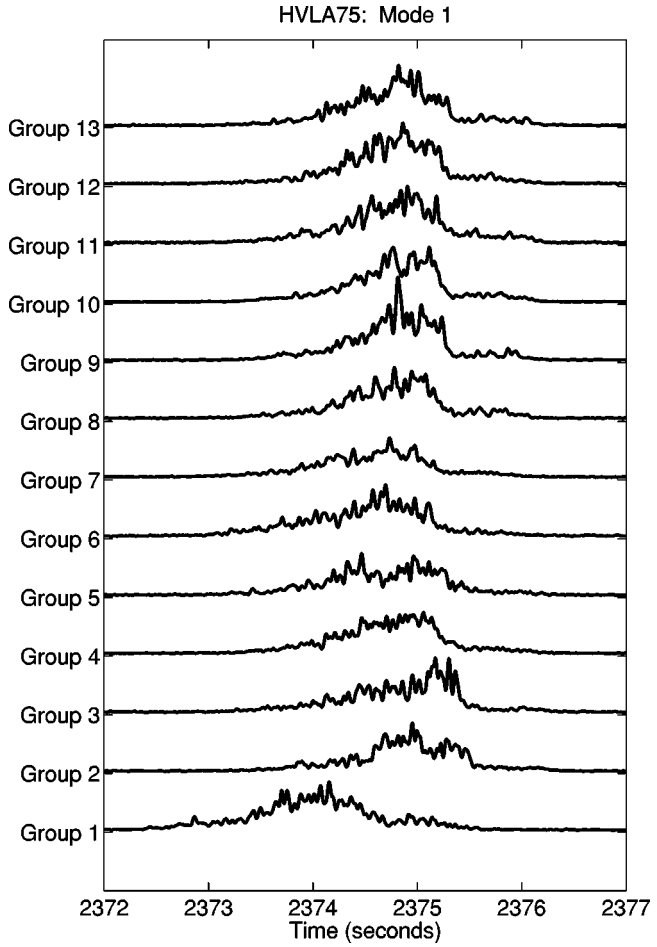


FIG. 17. Multipath intensity profiles for mode 1 for reception groups in the HVLA75 data set.

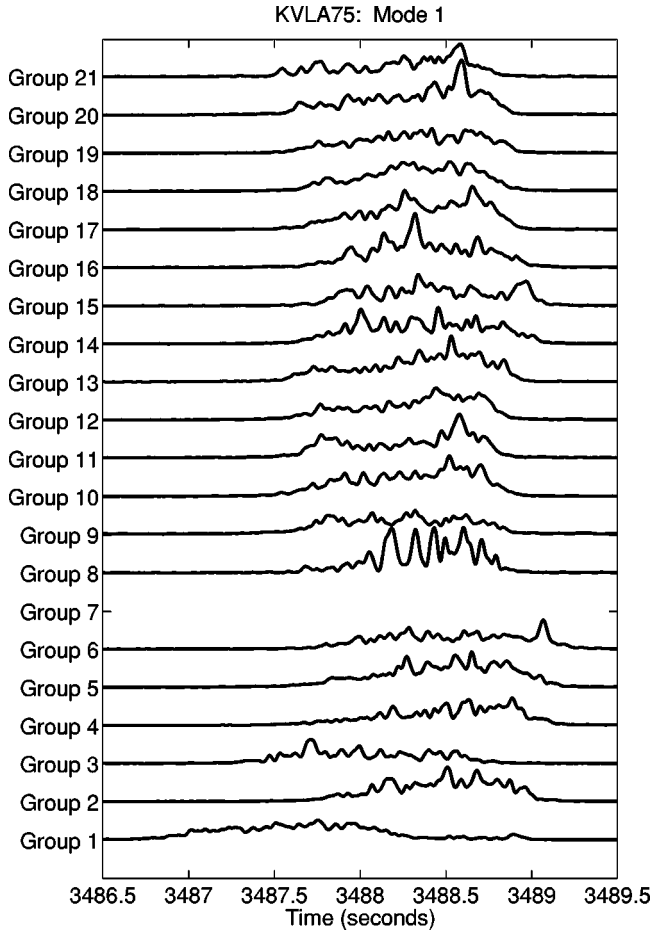


FIG. 18. Multipath intensity profiles for mode 1 for reception groups in the KVLA75 data set. Note that there are no good receptions at the Kiritimati VLA for reception group 7 (see Table I).

the HVLA75 data set. Dashed lines indicate the estimated noise floor, and the 3- and 8-dB thresholds above the noise floor. Note that the 8-dB threshold is high enough to avoid detecting the weaker arrivals that follow the primary arrivals in mode 1. A review of all reception groups indicates that an 8-dB threshold works reasonably well to detect the main arrivals and ignore the weaker trailing arrivals.

Figure 21 shows the mode time spreads, which are defined as the difference between the leading and falling edges determined by the 8-dB threshold. The HVLA75 and KVLA75 results shown in the plot represent the mean over all reception groups. For modes 1–10 in the HVLA75 data, the standard deviations of mean time spread lie between 0.13 and 0.3 s. The highest variance is associated with mode 1, which is due to the fact that the 8-dB threshold does not always suppress all of the weaker trailing arrivals. For the KVLA75 data, the standard deviations are somewhat smaller, with a minimum of 0.07 s and a maximum of 0.19 s.

Figure 21 shows that the average time spreads observed in ATOC and AST are on the order of 1.3 to 1.8 s. The spreads at the two ATOC receiver ranges (3515 and 5171 km) are quite comparable. Spread is a function of mode number, with the lowest modes showing somewhat less spread than the higher modes. Mode 1 in the HVLA28 data set shows the least amount of time spread, which is not surprising given that at 28 Hz the Hawaii mode 1 receptions are

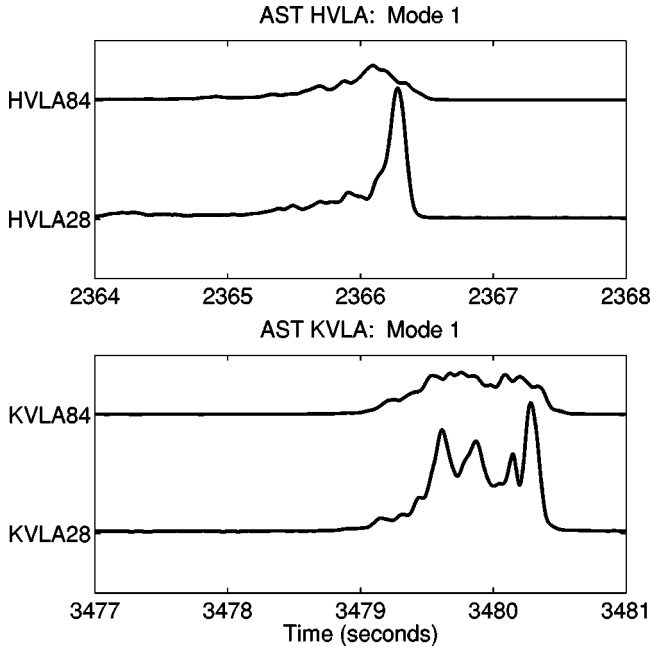


FIG. 19. Multipath intensity profiles for mode 1 at Hawaii and Kiritimati during the AST experiment. Both plots show a 4-s interval, but have different starting points.

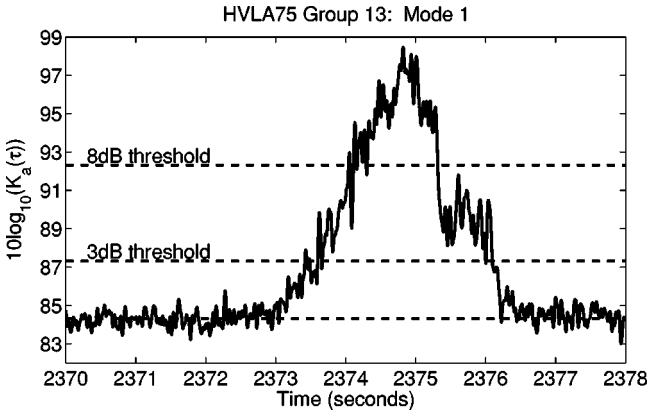


FIG. 20. Closeup of the multipath intensity profile for mode 1 in reception group 13 of the HVLA75 data set. Dashed lines mark the level of the noise floor, and the 3- and 8-dB thresholds above the noise floor.

dominated by one arrival. Mode 1 for the HVLA84 data set shows the most spread, but it is possible that this an artifact due to crosstalk in the processor designed for the degraded Hawaii VLA.

The multipath spread is a measure of the coherence bandwidth[41,42] of the channel, i.e., a time spread of T_m s implies a coherence bandwidth of $1/T_m$ Hz. With time spreads on the order of 1.5 s, the coherence bandwidth for the low modes is on the order of 0.67 Hz. This relatively short coherence bandwidth is consistent with the frequency-selective fading observed in the STFT mode estimates. The exception is the 28-Hz mode estimates, which show significantly less frequency-selective fading, even though their time spreads aren't substantially different. This indicates that at 28 Hz mode 1 is dominated by a few multipath arrivals that are temporally resolvable by the STFT processor (which has a resolution of 0.2 s).

B. Cross-mode coherence

In two seminal papers,[2,3] Dozier and Tappert predicted that scattering due to internal waves decorrelates the modes and results in an equipartitioning of energy at long ranges. Using the HVLA75 data, Wage *et al.*[20] showed that these

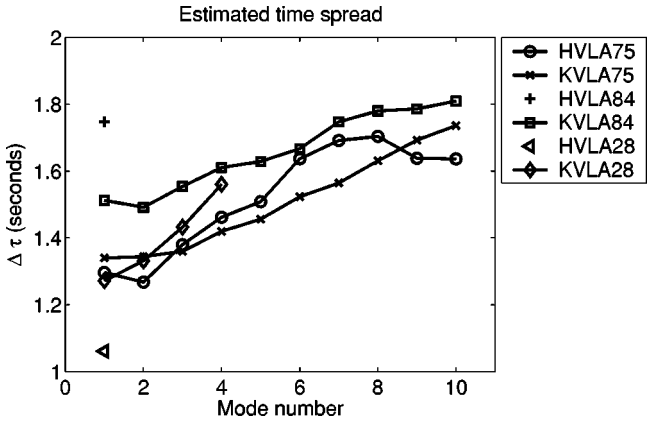


FIG. 21. Mode time spread for the ATOC and AST experiments. The HVLA75 and KVLA75 data represent the means over all reception groups in the experiment.

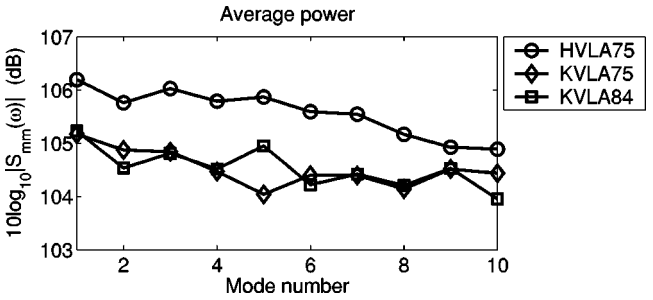


FIG. 22. Average power in the center frequency bin for modes 1–10 for the HVLA75, KVLA75, and KVLA84 data sets. The HVLA75 and KVLA75 results are for reception group 1.

predictions are valid for the lowest 10 modes at 3515 km. As discussed below, data from the Kiritimati array confirm that this result holds for a second, longer path.

A useful coherence metric is the magnitude-squared coherence (MSC), defined as:

$$\mathrm{MSC}(\omega)=\frac{|S_{mn}(\omega)|^2}{S_{mm}(\omega)S_{nn}(\omega)}, \tag{8}$$

where S_{mn} is the cross-power spectral density and S_{mm} and S_{nn} are the auto power spectral densities. By definition the MSC lies between 0 and 1. Carter[43] provides a thorough discussion of the statistical properties of the MSC. For the modes the cross-power spectrum can be estimated by averaging over L measurements in each STFT bin, e.g., an estimate of the cross spectrum for modes m and n is

$$\hat{S}_{mn}(\omega)=\frac{1}{L}\sum_{l=1}^{L}\hat{a}_m(\omega,l)\hat{a}_n^*(\omega,l), \tag{9}$$

where $\hat{a}_m$ and $\hat{a}_n$ are the amplitude estimates for modes m and n. Setting $m=n$ in Eq. (9) produces auto spectrum estimates. For the ATOC and AST data sets, the series of measurements needed to estimate the auto- and cross spectra are obtained by subsampling the output of the STFT processor. Specifically, samples are taken every 0.15 s, which corresponds to a 62.5% overlap of the 0.4-s processing window used in the STFT. The sampling interval is determined by the leading and falling edges defined by the 8-dB threshold, as discussed in Sec. IV A.

Using the HVLA75 data, Wage *et al.*[20] showed that the average power in modes 1–10 is approximately constant. Figure 22 demonstrates that this conclusion is also true for the Kiritimati propagation path. The plot shows the average power estimates ($10\log_{10}|\hat{S}_{mm}|$) for the center frequency bin in the HVLA75 (group 1), KVLA75 (group 1), and KVLA84 data sets. The average powers for the first 10 modes are equal within about 1 dB. Other bins and other reception groups show similar behavior. Dozier and Tappert's prediction of equipartioning of energy appears to be approximately true for the lowest 10 modes at megameter ranges.

Dozier and Tappert also predict that internal wave scattering will cause the modes to decorrelate. Again, the ATOC and AST data sets support this conclusion. For the HVLA75, KVLA75, and KVLA84 data sets, the maximum MSC between the low-order modes is 0.06. These results were obtained by averaging across all reception groups in each data

set. Note that the HVLA28, HVLA84, and KVLA28 data are excluded from this analysis because the diagonal loading required for mode processing introduces mode crosstalk, making mode cross-coherence estimation difficult. A set of PE simulations, which will be discussed more in the following section, also confirms that mode signals at 3.5 Mm or greater ranges are incoherent. The maximum cross-mode coherence for modes 1–10 estimated from narrow-band simulations at both 28- and 75 Hz is less than 0.1.

C. Temporal coherence

In previous work Wage *et al.*[20,40] demonstrated that the mode signals fluctuate significantly over time. Quantifying the temporal variability of these signals is an important prerequisite to developing a stochastic model for the mode arrivals. This section computes the coherence times of the low modes in the ATOC and AST data sets and compares them to estimates obtained from narrow-band PE simulations. Recall that each ATOC and AST transmission consisted of 40 M-sequences. The receiver recorded 10 four-period averages for each transmission. One way to quantify the temporal variability is to compute the MSC between the first four-period average in each transmission and successive four-period averages. In this study the MSC was estimated two ways: (1) using one bin of an FFT of the entire mode signal (between the leading and the falling edges), and (2) using samples from the desired bin of the STFT mode estimates (similar to how the cross-mode coherence estimation was implemented). Since the two approaches yield very similar results, only the STFT estimates are presented below. Note that in the following discussion the coherence time is defined as the lag time at which the MSC reaches 0.5. This threshold is somewhat arbitrary and may not be appropriate for all applications. Figures 23–26, which display the MSC as a function of lag time, can be used to determine the coherence time associated with any desired threshold.

Figure 23 shows the temporal coherence as a function of lag time for the 75-, 84-, and 28-Hz receptions at the two VLAs. The HVLA75 and KVLA75 results were obtained by averaging over all the reception groups listed in Table I. At 75- and 84 Hz, mode 1 decorrelates rather rapidly. The HVLA75 result shows the longest coherence time, reaching an MSC of 0.5 at approximately 8 min. Note that this is a slightly longer coherence time than previously reported; the previous result[20,40] was an MSC of 0.5 at 4.5 min. The difference is due to the revised mooring motion corrections discussed in Sec. III C. The new approach should do a better job of compensating for mooring-related phase changes. In addition, it is possible that this technique is removing a linear phase component associated with the internal wave fluctuations. The simulations discussed below shed more light on this issue.

Figure 23 also indicates that at 75 Hz the mode 1 signal at Kiritimati (5171-km range) is less coherent than at Hawaii (3515-km range). The MSC decays to 0.5 in approximately 6 min instead of 8 min. Both of the AST 84-Hz data sets show somewhat less coherence in mode 1 than the corresponding ATOC data sets. While this may be due to the increase in center frequency, it seems highly likely that residual source

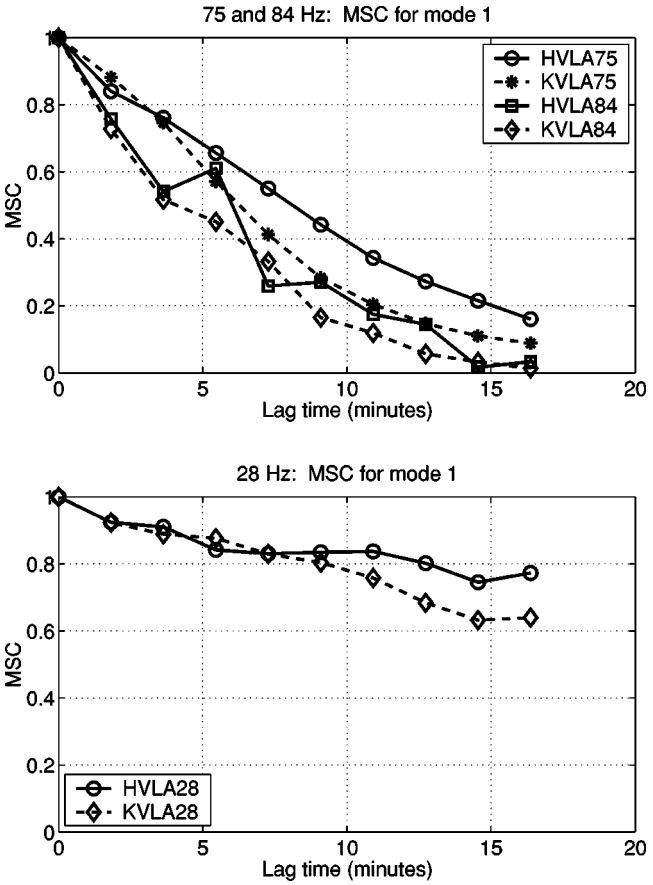


FIG. 23. Temporal coherence estimates for mode 1 in the ATOC and AST data sets. The ATOC results represent averages over all reception groups.

motion (not removed by the GPS-based correction) is a contributing factor. It is also important to remember that the mooring motion corrections for the 84-Hz receptions were derived from the peaks in the simultaneous 28-Hz receptions (since the lower frequency peaks are easier to track). Assuming that the temporal fluctuations at 28 Hz are independent of those at 84 Hz, the phase corrections derived from the 28-Hz data would not remove any of the internal-wave-induced variability at 84 Hz.

By comparison, the mode 1 signals at 28 Hz are much more coherent than at 75 Hz. Figure 23 also shows the MSC

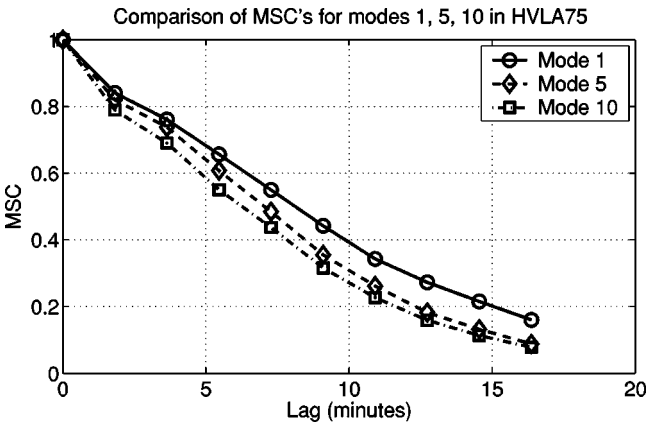


FIG. 24. Comparison of temporal coherence estimates for modes 1, 5, and 10 in the HVLA75 data set. Results are averages over all reception groups.

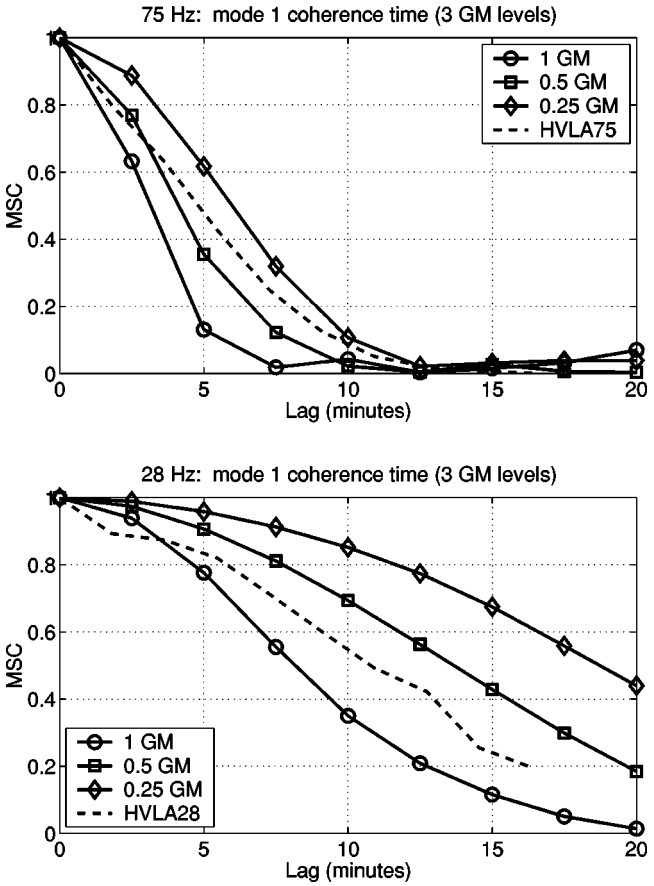


FIG. 25. Comparison of mode 1 coherence times for narrow-band PE simulations at 75- and 28 Hz with three GM levels: 1, 0.5, and 0.25 GM. The dashed line in each plot shows the Hawaii experimental results prior to the application of mooring corrections.

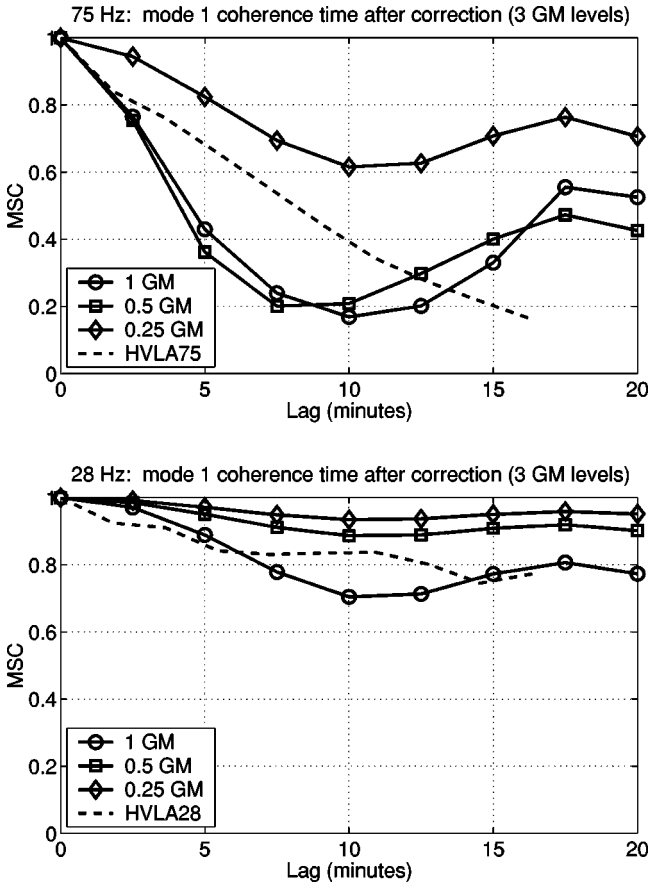


FIG. 26. Mode 1 coherence time for PE simulations at three GM levels after application of a linear phase correction. Subplots show results for 75- and 28 Hz. The dashed line in each plot shows the Hawaii experimental results after the application of mooring corrections.

as a function of lag time for the HVLA28 and KVLA28 data sets. As with the 75-Hz data, mode 1 appears to be slightly more coherent at the shorter (3515-km) range than at the longer range. At the maximum lag, the coherence is 0.77 for Hawaii and 0.64 for Kiritimati.

Figure 24 illustrates the general trend in coherence time as a function of mode number using the results for the HVLA75 data set. The plot shows the MSC for modes 1, 5, and 10. Mode 1 is the most coherent. The two higher-order modes decay to an MSC of 0.5 a few minutes earlier. This may be another consequence of how the mooring motion corrections were computed. The phase corrections were usually derived from mode 1 (assuming it had a peak that was trackable across all the four-period averages). Thus, it is possible that some of the internal wave variability associated with mode 1 is being removed, while this is not true for the other modes.

It is instructive to compare the experimental results to PE simulations of propagation through time-varying internal waves. The background environment for these simulations is identical to the one used in Sec. II C, and the suspended source is used. For this analysis, 30 realizations of sound-speed perturbations were generated for nine different lag times over a 20-min interval using the method of Colosi and Brown.[27] Using the perturbed environments, the RAM PE code[24] generated realizations of the narrow-band pressure field at 28- and 75 Hz for a range of 3515 km. Projecting the pressure field onto the mode functions yields the Fourier coefficient for each mode. These were averaged to obtain the auto- and cross spectra and finally the temporal coherence estimates.

Figure 25 shows the time coherence at 75 Hz for mode 1, derived from simulations using three different energy levels for the Garrett–Munk (GM) spectrum. As expected, the higher the GM strength, the lower the temporal coherence. Comparing Fig. 25 to the 75-Hz experimental results shown in Fig. 23 indicates that the measured mode 1 signals are more coherent than the simulations. Interestingly, the simulation results are comparable to the measured coherence for mode 1 obtained prior to applying the revised mooring corrections (shown as a dashed line in the figure). This suggests that these corrections, which are estimated by tracking arrival peaks, are removing more than just mooring motion.

Figure 25 also shows the simulation results for 28 Hz. As for the 75-Hz case, the temporal coherence for 28-Hz signals propagating through internal waves at 1 GM strength is lower than for those propagating through $\frac{1}{2}$ GM or $\frac{1}{4}$ GM internal waves. Also similar to 75 Hz, the 28-Hz simulations in Fig. 25 appear to be less coherent than the real data in Fig. 23. The dashed line in Fig. 25 indicates that if the mooring correction is not applied to the 28-Hz experimental data, the

estimated temporal coherence lies between the $\frac{1}{4}$ and $\frac{1}{2}$ GM simulation results.

The discrepancy between experimental and simulation results raises the possibility that the revised mooring corrections could be removing more than just mooring motion. To make a fairer comparison between experiment and simulation, a similar phase correction should be applied to the simulation data prior to coherence estimation. Since the simulations are narrow band, it is not possible to incorporate exactly the same type of phase correction as for the real data. Instead of using a linear fit to the phases of the detectable peaks in the STFT mode estimates, the phase correction for the simulated data is derived from a linear fit to the phase trend (across time) in the single frequency bin generated by the PE code. This removal of the linear trend is similar to what Colosi does for the finale coherence calculations.[19] Figure 26 shows the MSC estimates at 75- and 28 Hz, respectively, after the linear phase correction has been incorporated into the coherence calculation. For reference, the plots also include the results for the mooring-corrected experimental data.

At 28 Hz, there is good agreement between the phase-corrected simulations and the mooring-corrected data. The experimental data lies between the 0.5- and 1-GM results. At 75 Hz, the experimental measurement falls between the 0.25- and 0.5-GM results for most lag times; however, the trends in the simulated and experimental data don't really match. With the incorporation of the phase correction, the MSC is not monotonically decreasing. This is due to the fact that a linear fit to a nonlinear phase characteristic is being used for the correction. For many of the PE realizations, the least-squares solution for the phase trend results in a better fit near the endpoints than in the middle. Thus, the coherence decreases initially and then increases. The real data do not exhibit this behavior, but this is not surprising given that the phase corrections were implemented differently.

V. CONCLUSION

This article presented a detailed study of the mode arrivals observed in the ATOC and AST experiments. The analysis confirmed that the receptions at 5171 km share many of the same characteristics as the receptions at 3515 km analyzed previously. This study noted the difference in the cutoff of the ATOC and AST receptions. The simulations presented here suggest that bathymetric interaction near the Pioneer Seamount source is responsible for the generation of the weak, late arrivals observed in the ATOC receptions. Mode processing was implemented using both FFT and STFT techniques. The new approach to correcting for VLA motion worked well, resulting in longer mode coherence times than estimated in previous work. Simulations indicated that the simple phase correction for mooring motion likely removed some of the internal wave variability. Section IV provided a detailed set of mode statistics for the ATOC and AST data sets, including multipath intensity profiles, cross-mode coherence, and temporal coherence. At 75 Hz the mode time spreads are on the order of 1.5 s, giving a coherence bandwidth of 0.67 Hz. This explains the high degree of frequency-selective fading observed in short-time Fourier mode estimates. The time spread of the 28-Hz signals is somewhat lower, but these signals show significantly less frequency-selective fading. This suggests that at the lower frequency, the multipaths are temporally resolvable. At these ranges, the experimental analysis indicates that modes 1–10 have equal powers and that the coherence between these modes is effectively zero. Coherence times for mode 1 at 75 Hz are on the order of 8 min for the 3515-km range and 6 min for 5171-km range. At 28 Hz mode 1 is much more stable, with a magnitude-squared coherence of greater than 0.6 for the 20-min ATOC/AST transmission period.

This investigation raises several interesting issues for future study. First, Colosi *et al.*[19] computed similar coherence statistics for the total acoustic field (rather than splitting the signal into modes). A key question is, how do the mode statistics derived in this paper relate to the statistics of the overall field? The temporal coherence trends are quite similar,[44] but further work is needed to understand the relationship between these results. Second, it is intriguing that a simple correction designed to remove mooring motion also removed a nontrivial portion of the ocean variability, particularly in the 28-Hz case. This raises the question of whether more sophisticated signal processing can be designed to equalize the mode signals and produce observables for tomography.

ACKNOWLEDGMENTS

Kathleen Wage gratefully acknowledges the support of an Office of Naval Research Ocean Acoustics Entry-Level Faculty Award. Additionally, this research was supported by the Office of Naval Research (Grant N00014-95-1-0800 to APL-UW and Grant N00014-95-1-0589 to SIO) and by the Strategic Environmental Research and Development Program through Defense Advanced Research Project Agency Grant MDA972-93-1-0003. The two HLF-6A low-frequency acoustic sources used in the AST experiment were modified for operation at depths of up to 1300 m and a specialized handling system was developed for them as part of the Joint Environmental Test Initiative sponsored by SPAWAR PMW-182. The mode filters were designed using environmental profiles from the ECCO database, which is a contribution of the Consortium for Estimating the Circulation and Climate of the Ocean funded by the National Oceanographic Partnership Program. The authors acknowledge the significant contributions of Arthur Baggeroer, Theodore Birdsall, Kurt Metzger, Walter Munk, and Robert Spindel to the design and implementation of the ATOC and AST experiments. In addition they thank the anonymous reviewers for their suggestions.

[1] S. M. Flatté, R. Dashen, W. H. Munk, K. M. Watson, and F. Zachariasen, *Sound Transmission Through a Fluctuating Ocean* (Cambridge University Press, Cambridge, England, 1979).
[2] L. Dozier and F. Tappert, "Statistics of normal mode amplitudes in a random ocean. I. Theory," J. Acoust. Soc. Am. **63**, 353–365 (1978).
[3] L. Dozier and F. Tappert, "Statistics of normal mode amplitudes in a random ocean. II. Computations," J. Acoust. Soc. Am. **64**, 533–547 (1978).
[4] A. Sazontov, "Quasiclassical solution of the radiation transport equation in a scattering medium with regular refraction," Acoust. Phys. **42**, 487–494 (1996).

[5] A. Sazontov and V. Farfel, ''Fluctuation characteristics of the response of a horizontal array in a randomly inhomogenous ocean with short-time averaging,'' Sov. Phys. Acoust. **37**, 514–518 (1991).
[6] A. Sazontov and V. Farfel, ''Matched filtering of a narrowband pulse signal transmitted through a random waveguide channel,'' Sov. Phys. Acoust. **38**, 591–595 (1992).
[7] N. Vdovicheva, E. Gorodetskaya, A. Malekhanov, and A. Sazontov, ''Gain of a vertical antenna array in a randomly inhomogeneous oceanic waveguide,'' Acoust. Phys. **43**, 669–675 (1997).
[8] E. Gorodetskaya, A. Malekhanov, A. Sazontov, and V. Farfel, ''Effects of long-range propagation of sound in a random inhomogenous ocean on the gain loss of a horizontal antenna array,'' Acoust. Phys. **42**, 543–549 (1996).
[9] E. Y. Gorodetskaya, A. I. Malekhanov, A. G. Sazontov, and N. K. Vdovicheva, ''Deep-water acoustic coherence at long ranges: Theoretical prediction and effects on large-array signal processing,'' IEEE J. Ocean. Eng. **24**, 156–171 (1999).
[10] J. A. Colosi, S. M. Flatté, and C. Bracher, ''Internal-wave effects on 1000-km oceanic acoustic pulse propagation: Simulation and comparison with experiment,'' J. Acoust. Soc. Am. **96**, 452–468 (1994).
[11] J. A. Colosi and S. M. Flatté, ''Mode coupling by internal waves for multimegameter acoustic propagation in the ocean,'' J. Acoust. Soc. Am. **100**, 3607–3620 (1996).
[12] The ATOC Consortium, ''Ocean climate change: Comparison of acoustic tomography, satellite altimetry, and modeling,'' Science **281**, 1327–1332 (1998).
[13] P. F. Worcester, B. D. Cornuelle, M. A. Dzieciuch, W. H. Munk, B. M. Howe, J. A. Mercer, R. C. Spindel, J. A. Colosi, K. Metzger, T. G. Birdsall, and A. B. Baggeroer, ''A test of basin-scale acoustic thermometry using a large-aperture vertical array at 3250-km range in the eastern North Pacific Ocean,'' J. Acoust. Soc. Am. **105**, 3185–3201 (1999).
[14] B. D. Dushaw, B. M. Howe, J. A. Mercer, and R. C. Spindel, ''Multimegameter range acoustic data obtained by bottom-mounted hydrophone arrays for measurement of ocean temperature,'' IEEE J. Ocean. Eng. **24**, 202–214 (1999).
[15] K. D. Heaney and W. Kuperman, ''Very long-range source localization with a small vertical array,'' J. Acoust. Soc. Am. **104**, 2149–2159 (1998).
[16] M. Dzieciuch, P. Worcester, and W. Munk, ''Turning point filters: Analysis of sound propagation on a gyre-scale,'' J. Acoust. Soc. Am. **110**, 135–149 (2001).
[17] J. A. Colosi, E. K. Scheer, S. M. Flatté, B. D. Cornuelle, M. A. Dzieciuch, P. F. Worcester, B. M. Howe, J. A. Mercer, R. C. Spindel, K. Metzger, T. G. Birdsall, and A. B. Baggeroer, ''Comparisons of measured and predicted acoustic fluctuations for a 3250-km propagation experiment in the eastern North Pacific Ocean,'' J. Acoust. Soc. Am. **105**, 3202–3218 (1999).
[18] J. A. Colosi, F. Tappert, and M. A. Dzieciuch, ''Further analysis of intensity fluctuations from a 3252-km acoustic propagation experiment in the North Pacific Ocean,'' J. Acoust. Soc. Am. **110**, 163–169 (2001).
[19] J. A. Colosi, A. B. Baggeroer, B. D. Cornuelle, M. A. Dzieciuch, W. H. Munk, P. F. Worcester, B. D. Dushaw, B. M. Howe, J. A. Mercer, R. C. Spindel, T. G. Birdsall, K. Metzger, and A. M. G. Forbes, ''Analysis of multipath acoustic field variability and coherence for broadband basin-scale transmissions in the North Pacific Ocean,'' J. Acoust. Soc. Am. **117**, **1538–1564 (2005).**
[20] K. E. Wage, A. B. Baggeroer, and J. C. Preisig, ''Modal analysis of broadband acoustic receptions at 3515-km range in the North Pacific using short-time Fourier techniques,'' J. Acoust. Soc. Am. **113**, 801–817 (2003).
[21] P. F. Worcester, B. M. Howe, J. A. Mercer, M. A. Dzieciuch, and the Alternate Source Test (AST) Group, ''A Comparison of Long-Range Acoustic Propagation at Ultra-Low (28 Hz) and Very-Low (84 Hz) Frequencies,'' in *Proceedings of the U.S.-Russia Workshop on Experimental Underwater Acoustics*, edited by V. I. Talanov (Institute of Applied Physics, Russian Academy of Science, Nyzhny Novgorod, 2000), pp. 93–104.
[22] B. M. Howe, S. G. Anderson, A. Baggeroer, J. A. Colosi, K. R. Hardy, D. Horwitt, F. W. Karig, S. Leach, J. A. Mercer, J. Kurt Metzger, L. O. Olson, D. A. Peckham, D. A. Reddaway, R. R. Ryan, R. P. Stein, K. von der Heydt, J. D. Watson, S. L. Weslander, and P. F. Worcester, ''Instrumentation for the Acoustic Thermometry of Ocean Climate (ATOC) Prototype Pacific Ocean Network,'' in Proceedings of the 1995 IEEE/MTS Oceans Conference (San Diego, CA), pp. 1483–1500, October 1995.
[23] K. D. Heaney, ''Inverting for Source Location and Internal Wave Strength Using Long Range Ocean Acoustic Signals,'' Ph.D. thesis, University of California San Diego, 1997.
[24] M. D. Collins, *User's Guide for* RAM (Naval Research Laboratory, Washington, D.C.).
[25] S. Levitus and T. Boyer, *World Ocean Atlas 1994 Volume 4: Temperature*, 1994. NOAA Atlas NESDIS 4, Silver Springs, MD.
[26] S. Levitus, R. Burgett, and T. Boyer, *World Ocean Atlas 1994 Volume 3: Salinity*, 1994. NOAA Atlas NESDIS 3, Silver Springs, MD.
[27] J. A. Colosi and M. G. Brown, ''Efficient numerical simulation of stochastic internal-wave-induced sound-speed perturbation fields,'' J. Acoust. Soc. Am. **103**, 2232–2235 (1998).
[28] B. Sperry, H. Schmidt, and A. Baggeroer, ''Mode excitation for a source on a rough, elastic, sloping bottom,'' J. Acoust. Soc. Am. **101**, 3180 (1997).
[29] R. H. Ferris, ''Comparison of measured and calculated normal-mode amplitude functions for acoustic waves in shallow water,'' J. Acoust. Soc. Am. **52**, 981–988 (1972).
[30] F. Ingenito, ''Measurements of mode attenuation coefficients in shallow water,'' J. Acoust. Soc. Am. **53**, 858–863 (1973).
[31] C. Tindle, K. Guthrie, G. Bold, M. Johns, D. Jones, K. Dixon, and T. Birdsall, ''Measurements of the frequency dependence of normal modes,'' J. Acoust. Soc. Am. **64**, 1178–1185 (1978).
[32] J. R. Buck, J. C. Preisig, and K. E. Wage, ''A unified framework for mode filtering and the maximum *a posteriori* mode filter,'' J. Acoust. Soc. Am. **103**, 1813–1824 (1998).
[33] Wage (Ref. 20) discusses the mismatch issue in the context of designing filters for the ATOC HVLA75 data set.
[34] T. Yang, ''Broadband source localization and signature estimation,'' J. Acoust. Soc. Am. **93**, 1797–1806 (1993).
[35] C.-S. Chiu, C. W. Miller, and J. F. Lynch, ''Optimal modal beamforming of bandpass signals using an undersized sparse vertical hydrophone array: Theory and a shallow-water experiment,'' IEEE J. Ocean. Eng. **22**, 522–533 (1997).
[36] P. Casey, ''Mode Extraction for Long Range Underwater Acoustic Signals,'' Master's thesis, University of Auckland, February 1995.
[37] A. Voronovich, V. Goncharov, A. Y. Nikol'tsev, and Y. A. Chepurin, ''Comparative analysis of methods for the normal mode decomposition of a sound field in a waveguide: Numerical simulation and full-scale experiment,'' Sov. Phys. Acoust. **38**, 365–370 (1992).
[38] As indicated in Sec. II, the ATOC and AST experiments took place over a period of approximately 9 months. Temperature and salinity measurements were made at the deployment and the recovery of the VLAs, but no environmental measurements are available for the intervening months. This analysis uses monthly profiles from the ECCO database to derive the mode shapes and design the mode filters. ECCO stands for Estimating the Circulation and Climate of the Ocean. See ⟨http://ecco.ucsd.edu⟩ for more details.
[39] P. F. Worcester, personal communication (1995).
[40] K. E. Wage, A. B. Baggeroer, T. G. Birdsall, M. A. Dzieciuch, B. M. Howe, J. A. Mercer, K. Metzger, W. H. Munk, R. C. Spindel, and P. F. Worcester, ''A Comparative Study of Mode Arrivals at Megameter Ranges for 28 Hz, 75 Hz, and 84 Hz Sources,'' in Proceedings of the 2003 IEEE/MTS Oceans Conference, pp. 258–265, September 2003.
[41] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part III* (Wiley, New York, 1971).
[42] J. G. Proakis, *Digital Communications*, 4th ed. (McGraw-Hill, New York, 2001).
[43] G. C. Carter, ''Tutorial overview of coherence and time delay estimation,'' in *Coherence and Time Delay Estimation*, edited by G. C. Carter (IEEE, New York, 1993), pp. 1–27.
[44] Colosi plots magnitude coherence, rather than magnitude-squared coherence. To compare them directly, take the square root of the MSC results in Sec. IV.

Ocean acoustic wave propagation and ray method correspondence: Internal wave fine structure

Katherine C. Hegewisch, Nicholas R. Cerruti, and Steven Tomsovic
Department of Physics, Washington State University, Pullman, Washington 99164-2814

(Received 17 March 2004; revised 3 August 2004; accepted 4 August 2004)

Acoustic wave fields propagating long ranges through the ocean are refracted by the inhomogeneities in the ocean's sound speed profile. Intuitively, for a given acoustic source frequency, the inhomogeneities become ineffective at refracting the field beyond a certain fine scale determined by the acoustic wavelength. On the other hand, ray methods are sensitive to infinitely fine features. Thus, it is possible to complicate arbitrarily the ray dynamics, and yet have the wave field propagate unchanged. This feature raises doubts about the ray/wave correspondence. Given the importance of various analyses relying on ray methods, a proper model should, at a minimum, exclude all of the fine structure that does not significantly alter the propagated wave field when the correspondence to the ray dynamics is integral. We develop a simple, efficient, smoothing technique to be applied to the inhomogeneities—a low pass filtering performed in the spatial domain—and give a characterization of its necessary extent as a function of acoustic source frequency. We indicate how the smoothing improves the ray/wave correspondence, and show that the so-called ''ray chaos'' problem remains above a very low frequency ($\sim$15–25 Hz). [DOI: 10.1121/1.1854842]

PACS numbers: 43.30.Cq, 43.30.Ft, 43.20.Dk [AIT] Pages: 1582–1594

I. INTRODUCTION

As acoustic waves propagate long ranges through the deep ocean, they are refracted by inhomogeneities in the ocean's sound speed profile. Roughly speaking, in the earth's mid-latitudes, temperature and pressure effectively combine to form a wave guide in the depth coordinate that vertically confines the propagation.[1] In addition to this overall structure, the ocean behaves as a weakly turbulent medium[2] that multiply scatters the acoustic waves mainly in the forward direction. Whether one is intrinsically interested in waves propagating through weak turbulence or in the state of the ocean through tomography,[1] ray methods are relied upon at various stages and levels of complexity in the resulting experimental analyses.[3–5] It is therefore critical to understand the applicability and limits of these ray methods.

Ray methods can only capture the physics of refraction and reflection, unless a geometric theory of diffraction is explicitly added.[6,7] Intuitively, one expects refractive effects to dominate diffraction when sound speed inhomogeneities are larger than the acoustic wavelength of the source. On the other hand, due to their pointlike nature, rays are sensitive to structures at all scales. Thus, one should be suspicious of (nondiffractive) ray methods for models that have significant fine scale structures that are ineffective in refracting waves, but that fundamentally alter the rays themselves. Hence, certain fine scale structures in the model can be thought of as being physically irrelevant, i.e., having no influence on the wave propagation, and should be eliminated before applying a ray method analysis. The possibility of diffraction is very important, but should be dealt with separately and we do not discuss it further in this article.

Another serious challenge for the applicability of ray methods that has been recognized in the past fifteen years or so is the existence of ray chaos;[8] see also earlier work in the field of quantum chaos.[9–12] One typical argument goes that chaos introduces caustics, i.e., singularities in ray methods, at an exponentially increasing rate with propagation time (range). Ray methods must therefore break down on a logarithmically short propagation scale, which renders them essentially useless. A significant body of work has shown that this need not be the case, and methods can be developed which are accurate to much longer propagation scales.[13,14] Even so, detailed ray methods tend to become rather burdensome with the exponential proliferation of rays. Thus, resorting to statistical methods based on the chaotic properties of the rays is often attractive.

These two reservations about ray methods, the inclusion of physically irrelevant fine structures in the sound speed profile and ray chaos, have often been co-mingled. For example, it is possible to add very fine structure to a sound speed model that has no effect on propagating waves and yet generates chaotic rays as unstable as one wishes. Our point of view is that the two issues should be disentangled, necessarily beginning with the removal of the physically irrelevant fine structures, whose characterization depends on the acoustic wavelength. We will come back to the ray chaos question, but leave a more detailed and complete analysis for follow-up work to this paper.

Our purpose is, thus, to create a practical and easily implemented technique for smoothing inhomogeneities in a sound speed model, and to give prescriptions for the extent of smoothing needed as a function of source frequency. Toward these ends, it is not necessary to mimic a realistic ocean accurately with the model, but rather to include certain key features, such as a simple form for the waveguide confinement and the fluctuations due to the weak turbulence. It is more than sufficient to include scattering solely in the vertical spatial plane, to make the parabolic approximation[15] and

to neglect larger mesoscale structures. A simple ocean sound speed model uses Munk's canonical model[16] to account for large scale effects due to temperature, pressure and salinity, and an efficient implementation scheme by Colosi and Brown[17] to generate much smaller inhomogeneities due to the ocean's internal waves. Using their approach, the inhomogeneities have the statistics of the Garrett–Munk spectrum.[18] These features, though leading to a simplified model of the ocean, are more than adequate for investigating the length scale at which fluctuation features become important. Increased realism will be included in a future companion paper.[19]

The outline is as follows. In Sec. II, the ocean sound speed model and the method for acoustic propagation are presented. In the following section, we consider theoretical issues such as the convergence of the propagated wave field by asking the following question: "does adding more modes in the internal wave expansion cease altering the propagation beyond some maximum mode number?" In Sec. IV, a smoothing of the expression for the internal wave sound speed model is introduced. This smoothing is very similar to the application of a low-pass filter—it removes most of the structures in the sound speed model that are shorter than a certain scale—but it is done directly in the spatial domain so that ray methods can easily be applied. Sensibly, the optimal amount of smoothing necessary is a function of source frequency. We demonstrate the effects of smoothing on both the wave field propagation and on the phase space structures associated with the underlying ensemble of rays. This does not, in general, eliminate the consideration of ray chaos as the Lyapunov exponents are still positive, but it does remove a significant amount of the so-called "microfolding"[20] of the phase space structures. We discuss how this can markedly improve the ray/wave correspondence.

II. THE ACOUSTIC PROPAGATION MODEL

In a medium such as the ocean where density fluctuations are small, the wave equation accurately describes the acoustic waves in which we are interested. The governing equation is

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

where $\mathrm{Re}\{\Phi(\vec{r},t)\}$ is the acoustic pressure and $c(\vec{r},t)$ is the sound speed at a location $\vec{r}$ and time t. The mean sound speed is roughly 1.5 km/s and if we consider a water parcel, the sound passes through it far faster than any variation in $c(\vec{r},t)$ due to the internal waves; i.e. the internal waves travel several orders of magnitude more slowly than the acoustic waves. Hence, it is reasonable to 'freeze' the state of the ocean and let $c(\vec{r},t)=c(\vec{r})$.

In anticipation of treating long range propagation, we assume that the scattering in the azimuthal direction is negligible and the important components of the acoustic wave field propagation take place in two spatial dimensions with $\vec{r}=(z,r)$, where z is depth in the ocean and r is range from the source. Consider a constant frequency source, i.e., a pure sinusoidal source of angular frequency $\omega=2\pi f$ with frequency f, whose amplitude is constant in time. Then, the wave field has a frequency response, $\Phi_\omega(z,r)$, where $\Phi(z,r,t)=\Phi_\omega(z,r)e^{-i\omega t}$. With this assumption, the wave equation reduces to the Helmholtz equation in cylindrical coordinates

$$\nabla^2\Phi_\omega(z,r)+k^2(z,r)\Phi_\omega(z,r)=0, \tag{2}$$

where the wave number $k(z,r)=\omega/c(z,r)$.

A. The parabolic equation

For long range propagation, waves that propagate too steeply with respect to the horizontal strike the ocean bottom and are strongly attenuated. Since the surviving waves propagate at small angles with respect to the horizontal, a Fresnel approximation[15] is possible which expresses the acoustic frequency response as the product of an outgoing cylindrical wave, $e^{ik_0r}/\sqrt{r}$ and a slowly varying envelope function, $\Psi_\omega(z,r)$, where the horizontal wave number $k_0\approx\omega/c_0$. Thus,

$$\Phi_\omega(z,r)=\Psi_\omega(z,r)\frac{e^{ik_0(\omega)r}}{\sqrt{r}}. \tag{3}$$

Substituting Eq. (3) into Eq. (2) and dropping two small terms gives the parabolic equation

$$\frac{i}{k_0}\frac{\partial}{\partial r}\Psi_\omega(z,r)=-\frac{1}{2k_0^2}\frac{\partial^2}{\partial z^2}\Psi_\omega(z,r)+V(z,r)\Psi_\omega(z,r). \tag{4}$$

Since the sound speed can be decomposed into the reference sound speed, c_0, and fluctuations, δc, about the reference: $c(z,r)=c_0+\delta c(z,r)$ with $\delta c(z,r)\ll c_0$, the potential is approximated as follows:

$$V(z,r)=\frac{1}{2}\left(1-\left(\frac{c_0}{c(z,r)}\right)^2\right)\approx\frac{\delta c(z,r)}{c_0}. \tag{5}$$

In our calculations, we use the last form of Eq. (5) for simplicity. Notice that there is a direct analogy between this parabolic equation and the quantum mechanical Schrödinger equation through the substitutions: $t\to r$, $m\to 1$, and $\hbar\to 1/k_0$. We use a symmetric split-operator, fast-Fourier-transform method to propagate the wave field;[21,22] see Appendix A.

The two terms neglected on the right side of Eq. (4) are

$$\frac{1}{8k_0^2r^2}\Psi_\omega(z,r)+\frac{1}{2k_0^2}\frac{\partial^2}{\partial r^2}\Psi_\omega(z,r). \tag{6}$$

The first term falls off rapidly with range since $k_0r\gg 1$. The second term is dropped due to the parabolic approximation which assumes that for a slowly varying envelope function, the curvature is weak. Note that up to this point, we have also dropped other terms from the propagation equation in assuming negligible azimuthal scattering and negligible time dependence of the internal waves. See the discussion in Ref. 23 for more details on all of the terms that have been dropped and an order of magnitude estimate for the size of the various contributions.

B. Ocean sound speed model

A simple model for the speed of sound in the ocean consists of two main components. The first component of the model is an adiabatic, large scale behavior which is responsible for creating the ocean's "sound channel"—an effective wave guide for acoustic propagation in the deep ocean. This general behavior has a minimum sound speed at the sound channel axis, and varies slowly with latitude and season, with the sound channel axis moving toward the surface for higher latitudes and colder seasons. Mesoscale variability is neglected in this study. The second component of the model is local fluctuations in the sound speed due to the ocean's internal waves. These fluctuations are much smaller in magnitude than the wave guide confining behavior, but describe the range dependence. The model potential $V(z,r)$ takes the form

$$V(z,r)=\frac{\delta c(z,r)}{c_0}=\frac{\delta c_{\mathrm{wg}}(z)}{c_0}+\frac{\delta c_{\mathrm{iw}}(z,r)}{c_0}, \tag{7}$$

where δc_{wg} represents the change of the sound speed due to the wave guide, which we take to be range independent, and δc_{iw} represents the fluctuations due to internal waves.

1. The confinement/wave guide

In the ocean, the main effects of pressure, temperature, and salinity create a minimum in the sound speed. Since sound bends toward regions of lower velocity, the shape of the sound speed profile refracts propagating waves toward the sound channel axis. This effect is captured in a smooth, average model proposed by Munk[16] and is known as Munk's canonical model,

$$\frac{\delta c_{\mathrm{wg}}}{c_0}=\frac{B\gamma}{2}[e^{-\eta(z)}-1+\eta(z)], \tag{8}$$

where $\eta(z)=2[z-z_a]/B$, z_a is the sound channel axis, B is the thermocline depth scale giving the approximate width of the sound channel, and γ is a constant representing the overall strength of the confinement. This model has its minimum speed at $z=z_a$ and captures the right exponential and linear trends near the surface and bottom. The parameters are chosen to be $B=1.0$ km, $z_a=1.0$ km, $c_0=1.49$ km/s and $\gamma=0.0113$ km^{-1}, which are roughly consistent with the well known environmental measurements performed in the SLICE89 experiment.[24,25]

2. Internal wave sound speed fluctuations

Internal wave fluctuations perturb the sound speed in the ocean through the resultant vertical motions of water parcels. They are responsible for multiple, weak, forward scattering of acoustic waves. A numerical scheme has been introduced by Colosi and Brown,[17] which allows the efficient computation of a random ensemble of individual realizations of the typical sound speed fluctuations. This scheme conforms to the Garrett–Munk spectral and statistical phenomenological description of the internal waves[18,26] and has the form

$$\frac{\delta c_{\mathrm{iw}}}{c_0}=\sum_{j=1}^{J_{\max}}\sum_{k_r} e_{j,k_r}\exp\left(-\frac{3z}{2B}\right)\sin(j\pi\xi(z)), \tag{9}$$

where we took $\xi(z)=e^{-z/B}-e^{-H/B}$ with H the depth of the ocean. The prefactor e_{j,k_r} includes a random phase and magnitude factor for each j and k_r in the sum; see Appendix B for further details and to infer a definition of e_{j,k_r}. Since the frequency of vertical motions lie between the inertial frequency (due to the earth's rotation) and the buoyancy frequency (due to the local stratification), the sum over the horizontal wave vector k_r has terms representing the superposition of internal waves with wavelengths in the range of 1–100 km. A maximum for the j-summation has been chosen as $J_{\max}=180$, which gives structure down to the scale of roughly a meter. The modes, $\sin(j\pi\xi(z))$, are connected to the buoyancy profile which is assumed to have an exponential form. Although the form given in the text above for $\xi(z)$ does not vanish precisely at the surface, it is sufficient for our purposes.

The model should enforce that both the function δc_{iw} and its derivative vanish sufficiently smoothly at the surface. Thus, a surface filter is introduced which consists of multiplying Eq. (9) by a continuous function of depth with the properties that it vanishes above the ocean's surface, is unity below 200 m, and has continuous first and second derivatives. In this way, $\delta c_{\mathrm{iw}}/c_0$ and its derivative vanish at the surface and are fully, smoothly restored below 200 m. Since the upper 200 m of the ocean can be quite complex with storms, seasonal fluctuations and latitudinal variability, there is no simple, general sound speed model near the surface; the surface filter is adequate for our purposes. We will propagate waves for which very little energy will enter this region, and thus, little effect of this surface smoothing will be relevant. The specific form we have chosen for the surface filter is

$$g(z;z_{\mathrm{st}},\tau_{\mathrm{st}})=\begin{cases}0 & \text{for } z'\leqslant -1/2\\ h(z') & \text{for } |z'|\leqslant 1/2\\ 1 & \text{for } z'\geqslant 1/2\end{cases}, \tag{10}$$

where $z'=(z-z_{\mathrm{st}})/\tau_{\mathrm{st}}$, the width is $\tau_{\mathrm{st}}=200$ m, the center is $z_{\mathrm{st}}=\tau_{\mathrm{st}}/2=100$ m, and the smooth function in between is

$$h(z)=\frac{1}{2}+\frac{9}{16}\sin(\pi z)+\frac{1}{16}\sin(3\pi z). \tag{11}$$

C. Initial wave field

The parabolic equation requires an initial wave field $\Psi_\omega(z,r=0)$ as input, which can then be propagated to the desired range of interest. It is important to understand the connection between the initial wave field and the localized, continuous wave source. Typical sources can be thought of as point sources whose acoustic energy disperses broadly. However, due to the previously mentioned fact that all the steeply propagating waves are strongly attenuated, we can instead propagate only that wave energy moving sufficiently close to the horizontal (within a spread of angles from the sound channel axis) that would have avoided the ocean's surface and bottom. Restricting the propagation angles limits the size of the vertical wave vector and necessarily creates "uncertainty" in the location of the point source. For our

zpurposes, it is appropriate to choose $\Psi_\omega(z,0)$ to be a minimum uncertainty wave packet. This implies using a normalized Gaussian wave packet of the form

$$\Psi_\omega(z,0)=\left(\frac{1}{2\pi\sigma_z^2}\right)^{1/4}\exp\left(-\frac{(z-z_0)^2}{4\sigma_z^2}+ik_{0z}(z-z_0)\right), \tag{12}$$

where z_0 centers the field, σ_z is the standard deviation of the Gaussian intensity and k_{0z} gives the propagating field an initial wave number in the z-direction. In all our calculations, we set $k_{0z}=0$, which maximizes the horizontally propagating energy, and $z_0=z_a$, which centers the energy on the sound channel axis.

A Fourier transform of Eq. (12) yields a complex Gaussian distribution of initial vertical wave numbers, k_z, centered at k_{0z} with standard deviation in intensity, σ_k. Since σ_z^2 and σ_k^2 are the variances of the intensity and not the amplitude of the wave, their relation is $\sigma_z^2=1/4\sigma_k^2$. By a simple geometrical argument, a vertical wave number can be related to the horizontal wave number by $k_z=k_0\tan\theta$, where θ is the angle with respect to the sound channel axis. In the next subsection, it is seen that $p=\tan\theta$ is a generalized momentum for a classical ray corresponding to the wave. Classical rays with the maximum angle $\theta_{\max}$ just barely graze the surface or bottom, and thus, rays are limited in their vertical wave numbers. Yet, for Gaussian wave packets, all wave numbers are in principle present, though most are weighted negligibly by the tails. It is the width, σ_k, which determines if the wave contains wave numbers large enough for a substantial amount of the wave to hit the surface or the ocean floor. One can determine a proper Gaussian width, in order for only the Gaussian tails to pass the surface or bottom, in analogy with the limiting classical rays by letting the maximum classical wave number $k_0\tan\theta_{\max}$ correspond to three standard deviations out in the initial Gaussian wave number distribution, i.e., set $3\sigma_k=k_0\tan\theta_{\max}$. Then

$$\sigma_z^2=\frac{9}{4k_0^2\tan^2\theta_{\max}}. \tag{13}$$

The explicit dependence of σ_z on the angular frequency, ω, of the continuous wave source is realized using the approximate relation $k_0\approx\omega/c_0$.

The specific choice of $\theta_{\max}$ depends on the vertical confinement. For the background confinement in Eq. (8), those rays departing the sound channel axis with an angle of $\theta=\pi/18$ (10°) come within 150 m of the surface, and those with $\theta=\pi/15$ (12°) come within 30 m. The addition of internal waves to the sound speed model causes some rays to have a closer approach to the surface, so we will most often use $\theta_{\max}=10°$ in this paper.

D. The classical rays

From the parabolic equation, one can derive a Hamiltonian system of equations for the position, z, and generalized momentum, p, of the collection of rays corresponding to the wave propagation. The Hamiltonian is given by $H=p^2/2+V(z,r)$ and the potential is $V(z,r)=\delta c(z,r)/c_0$. The equations are

$$\begin{aligned}\frac{dz}{dr}&=\frac{\partial H}{\partial p}=p,\\ \frac{dp}{dr}&=-\frac{\partial H}{\partial z}=-\frac{\partial V(z,r)}{\partial z}.\end{aligned} \tag{14}$$

Since $dz/dr\approx\Delta z/\Delta r=\tan\theta$, the generalized momentum is $p=\tan\theta$. The classical action T is calculated by imposing the initial condition $T_0=0$ and using the relationship

$$\frac{dT}{dr}=p\frac{dz}{dr}-H. \tag{15}$$

Through the parabolic approximation, the classical action is directly related to the travel time, τ, of the acoustic waves, where $T=c_0\tau-r$.

The relevant rays to the wave propagation are those appropriate for a Gaussian wave packet,[14,27] which implies initial conditions in the neighborhood of (z_0,p_0). Since $k_{0z}=k_0p_0=0$ for the wave packet in Eq. (12), ray calculations are done in a neighborhood of $p_0=0$. However, z_0 is taken to be on the sound channel axis, z_a.

The addition of range dependent internal wave effects to the sound speed model causes the classical rays to be chaotic.[8] The stability matrix contains the information about whether the rays are unstable (chaotic) or not.[28] At a fixed r, one has

$$\begin{pmatrix}\delta p_r\\ \delta z_r\end{pmatrix}=Q_r\begin{pmatrix}\delta p_0\\ \delta z_0\end{pmatrix}, \tag{16}$$

where the stability matrix,

$$Q_r=\begin{pmatrix}q_{11}&q_{12}\\ q_{21}&q_{22}\end{pmatrix}=\begin{pmatrix}\left.\frac{\partial p_r}{\partial p_0}\right|_{z_0}&\left.\frac{\partial p_r}{\partial z_0}\right|_{p_0}\\ \left.\frac{\partial z_r}{\partial p_0}\right|_{z_0}&\left.\frac{\partial z_r}{\partial z_0}\right|_{p_0}\end{pmatrix}. \tag{17}$$

Elements of this matrix evolve according to

$$\frac{d}{dr}Q_r=K_rQ_r, \tag{18}$$

where Q_r at $r=0$ is the identity matrix, and

$$K_r=\begin{pmatrix}-\frac{\partial^2H}{\partial z_r\partial p_r}&-\frac{\partial^2H}{\partial z_r^2}\\ \frac{\partial^2H}{\partial p_r^2}&\frac{\partial^2H}{\partial z_r\partial p_r}\end{pmatrix}. \tag{19}$$

The system of differential equations, Eqs. (14), (15) and (18), are solved using a 4th order Runge–Kutta method (where we have taken $\Delta r=100$ m in all calculations).

The Lyapunov exponent, μ, is a measure of the rate at which the rays are deviating under small perturbations. The relationship between the Lyapunov exponent and the matrix Q_r comes through the trace (sum of the diagonal elements) of Q_r,

$$\mu\equiv\lim_{r\to\infty}\frac{1}{r}\ln|\mathrm{Tr}(Q_r)|. \tag{20}$$

If $|\mathrm{Tr}(Q_r)|$ grows exponentially, the Lyapunov exponent is nonvanishing and positive, and the corresponding trajectory is chaotic.

III. THEORETICAL CONSIDERATIONS

Wave propagation should become increasingly insensitive to smooth perturbations as the scale of the perturbations decreases to the order of the smallest wavelength in the source and beyond. This issue does not arise in the horizontal coordinate of the internal wave model in Eq. (9), since the fluctuation scales are much longer than the horizontal projections of typical source wavelengths. However, this is an issue for the vertical fluctuations since Eq. (9) is a weighted superposition of a large number of vertical internal wave modes and presumably contains more detail than is necessary for accurate wave propagation. There comes a point in the summation beyond which the vertical modes begin to add physically irrelevant features to the sound speed inhomogeneities for a given source frequency. To determine the transition point where this occurs requires an understanding of the minimum wavelength structure in the propagating wave field, and an understanding of the power spectrum of individual vertical internal wave modes. The transition point, though, is not the only issue since higher modes contain a mix of physically relevant and irrelevant structures. These issues as well as their interplay are discussed here.

A. The vertical mode number transition

Intuitively, the vertical structures in the sound speed model responsible for refracting the wave are those that are larger than the minimum vertical wavelength, $\lambda_{\min}$, in the initial wave packet. Expressions for $\lambda_{\min}$ can be obtained by using $\lambda = 2\pi/k$ and the previously noted geometrical relation $k_z = k_0 \tan\theta$,

$$\lambda_{\min} = \frac{2\pi}{k_0 \tan\theta_{\max}} = \frac{\lambda_0}{\tan\theta_{\max}} = \frac{c_0}{f \tan\theta_{\max}}. \tag{21}$$

Recall that the ocean waveguide forces $\theta_{\max}$ to be small so that the minimum vertical wavelength is always enhanced over the source wavelength, λ_0. For $\theta_{\max} = 10°$, this enhancement is roughly a factor of 6. As a practical example, we note that some of the experiments conducted by the Acoustic Engineering Test (AET)[3,4] use a broadband 75 Hz source. A pure 75 Hz source has a 20 m source wavelength. Thus, if the energy stripping due to the ocean surface and bottom is consistent with $\theta_{\max} = 10°$, then the wave propagation would have a minimum vertical wavelength scale of roughly 110 m.

The vertical structures in the sound speed model in Eq. (9) arise through the superposition of vertical modes of the form $e^{-3z/2B}\sin(j\pi(e^{-z/B} - e^{-H/B}))$. Since the argument of the sine is nonlinear, each vertical mode contributes different oscillation lengths at different depths. The monotonicity of the argument illustrates that each mode has a "chirped" structure, i.e., each mode oscillates more and more rapidly as the surface is approached. To make this more precise, an expansion of the argument of the sine reveals that the local oscillation length as function of depth is

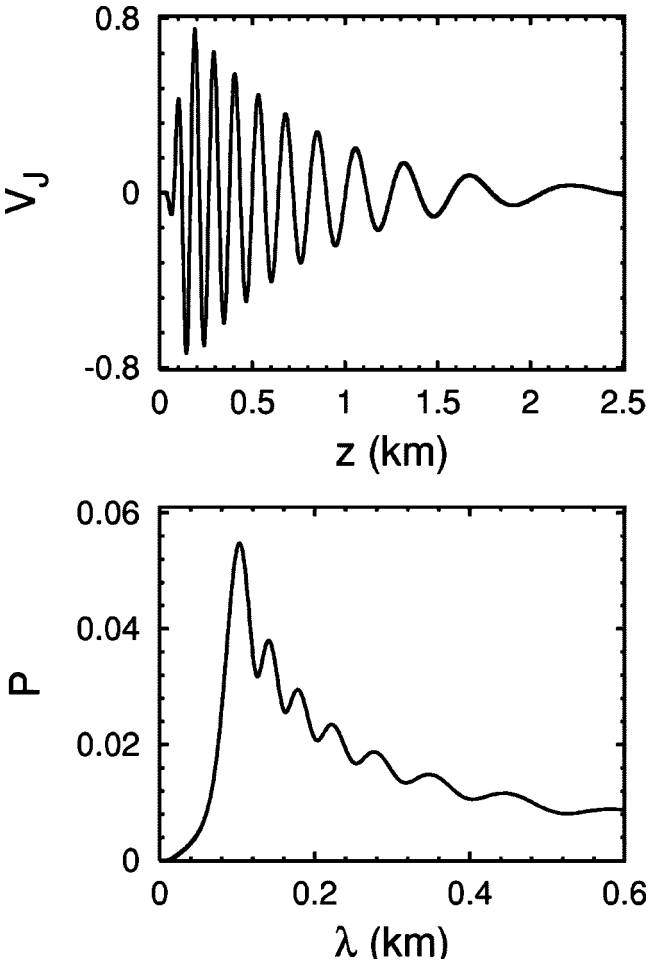


FIG. 1. Example of the single vertical internal wave mode for $j = 25$. The upper plot illustrates its depth dependence, $V_j = g(z; z_{\mathrm{st}}, \tau_{\mathrm{st}}) e^{-3z/2B} \sin(j\pi(e^{-z/B} - e^{-H/B}))$, where g is the surface filter defined in Eq. (10) and the lower plot is the power spectrum, P, of V_j.

$$\lambda_{\mathrm{iw}}(z, j) = \frac{2Be^{z/B}}{j}. \tag{22}$$

Therefore, the jth internal wave mode contributes its shortest length contribution of $2B/j$ near the surface, with longer length scales at increasing depth. Each mode gives contributions to the sound speed fluctuations over a broad range of scales.

Figure 1 illustrates the depth dependence and power spectrum of an internal wave mode. The power spectrum has a fairly sharp high frequency (short length scale) cutoff from the structures added near the ocean surface and a slowly decaying tail for the lower frequencies (longer length scales). The broad tail for an individual mode indicates that many different modes contribute to a particular size feature in the internal wave model.

The transition vertical mode number J_{trans} can be identified as that point where the vertical modes begin to introduce structure smaller than $\lambda_{\min}$. Thus, setting Eqs. (21) and (22) equal to each other and solving for j gives

$$J_{\mathrm{trans}} = \frac{2B \tan\theta_{\max}}{\lambda_0} = \frac{2Bf \tan\theta_{\max}}{c_0}. \tag{23}$$

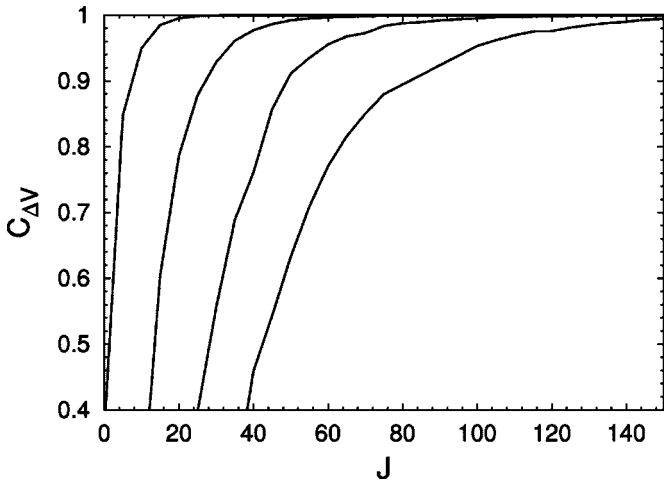


FIG. 2. $C_{\Delta V}(r)$ as a function of J for the source frequencies of 25, 75, 150, and 250 Hz (corresponding to the curves from left to right, respectively) at a range of 1000 km.

The calculation of J_{trans} does not reflect that each vertical mode is weighted in Eq. (9) by the coefficients e_{j,k_r}, which we numerically found to have root mean square decay $\sqrt{\Sigma_{k_r}|e_{j,k_r}|^2} \sim j^{-1.1}$ for large j. Thus, the higher vertical modes have a slowly decreasing weighting. The acid test of the effects of both the diminishing amplitudes and the detectability of features by the wave is to look at the sensitivity of the wave field to variations in the value for the j-summation cutoff in Eq. (9).

B. Wave field convergence

We can investigate the convergence of the wave field propagation by using different values for the j-summation cutoff in Eq. (9) to generate various sound speed media. The value of the cutoff leading to a converged wave field, denoted by J_ω, is the minimum number such that by including higher modes there is relatively little change in the wave propagation. We do not have a simple intuitive argument that gives an expression for J_ω, but instead rely on numerical simulations to determine reasonable values.

In order to discuss quantitatively what is meant by 'little change to the wave propagation', it is necessary to have a measure of the similarity of two wave fields. An ideal measure is the absolute value squared of the overlap (inner product). For two sound speed potentials that differ by ΔV, the quantity $C_{\Delta V}$ is defined as

$$C_{\Delta V}(r) = \left| \int \mathrm{d}z \Psi_\omega^{*\Delta V}(z,r) \Psi_\omega(z,r) \right|^2, \tag{24}$$

where $\Psi_\omega(z,r)$ is understood to be the wave field propagated to range r with the full potential and $\Psi_\omega^{*\Delta V}(z,r)$ is the same initial state propagated using the potential which differs from the full potential by ΔV. It is convenient to normalize the propagating wave fields to unity since this is preserved under the unitary propagation of the parabolic equation. With this choice, the measure gives unity only if the two wave fields are identical. The greater the reduction from unity, the greater the differences between the two propagations, i.e. the lower the faithfulness or fidelity of the propagations.

TABLE I. A comparison of key parameters for a few viable long range propagation frequencies. Both J_ω and λ_s^{opt} (see the next section) were determined using a conservative 0.99 criterion for the value of the $C_{\Delta V}$ at 1000 km in Figs. 2 and 6. Other choices for the criterion, propagation range, etc. could lead to somewhat greater differences; however the dependences are rather weak. For each calculation, $\theta_{\max}=10°$. Note the minimum wavelength feature, $\lambda_{\min}$, in the initial wave packet is extremely close to λ_s^{opt}.

Frequency (Hz)	J_{trans}	J_ω	λ_0 (km)	$\lambda_{\min}$ (km)	λ_s^{opt} (km)
25	6	20	0.060	0.340	0.308
75	18	50	0.020	0.113	0.106
150	36	90	0.010	0.056	0.060
250	60	145	0.006	0.034	0.034

The full wave propagation is compared to wave propagation for various values of mode number cutoff $J \leqslant J_{\max}$. Thus, ΔV is the internal wave sum for j in the interval $[J+1, J_{\max}]$. Since deviations of $C_{\Delta V}(r)$ from unity indicate that features in the modes $[J+1, J_{\max}]$ were detectable by the wave propagation, the value of J where $C_{\Delta V}(r)$ breaks appreciably from unity determines J_ω.

Sound waves with source frequencies of 25, 75, 150, and 250 Hz were propagated to $r=1000$ km; see Appendix A for details regarding the propagation. Figure 2 demonstrates the dependence of $C_{\Delta V}(r)$ on J. To interpret this figure, consider the curve for 75 Hz. Above $J=50$, $C_{\Delta V}(r) \geqslant 0.99$ and remains near unity. Thus, we can say that here $J_\omega \approx 50$. Using higher internal wave modes adds no more realism, and only slows down the calculations. A similar argument for the other frequencies gives the values of J_ω listed in Table I. Note that J_ω increases just a little more slowly than linear in source frequency due, in part, to the decreasing weightings.

Since $C_{\Delta V}$ is inherently range dependent, determining J_ω from a plot of $C_{\Delta V}$ for a single range is potentially inappropriate for longer ranges. Yet, Fig. 3 illustrates that the range dependence of $C_{\Delta V}$ is fairly weak for a frequency of 75 Hz. Increasing the range from 1000 km to 4000 km for $J_\omega=50$ decreases $C_{\Delta V}$ very little from 0.99 to 0.96. Thus,

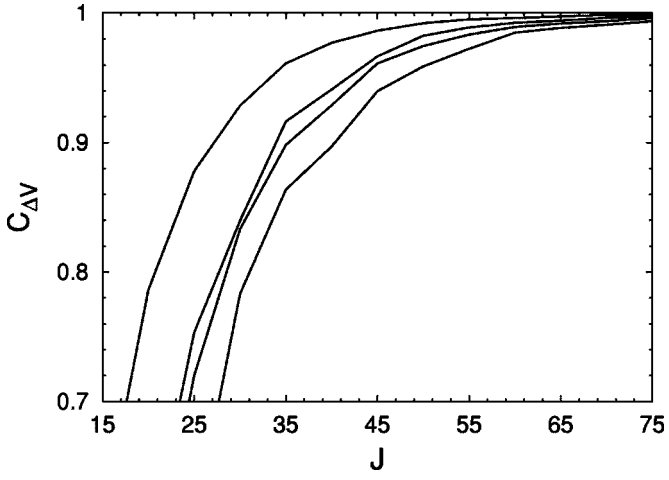


FIG. 3. $C_{\Delta V}(r)$, as a function of J for the ranges of 1000, 2000, 3000, and 4000 km (corresponding to the curves from left to right, respectively) for a frequency of 75 Hz.

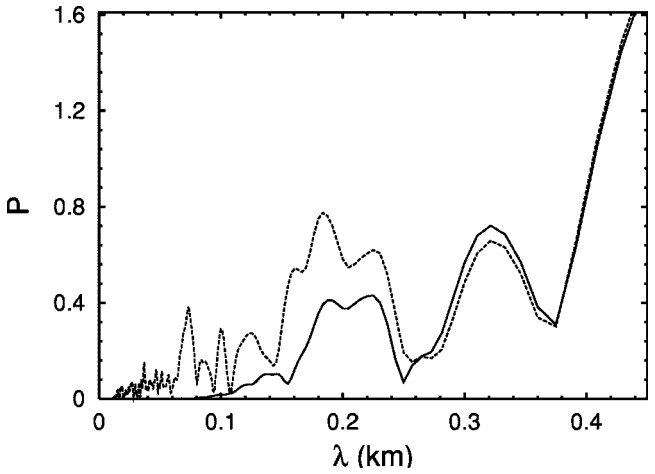


FIG. 4. Effects of the amount of smoothing on the power spectrum, P, of $\delta c_{\mathrm{iw}}/c_0$. The dashed line is the power spectrum of the unsmoothed full potential and the solid line is the power spectrum of the smoothed full potential for $\lambda_s = 0.20$ km.

$J_\omega = 50$ is a conservative choice even for ranges up to 4000 km.

For reasonable source frequencies, it is clear that even an optimal choice for J_ω will leave a significant amount of oscillations in the model on a scale much smaller than $\lambda_{\min}$ (since $J_\omega \gg J_{\mathrm{trans}}$). Presumably, these oscillations are physically irrelevant for the wave propagation, but it is worthwhile studying more precisely where the cutoff lies within the context of long range propagation.

IV. FILTERING THE PHYSICALLY IRRELEVANT FEATURES

Since we have taken a smooth background sound speed model, the physically irrelevant features of the sound speed model can be removed by filtering the high frequency components from the internal wave sound speed model, $\delta c_{\mathrm{iw}}(z,r)$. The ideal approach would be through the application of a low pass filter: Fourier transform $\delta c_{\mathrm{iw}}(z,r)$ for a fixed range to a frequency domain, apply a filter that removes the high frequencies and Fourier transform back to give the physically relevant portion of $\delta c_{\mathrm{iw}}(z,r)$. There are several drawbacks with respect to proceeding this way. The Fourier transforming back and forth is computationally expensive, creates a problematic ocean surface, and severely complicates the ray correspondence; the same would be true using a convolution technique. Instead, we develop a smoothing that can be directly applied to each vertical mode in the spatial z domain and serves as a very good approximation to a low-pass filtering in the frequency domain. It takes advantage of the monotonicity of the chirped structure of the individual internal wave modes. The spatial filtering method simplifies the ray equations enormously and allows first and second derivatives to be evaluated exactly, as opposed to numerically, which is an unstable operation.

A. The smoothing

Due to the precise oscillatory nature of each vertical mode, a good approximation to a low-pass filter can be accomplished by removing the upper portion of the vertical

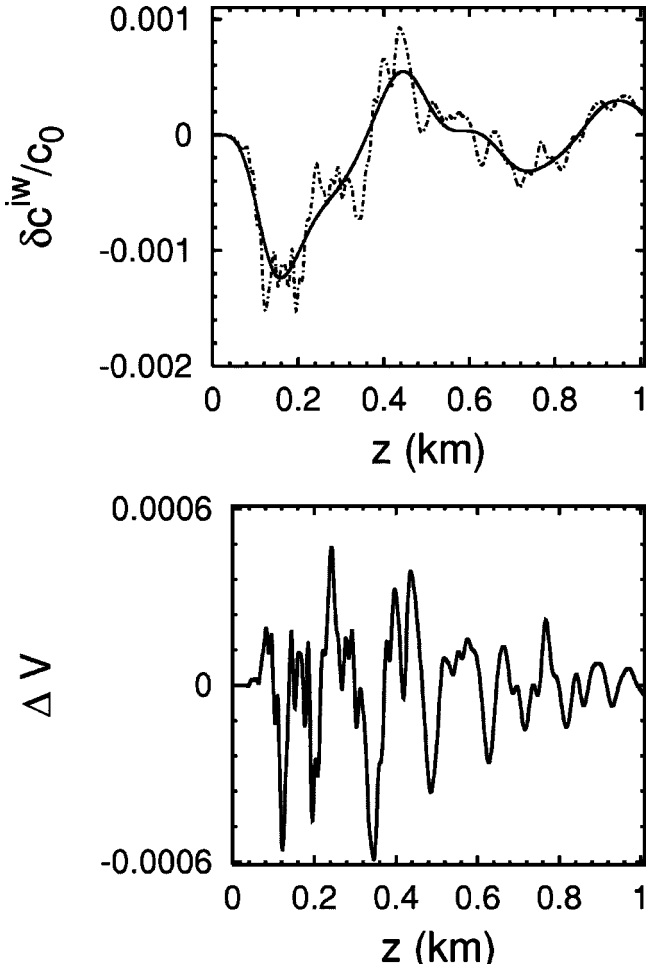


FIG. 5. Effects of the amount of smoothing on the full potential, $\delta c_{\mathrm{iw}}/c_0$. In the upper panel, the dashed line is the unsmoothed potential and the solid line is the smoothed potential for $\lambda_s = 0.20$ km. In the lower panel, the difference, ΔV, between the smoothed and unsmoothed potential is displayed.

mode that contains oscillations that are smaller than the smoothing parameter, λ_s. This involves multiplying each vertical mode by the function $g(z;z_{\mathrm{sm}},\tau_{\mathrm{sm}})$ defined in Eq. (10). This filter is centered at the depth such that the local length scale is λ_s, which by inversion of Eq. (22), gives the mode-dependent depth $z_{\mathrm{sm}} = B\ln(j\lambda_s/2B)$. Note that j must exceed $2B/\lambda_s$ in order for the filter to be below the ocean surface, which is where it begins to have an effect. This is consistent with the shortest length contribution of each mode being $2B/j$ at the surface. The width of the filter is carefully chosen to be $\tau_{\mathrm{sm}} = 2.0\lambda_s$ so that it does not cut off too sharply thereby introducing high frequency components into the model. If the width were chosen much greater, amplitudes of physically relevant length scales would be reduced.

Figure 4 shows the power spectrum of the sound speed model with and without smoothing; it is illustrated with a value, $\lambda_s = 0.2$ km. The power spectrum remains relatively unchanged for length scales greater than 0.2 km, but the length scales below 0.2 km are significantly dampened out of the model. This is evidence that a smoothing parameter of $\lambda_s = 0.2$ km is doing exactly what it was designed to do: it is filtering out features on scales below 0.2 km, but leaving features above 0.2 km in the model. Figure 5 shows the

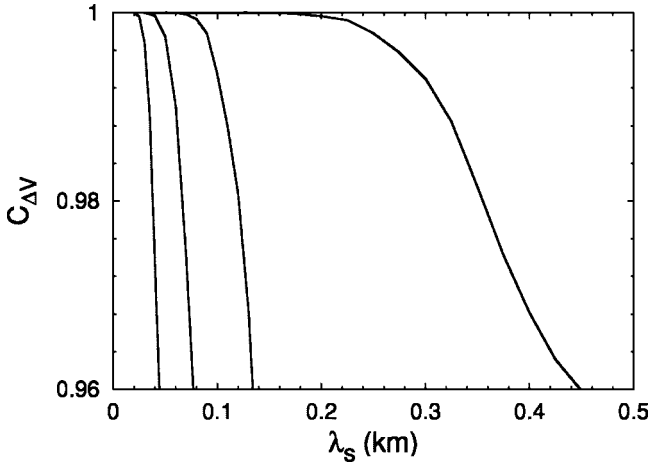


FIG. 6. $C_{\Delta V}(r)$ as a function of λ_s for source frequencies of 25, 75, 150, and 250 Hz (corresponding to the curves from right to left, respectively) at a range of 1000 km.

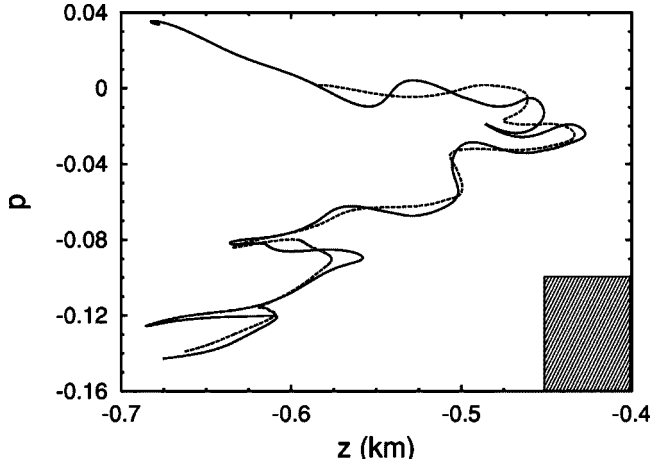


FIG. 7. Smoothed phase space manifold. The solid line is the phase space plot for a dense set of trajectories with launch angle $\theta \in [4°, 8°]$ propagated for 50 km in the unsmoothed ocean model. The dashed line is for the same set of trajectories, but for a smoothing parameter of $\lambda_s = 0.10$ km. All the trajectories started on the sound channel axis. The hatched rectangle is a reference area for physically irrelevant microfolds and has an area $\lambda_0/2\pi$, which corresponds to a 75 Hz source.

smoothed sound speed potential and the portion of the potential, ΔV, filtered by the smoothing. It is clear from these figures that the oscillations in the unsmoothed potential which have a length scale of less than 0.2 km have been removed, while larger oscillations have been preserved.

B. Estimating the optimal smoothing parameter

The optimal smoothing parameter, λ_s^{opt}, would be such that only those features in the model that are not detectable by the wave would be removed. Intuitively, λ_s^{opt} would be very close to $\lambda_{\min}$ of Eq. (21). In order to test this intuition, we again use $C_{\Delta V}(r)$ defined in Eq. (24), where here ΔV is the high frequency portion of the internal wave sum, which the smoothing removes, and the other potential is the full unsmoothed sound speed model. λ_s^{opt} is determined to be the maximum value of λ_s up to which $C_{\Delta V}$ remains nearly unity but deviates significantly beyond.

As in the previous section, source frequencies of 25, 75, 150, and 250 Hz were propagated to $r = 1000$ km with J chosen for each frequency to be that value of J_ω in Table I. Figure 6 demonstrates the dependence of $C_{\Delta V}(r)$ on different values of λ_s and its interpretation is similar to that done for Fig. 2. Consider the curve for 75 Hz. Above $\lambda_s \approx 0.1$ km, $C_{\Delta V}$ breaks significantly from unity giving the optimal smoothing of the sound speed model for a 75 Hz source to be $\lambda_s^{\text{opt}} \approx 0.1$ km. Smoothing less than this allows high frequency features to remain in the model which have no effect on the wave propagation. Table I summarizes the results which all agree closely with the intuitive idea that $\lambda_s^{\text{opt}} \approx \lambda_{\min} = \lambda_0/\tan\theta_{\max}$.

For a fixed λ_s, the higher source frequencies lead to a reduced value of $C_{\Delta V}$. This indicates that the high frequency components of ΔV are more detectable by a high frequency source than by a low frequency source. This fully supports the age-old intuitive concept that high frequency waves can detect smaller features than low frequency waves, and that the appropriate detection scale is a wavelength. A long range propagation experiment utilizing a source frequency f only detects that portion of the internal wave power spectrum with features longer than the length scale $\lambda_{\min} = c_0/f\tan\theta_{\max}$.

C. Effects of smoothing on phase space structures

Classical ray methods can be used to reconstruct propagating wave fields in detail through the use of semiclassical Greens functions.[29] The semiclassical approximation to the wave field is

$$\Psi_{\text{sc}}(z,r;k_0) = \sum_j A_j(z,r)\exp[ik_0 T_j(z,r) - i\pi\nu_j/2], \tag{25}$$

where the sum is over all ray paths labeled by j that begin at the source and end at a depth z for a given range r. The phase contribution of a path is related to its classical action, T_j, the source wave number, k_0, and the number of caustics, ν_j. The amplitude contribution of a path, A_j, is related to its stability matrix elements; see Ref. 30 for a readable account. This discrete set of paths becomes continuous if we consider all z. Thus, there is a continuous set of rays that underlies the full construction of $\Psi_{sc}(z,r;k_0)$ at a given range. A powerful analysis of the properties of this set comes by considering the rays in the phase space formed by all allowable points given by position and conjugate momentum. Viewed in phase space, the continuous set of rays underlying the wave field propagation (in the single degree of freedom problem being discussed here) forms a continuous, self-avoiding line which is called a manifold. As the range increases, the manifold evolves into a rather wild-looking "spaghetti". The more chaotic the system, the wilder the appearance of the manifold.

The construction of Eq. (25) relies on the use of stationary phase approximations, which can only be applied reliably when the phase between successive stationary phase points is greater than order unity. Care must be taken in defining the meaning of successive in this context. Thus, Eq. (25) breaks down when $[T_j(z,r) - T_{j'}(z,r)] < k_0^{-1} = \lambda_0/2\pi$ where j and j' are the classical paths/rays corresponding to successive stationary phase points. We term this the "area-$(\lambda_0/2\pi)$

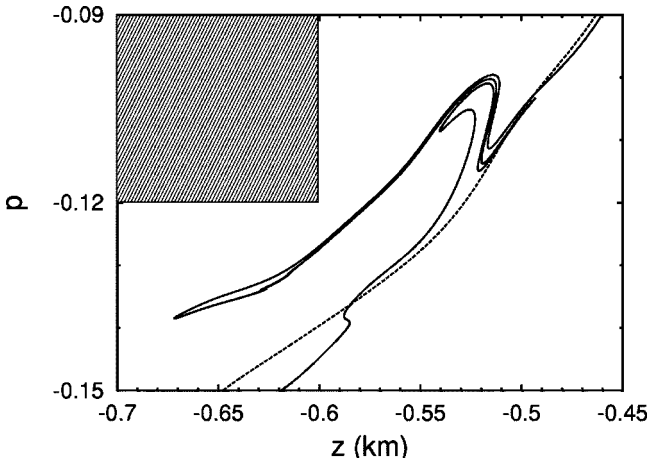


FIG. 8. Smoothing of the microfolds. The solid line is the phase space plot for a dense set of trajectories with launch angle $\theta \in [7°,8°]$ propagated for 100 km in the unsmoothed ocean model. The dashed line is for the same set of trajectories, but for a smoothing parameter of $\lambda_s = 0.10$ km. All the trajectories started on the sound channel axis. The hatched rectangle is a reference area for physically irrelevant microfolds and has an area $\lambda_0/2\pi$, corresponding to a 75 Hz source.

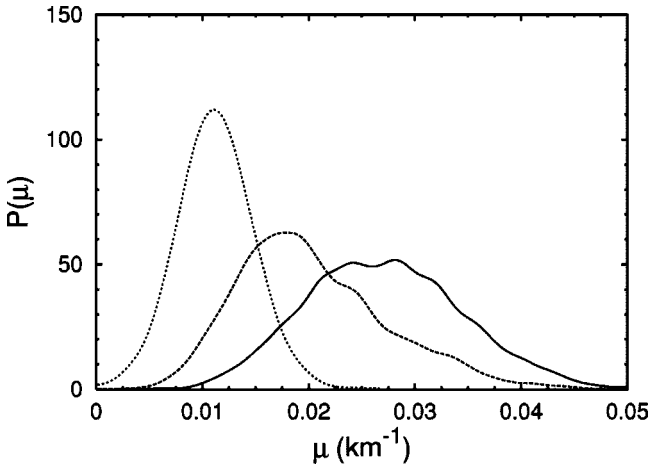


FIG. 9. Probability distribution of finite-range Lyapunov exponents. The range of propagation is 1000 km and each probability distribution consists of 4000 trajectories within a uniform distribution of launch angles. For the solid line, $|\theta| \in [0°,2°]$, for the dashed line, $|\theta| \in [4°,6°]$, and for the dotted line, $|\theta| \in [8°,10°]$. Each probability distribution was obtained by averaging over a Gaussian window of the corresponding histogram. The smoothing parameter is $\lambda_s = 0.10$ km and all the trajectories started on the sound channel axis.

rule'' (the translation to this problem of the area-$\hbar$ rule of Refs. 9 and 10). See Refs. 13 and 14 for a detailed presentation of the breakdown of the stationary phase approximation in quantum chaotic systems.

The breakdown of stationary phase is intimately related to how the manifold winds and folds its way through phase space. The difference in the classical action for two rays is related to the areas in phase space between the folds of the evolving manifold and the vertical line of the final depth, z, whose intersections with the manifold specify the rays. If these areas become smaller than $\lambda_0/2\pi$, then stationary phase breaks down for that pair of rays and we say that the two stationary phase points are coalescing. By drawing the manifold and filling in areas of $\lambda_0/2\pi$ in the folds, one can immediately see where problems, such as caustics which produce infinite amplitudes, will be occurring in the semiclassical construction. In the simplest case of two coalescing points, an Airy function uniformization is possible. However, if so many coalescing pairs occur that they cannot be isolated from each other, uniformization effectively is no longer possible, and the semiclassical approximation has broken down.

In the work of Simmen, Flatté, and Wang,[20] they show how the fine features in the internal wave field lead to a phenomenon they termed ''microfolding'' in which tiny folds are densely found along the manifold. Clearly, for typical source frequencies in long range propagation, the neighborhoods of the microfolds violating the area-$(\lambda_0/2\pi)$ rule overlap everywhere with each other. Thus, one anticipates a dense set of singularities in the semiclassical approximation rendering the approach useless.

Herein lies the advantage of smoothing the ocean sound speed model of physically irrelevant features before making the ray correspondence. Presumably, the bulk of the microfolding is related to fine features which are ignored by the wave propagation. The smoothed system contains only that structure necessary to describe the wave propagation so it should have fewer microfolds. Figure 7 illustrates the effects of smoothing on a set of trajectories. One can see that the smoothed manifold tracks the unsmoothed manifold along its length very well. A more detailed example of microfolding is illustrated in Fig. 8 for a range of 100 km. Notice how the smoothed manifold completely eliminates this particular microfolded structure for a smoothing parameter of $\lambda_s = 0.1$ km (appropriate for 75 Hz). Eleven, nonisolated pairs of coalescing stationary phase points were eliminated by the smoothing. Only a well behaved piece of the manifold with no coalescing pairs remains. Thus, there are fewer locations leading to singularities and breakdown in the semiclassical construction for the smoothed system, yet it is describing the same propagated wave. We leave the full semiclassical reconstruction for future work.

D. Effects of smoothing on Lyapunov exponent

The following question naturally poses itself from the results of the previous section, ''if smoothing the inhomogeneities reduces the number of folds, perhaps it is eliminating the ray chaos that was discovered in Ref. 8?'' This turns out not to be the case. The Lyapunov exponents for smoothed systems do not vanish. The Lyapunov exponent, μ, as defined in Eq. (20), requires the infinite range limit, which due to the maximum range of the ocean, is not very sensible. Instead, it is much more relevant to work with finite-range Lyapunov exponents.[31,32] The stability matrix, Q_r, as defined in Eq. (17), is calculated for a classical ray starting on the sound channel axis with an initial angle θ and propagated for a range r. If $|\mathrm{Tr}Q_r|$ is growing exponentially with range, then the ray is unstable or chaotic and the following relationship can be inverted to obtain the finite-range Lyapunov exponent:

$$|\mathrm{Tr}Q_r| = e^{\mu r} + e^{-\mu r}. \tag{26}$$

Excluding a few highly abstract systems, this μ fluctuates as a function of range and from one ray to the next. In fact, for

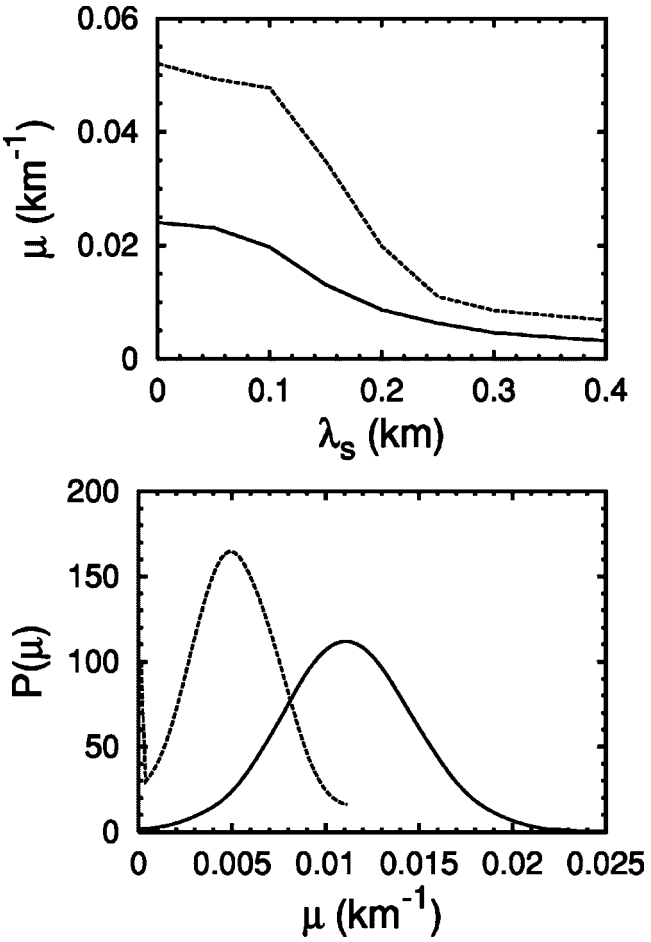


FIG. 10. Average Lyapunov exponents, μ_0 and $\overline{\mu}$, and the probability distribution as a function of the smoothing parameter, λ_s. Both plots are for a range of propagation of 1000 km. The upper plot is an average of 2000 trajectories within a uniform distribution of launch angle $\theta \in [-10°, 10°]$. The solid line is μ_0 and the dashed line is $\overline{\mu}$. The lower plot is the same as the previous plot except that $|\theta| \in [8°, 10°]$ in both curves and the smoothing parameter is varied. The solid line is for a smoothing parameter of $\lambda_s = 0.10$ km and the dashed line is for a smoothing parameter of $\lambda_s = 0.30$ km. A narrow peak near the origin exists in the dashed curve, which indicates a non-negligible fraction of stable trajectories.

typical chaotic systems and the internal wave problem here, $|\mathrm{Tr}Q_r|$ is close to being lognormally distributed, or from a different point of view, the finite-range Lyapunov exponents give something close to a Gaussian density.[31,32] The finite-range Lyapunov exponents are launch angle dependent.[33] Figure 9 shows histograms of the finite-range Lyapunov exponents for a range of 1000 km for a range of ray angles.

It turns out that the mean of the finite-range Lyapunov exponents is the usual infinite-limit Lyapunov exponent (as long as one has propagated beyond a transient range of a few Lyapunov lengths). Letting the brackets $\langle\ \rangle$ denote averaging over many rays,

$$\mu_0 = \frac{1}{r}\langle \ln|\mathrm{Tr}Q_r|\rangle. \tag{27}$$

If one averages before taking the natural logarithm, one gets a second stability exponent which is not the Lyapunov exponent, but rather a related one:

$$\overline{\mu} = \frac{1}{2r}\ln(\langle|\mathrm{Tr}Q_r|^2\rangle). \tag{28}$$

The relationship between μ_0 and $\overline{\mu}$ for a Gaussian density is through the variance of the distribution of the finite-range Lyapunov exponents,

$$\sigma_\mu^2 = \frac{\overline{\mu} - \mu_0}{r}. \tag{29}$$

These two stability exponents fix the Gaussian density completely. Figure 10 illustrates the dependence of μ_0, $\overline{\mu}$ and the distribution on the smoothing parameter λ_s.

Although, there is still ray chaos, the Lyapunov exponent is monotonically decreasing with increased smoothing, but unless a smoothing greater than 0.10 km is applied, μ_0 does not decrease appreciably. At some point, beyond a smoothing somewhere in the neighborhood of 0.3–0.5 km, a large fraction of the rays behave stably. Note in Fig. 10 that for $\lambda_s = 0.30$ km, a significant fraction of the rays have become stable, i.e., they have a Lyapunov exponent equal to zero. Using the relation between frequency and optimal smoothing, for source frequencies in the neighborhood of 15–25 Hz, there is a transition below which the ray chaos problem due to the internal wave inhomogeneities effectively disappears and above which it remains important over ocean basin scale propagation ranges. Though the background profile used for this study is somewhat simplistic, surprisingly these results seem to be consistent with some very low frequency experiments. In particular, data from the Alternate Source Test (AST) clearly shows that 28 Hz receptions have a more stable arrival pattern than the 84 Hz receptions for transmission over a 5000 km range.[34]

V. DISCUSSION

In probing the state of the ocean, it is important to understand what information is carried in the wave propagation for a given source frequency. Intuitively, fluctuations in the ocean sound speed on scales shorter than an acoustic wavelength should be ineffective sources of refraction for a sound wave in the ocean. Though, parabolic equation simulations are unaffected by the inclusion of physically irrelevant fine scale fluctuations in the sound speed model (except for the resulting slower computation time), this inclusion worsens the correspondence of ray methods to the wave propagation. On the other hand, ray methods are sensitive to infinitely fine scale structures. Those fine structures that are not detectable by the wave propagation lead to physically irrelevant microfolds in the phase space manifolds for the rays. These folds lead to unwanted singularities and the breakdown of semiclassical approximations. Smoothing of the internal wave sound speed model allows a significant reduction in the extent of microfolding and this must lead to a better ray/wave correspondence.

In our study, we noted that the chirped structure of each of the internal wave modes gives contributions to the sound speed fluctuations over a broad range of scales. Thus, limiting the number of vertical modes used in an internal wave sound speed model does not fully resolve the issue of physically irrelevant fine structure. For the specific construction of

Colosi and Brown, our calculations gave frequency dependent values for the number of vertical modes J_ω necessary in the model. For frequencies of {25,75,150,250} Hz, we found that the wave field propagation is essentially converged for $J_\omega=\{20,50,90,145\}$, respectively. However, for the same set of frequencies, modes greater than the transition modes $J_{\text{trans}}=\{6,18,36,60\}$, respectively, add structures on a finer scale than $\lambda_{\min}$. Hence, each mode contains a large spread of frequency contributions so that a low-pass filtering of each vertical mode is needed.

In order to remove physically irrelevant structures, we constructed an approximation in the position domain to a low-pass filter by taking advantage of the monotonicity of the chirped structure of each mode. The accuracy of this approximation (though not shown in this paper) was very good for individual modes. The spatial filtering method that we developed gives three main advantages: reducing required computations, better behavior in the neighborhood of the ocean's surface, and simplicity with respect to making the ray correspondence. With this study, it was found that the vertical scale of interest for the vertical fluctuations is not the source wavelength, $\lambda_0=c_0/f$, but rather the minimum vertical wavelength present in the wave field, which contains the additional projection factor $(\tan\theta_{\max})^{-1}$; see Eq. (21). $\theta_{\max}$ is the largest angle with respect to the horizontal that waves can propagate without being stripped out by bottom interactions and is typically in the neighborhood of $10°-12°$ in the ocean's mid-latitudes. For these values of $\theta_{\max}$ the minimum vertical wavelength is roughly 5–6 times λ_0; i.e., relevant vertical structures are much larger than that implied by λ_0.

Additionally, from the results in Table I, J_ω scales more slowly with increasing frequency than J_{trans}. This appears to be due to the decreasing weighting of the terms in Eq. (9), which directly influences the convergence of the wave field propagation and the value of J_ω. If this trend were to continue, then at a sufficiently high frequency, J_{trans} would overtake J_ω in value. Beyond this frequency, low-pass filtering would no longer serve any purpose; one could simply choose an appropriate J_ω. We do not attempt to extrapolate to that point here using our calculations and model, but note that wherever it is, the frequency would be so high that very long-range acoustic propagation would not be possible due to dissipation. However, it may be useful in the context of short range acoustic experiments using much higher frequencies to establish a cross-over frequency with a more realistic model.

We found that smoothing the internal wave sound speed fluctuations does not, in general, eliminate the problems associated with ray chaos. The Lyapunov exponents are positive and significant unless the smoothing scale exceeds 300–500 m. Thus, in this simplified model, ray chaos continues to be an important issue for source frequencies above the 15–25 Hz range.

A number of difficulties arise in the study of chaotic systems. For example, the exponential prolification of rays, makes it impractical to carry out ray methods. A common technique to overcome these difficulties is to apply various statistical methods whose justification derives from the chaos itself. However, even if you wish to apply these statistical methods, the validity of semiclassical techniques is still an issue.

Though it is known in the literature that the background sound speed profile can dramatically affect the complexity of the ray dynamics, it is still a question for investigation as to how significant these effects are on the wave propagation. Here we use Munk's canonical model as a simple, smooth background profile, which is sufficient for a study of the removal of physically irrelevant structures. However, before inferring detailed properties of long range experimental data, it would be good to employ a more realistic background sound speed wave guide. In fact, this would require a method for removing fine scale structures from the background in addition to the internal wave model and would not likely be subject to as simple a spatial filtering scheme as we used for the internal waves. We will address these issue in a forthcoming paper.

A number of interesting questions remain or emerge from our results. Our computations did not use pulsed sources, which can be expressed as an integral over a range of frequencies. It would seem reasonable to assume that the dynamics should be smoothed less for higher frequencies than for appreciably lower frequencies. How much attention must be paid to this issue? Can one make the crude approximation of using smoothing for the center frequency of a pulse?

In pulsed experiments, the early arrivals form branches which correspond to wave energy propagating at the larger angles near $\theta_{\max}$. Depending on the range of propagation, the late arrival portion of the signal may be constrained to a narrower range of propagation angles. Is more smoothing appropriate for this portion due to the $\theta_{\max}$ factor in $\lambda_{\min}$?

The measure $C_{\Delta V}$ is quite generally a function of range. Yet, we mainly used 1000 km propagation in our calculations to determine the optimal amount of smoothing and neglected the range dependence; we did note however a weak range dependence. Recall that several approximations are made in order to arrive at the parabolic equation or other one-way approximation variants. The neglected terms may also put range dependence in the propagation, and it would not make sense to try to be more accurate with the smoothing than the level of these other approximations. A more detailed understanding of the effects of neglected terms would be desirable.

Although there is significantly less microfolding for the smoothed than for the unsmoothed potentials, there is still uncertainty as to how much improvement is gained for the optimal smoothing. This could be made clear by carrying out the full detailed semiclassical construction and comparing it to the wave field propagation; we will carry this out in Ref. 19. A much deeper understanding would come from a full theory based on applying the area-$(\lambda_0/2\pi)$ rule discussed in Sec. IV C. It would give the most precise answers possible to questions of which structures are physically irrelevant and which method removes them in the most optimal way. We are pursuing this investigation because only by separating out the physically irrelevant fine structures can we begin to fully address the ray chaos conundrum and know whether it can be overcome.

ACKNOWLEDGMENTS

We gratefully acknowledge M. G. Brown for a critical reading of the manuscript, and for the support of the Office of Naval Research through Grant No. N00014-98-1-0079 and the National Science Foundation through Grant No. PHY-0098027.

APPENDIX A: THE SPLIT-OPERATOR, FAST FOURIER TRANSFORM METHOD

The parabolic equation in Eq. (4) describes the propagation of an acoustic wave with Hamiltonian $H=p^2/2+V$, where $p^2/2$ and V denote the kinetic and potential energies. A wave field can be advanced exactly through the application of the unitary propagation operator $\exp(-ik_0\int H dr)$. The split-operator Fourier transform method[22] approximates this operator using $e^{A+B}\approx e^{A/2}e^{B}e^{A/2}$, where A is taken to be $-ik_0\int(p^2/2)dr=-(i/k_0)\int(k^2/2)dr$ and B is taken to be $-ik_0\int V(z,r)dr$. Inserting a Fourier transform identity and rearranging terms before integrating gives a formula for the propagation of a wave field, $\Psi_\omega(z,r)$, at a range r to a wave field, $\Psi_\omega(z,r')$, at a range $r'=r+\Delta r$,

$$\Psi_\omega(z,r')=F^{-1}[e^{A/2}F[e^{B}F^{-1}[e^{A/2}F[\Psi_\omega(z,r)]]]], \tag{A1}$$

where F and F^{-1} are the forward and backward Fourier transforms, respectively. This equation has error $O((k_0\,\Delta r)^3)$ due to the operator approximation. We approximate the integral $\int_r^{r'}V(z,r)dr\approx\Delta r[V(z,r)+V(z,r')]/2$ and perform the integration $\int_r^{r'}(k^2/4)dr=\Delta r k^2/4$.

The wave fields in this paper are calculated over a vertical grid of $z\in[-2,7]$ km. The reflection boundary condition at the surface is not enforced in favor of the wave experiencing a soft reflection from the potential rather than a hard reflection from the surface. Wave energy which is reflected from the surface is eventually absorbed by the bottom in long range propagations so that this energy is negligible at a receiver. The soft reflections of the wave are due only to the background portion of the potential [Munk's canonical model in Eq. (8)] whose effects have been extended above the surface, $z<0$. The internal wave fluctuations from Eq. (9) are cut off by the surface filter in Eq. (10) so that they do not have an effect on the wave above the surface.

The grid size for the propagation is chosen to be dependent on the source frequency (to ensure proper sampling of the source in the horizontal and vertical directions) and the maximum number of vertical modes, J (to ensure proper sampling of the smallest wavelengths in each vertical mode). The grid number in the depth direction is purposely chosen to be a power of 2 to allow the use of a fast Fourier transform for the split-operator Fourier method. Specifically, for the source frequencies 25,75,150,250 Hz, we chose $\Delta r=0.01,0.01,0.005,0.0025$ km and $\Delta z=9/N$ km, where $N=1024,2048,2048,2048$, respectively. These values are large enough to guarantee proper convergence of the split-operator method for the propagation.

APPENDIX B: IMPLEMENTATION OF INTERNAL WAVE SOUND SPEED MODEL

The efficient numerical scheme devised by Colosi and Brown[17] generates a random ensemble of internal wave effects for the sound speed model, $\delta c_{\mathrm{iw}}(z,r)/c_0$, through the following equation:

$$\frac{\delta c_{\mathrm{iw}}(z,r)}{c_0}=\frac{24.5}{g}\frac{2B}{\pi}N_0^2\sqrt{\frac{E\Delta k_r}{M}}\exp(-3z/2B)\times\sum_{j=1}^{J_{\max}}\sum_{k_r=k_{\min}}^{k_{\max}}\sin(j\pi\xi(z))\times\sqrt{\frac{I_{j,k_r}}{j^2+j_*^2}}\cos(\phi_{j,k_r}+k_r r) \tag{B1}$$

where

$$k_j I_{j,k_r}=\frac{1}{\beta^2+1}+\frac{1}{2}\frac{\beta^2}{(\beta^2+1)^{3/2}}\ln\left(\frac{\sqrt{\beta^2+1}+1}{\sqrt{\beta^2+1}-1}\right). \tag{B2}$$

A single random seed generates the random phases, $\phi_{j,k_r}\in[0,2\pi)$, for each internal wave with vertical mode, j, and horizontal wave number, k_r. These random phases give the ocean a different internal wave realization for each random seed. All calculations in this paper were done with a single realization of the internal wave field, but all results are similar for averages over an ensemble of internal wave fields as well. Each internal wave in the superposition has the statistics of the Garrett–Munk spectrum. The full Garrett–Munk energy of $E=6.3\times10^{-5}$ has been used in all calculations. Our calculations are done for a latitude of 30° so that the inertial frequency is $f_i=1$ cycle per day. The buoyancy profile is assumed to have the form $N(z)=N_0e^{-z/B}$, where $N_0=1$ cycle per 10 min is the buoyancy frequency at the surface. We considered the depth of the ocean to be $H=5.0$ km, even though we extended the propagation range to the region $[-2,7]$ km for the reasons described in Appendix A.

The particular functional forms and constants used in this paper are as used by Colosi and Brown. Some of these forms and constants have already been identified in the body of the paper [i.e., near Eq. (9)], while the others are listed here. We took the gravitational acceleration $g=9.81$ m/s^2, $M=(\pi j_*-1)/2j_*^2$ and the principle mode number $j_*=3$. We took 512 horizontal internal wave numbers equally spaced by Δk_r for $k_r\in 2\pi[0.01,1.0]$ cycles per km. In the expression for I_{j,k_r}, we took $k_j=f_i\pi j/N_0B$ and the ratio $\beta=k_r/k_j$.

[1] W. H. Munk, P. F. Worcester, and C. Wunsch, *Ocean Acoustic Tomography* (Cambridge University Press, Cambridge, 1995).
[2] Y. V. Lvov and E. G. Tabak, "Hamiltonian formalism and the Garrett–Munk spectrum of internal waves in the ocean," Phys. Rev. Lett. **87**, 168501 (2001).
[3] P. F. Worcester, B. D. Cornuelle, M. A. Dzieciuch, W. H. Munk, B. M. Howe, J. A. Mercer, R. C. Spindel, J. A. Colosi, K. Metzger, T. G. Birdsall, and A. B. Baggeroer, "A test of basic-scale acoustic thermometry using a large-aperture vertical array at 3250-km range in the eastern North Pacific Ocean," J. Acoust. Soc. Am. **105**, 3185–3201 (1999).
[4] J. A. Colosi, E. K. Scheer, S. M. Flatté, B. D. Cornuelle, M. A. Dzieciuch, W. H. Munk, P. F. Worcester, B. M. Howe, J. A. Mercer, R. C. Spindel, K. Metzger, T. G. Birdsall, and A. B. Baggeroer, "Comparisons of measured

and predicted acoustic fluctuations for a 3250-km propagation experiment in the eastern North Pacific Ocean,'' J. Acoust. Soc. Am. **105**, 3202–3218 (1999).
[5] B. D. Dushaw, G. Bold, C.-S. Chui, J. A. Colosi, B. D. Cornuelle, Y. Desaubies, M. A. Dzieciuch, A. Forbes, F. Gaillard, J. Gould, B. M. Howe, M. Lawrence, J. Lynch, D. Menemenlis, J. Mercer, P. Mikhaelvsky, W. H. Munk, I. Nakano, F. Schott, U. Send, R. Spindel, T. Terre, P. F. Worcester, and C. Wunsch, ''Observing the ocean in the 2000's: A strategy for the role of acoustic tomography in ocean climate observation,'' in *Observing the Oceans in the 21st Century*, edited by C. J. Koblinsky and N. R. Smith, GODAE Project Office and Bureau of Meteorology, Melbourne, 2001, pp. 391–418.
[6] J. B. Keller, ''A Geometrical Theory of Diffraction,'' in *Calculus of Variations and its Applications, Proceedings of Symposia in Applied Mathematics*, edited by L. M. Graves (McGraw-Hill, New York, 1958), Vol. 8, pp. 27–51.
[7] J. B. Keller, ''Geometrical Theory of Diffraction,'' J. Opt. Soc. Am. **52**, 116–130 (1962).
[8] D. R. Palmer, M. G. Brown, F. D. Tappert, and H. F. Bezdek, ''Classical chaos in nonseparable wave propagation problems,'' Geophys. Res. Lett. **15**, 569–572 (1988).
[9] M. V. Berry, N. L. Balazs, M. Tabor, and A. Voros, ''Quantum Maps,'' Ann. Phys. (N.Y.) **122**, 26–63 (1979).
[10] M. V. Berry and N. L. Balazs, ''Evolution of semiclassical quantum states in phase space,'' J. Phys. A **12**, 625–642 (1979).
[11] G. M. Zaslavsky, ''Stochasticity in quantum systems,'' Phys. Rep. **80**, 157–250 (1981).
[12] G. P. Berman and G. M. Zaslavsky, ''Condition of stochasticity of quantum nonlinear systems,'' Phys. Lett. A **91**, 450–460 (1978).
[13] M. A. Sepúlveda, S. Tomsovic, and E. J. Heller, ''Semiclassical propagation: How long can it last?,'' Phys. Rev. Lett. **69**, 402–405 (1992).
[14] S. Tomsovic and E. J. Heller, ''Long-time semiclassical dynamics of chaos: The stadium billiard,'' Phys. Rev. E **47**, 282–299 (1993).
[15] F. D. Tappert, ''The parabolic approximation method,'' in *Wave Propagation and Underwater Acoustics*, edited by J. B. Keller and J. S. Papadakis, *Topics in Current Physics* (Springer-Verlag, Berlin, 1977), Vol. 8.
[16] W. H. Munk, ''Sound channel in an exponentially stratified ocean with applications to SOFAR,'' J. Acoust. Soc. Am. **55**, 220–226 (1974).
[17] J. A. Colosi and M. G. Brown, ''Efficient numerical simulation of stochastic internal-wave induced sound-speed perturbation fields,'' J. Acoust. Soc. Am. **103**, 2232–2235 (1998).
[18] W. H. Munk and C. Garrett, ''Internal waves in the ocean,'' Annu. Rev. Fluid Mech. **11**, 339–369 (1979).
[19] K. C. Hegewisch, N. R. Cerruti, and S. Tomsovic, preprint.
[20] J. Simmen, S. M. Flatté, and G.-Yu Wang, ''Wavefront folding, chaos and diffraction for sound propagation through ocean internal waves,'' J. Acoust. Soc. Am. **102**, 239–255 (1997).
[21] R. H. Hardin and F. D. Tappert, ''Application of the split-step Fourier method to the numerical solution of nonlinear and variable coefficient wave equations,'' SIAM Rev. **15**, 423 (1973).
[22] J. A. Fleck and M. D. Feit, ''Time-dependent propagation of high energy laser beams through the atmosphere,'' Appl. Phys. **10**, 129–160 (1976).
[23] S. M. Flatté, R. Dashen, W. H. Munk, K. M. Watson, and F. Zachariasen, *Sound Transmission Through a Fluctuating Ocean* (Cambridge University Press, Cambridge, 1979).
[24] B. M. Howe, J. A. Mercer, R. C. Spindel, P. F. Worcester, J. A. Hildebrand, W. S. Hodgkiss, Jr., T. Duda, and S. M. Flatté, ''Slice 89: A single-slice tomography experiment,'' in *Ocean Variability and Acoustic Propagation* edited by J. Potter and A. Warn-Varnas (Kluwer, Academic, 1991), pp. 81–86.
[25] T. F. Duda, S. M. Flatté, J. A. Colosi, B. D. Cornuelle, J. A. Hildebrand, W. S. Hodgkiss, Jr., 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).
[26] M. G. Brown and J. Viechnicki, ''Stochastic ray theory for long-range sound propagation in deep ocean environments,'' J. Acoust. Soc. Am. **104**, 2090–2104 (1998).
[27] D. Huber, E. J. Heller, and R. G. Littlejohn, ''Generalized Gaussian wave packet dynamics, Schrödinger equation, and stationary phase approximation,'' J. Chem. Phys. **89**, 2003–2014 (1988).
[28] M. G. Brown, J. A. Colosi, S. Tomsovic, A. L. Virovlyanksky, M. A. Wolfson, and G. M. Zaslavsky, ''Ray dynamics in long-range deep ocean sound propagation,'' J. Acoust. Soc. Am. **113**, 2533–2547 (2003).
[29] L. B. Schulman, *Techniques and Applications of Path Integration* (Wiley, New York, 1981).
[30] E. J. Heller, ''Wavepacket dynamics and quantum chaology'' in *Chaos and Quantum Physics*, edited by M.-J. Giannoni, A. Voros, and J. Zinn-Justin, Les Houches Session LII 1989 (Elsevier, Amsterdam, 1991).
[31] P. Grassberger, R. Badii, and A. Politi, ''Scaling laws for invariant measures on hyperbolic and non-hyperbolic attractors,'' J. Stat. Phys. **51**, 135–178 (1988).
[32] M. A. Wolfson and S. Tomsovic, ''On the stability of long-range sound propagation through a structured ocean,'' J. Acoust. Soc. Am. **109**, 2693–2703 (2001).
[33] F. J. Beron-Vera and M. G. Brown, ''Ray stability in weakly range-dependent sound channels,'' J. Acoust. Soc. Am. **114**, 123–130 (2003).
[34] P. F. Worcester, B. M. Howe, J. A. Mercer, M. A. Dzieciuch, and the Alternate Source Test (AST) Group (T. G. Birdsall, B. M Howe, J. A. Mercer, K. Metzger, R. C. Spindel, and P. F. Worcester), ''A comparison of long-range acoustic propagation at ultra-low (28 Hz) and very-low (84 Hz) frequencies,'' in *Proceedings of the US–Russia Workshop on Experimental Underwater Acoustics*, edited by V. I. Talanov and N. Novgorod (Institute of Applied Physics, Russian Academy of Sciences, 2000), pp. 93–104.

Ray chaos, travel time modulation, and sensitivity to the initial conditions

I. P. Smirnov and A. L. Virovlyansky
Institute of Applied Physics, Russian Academy of Science, 46 Ul'yanov Street, 603950 Nizhny Novgorod, Russia

G. M. Zaslavsky
Courant Institute of Mathematical Sciences, New York University, 251 Mercer St., New York, New York 10012, and Department of Physics, New York University, 2-4 Washington Place, New York, New York 10003

(Received 17 March 2004; revised 9 September 2004; accepted 9 September 2004)

Using a model of underwater acoustic waveguide with a periodic range dependence, travel times of chaotic rays at long-range sound propagation in the ocean are studied. It is shown that the travel time as a function of the initial momentum and propagation range in the unperturbed (range-independent) waveguide displays a scaling law. Some properties predicted by this law still persist in periodically nonuniform waveguides with chaotic ray trajectories. Ray tracing has demonstrated that the coexistence of chaotic and regular rays causes an appearance of a gap in the timefront, representing ray arrivals in the time–depth plane. Besides the gap, a ''focusing'' of ray travel times has been found within a comparatively small temporal interval preceding the gap. In numerical simulation of a sound pulse propagation this phenomenon reveals itself in the appearance of a bright spot in the distribution of acoustic energy in the time–depth plane. It has been shown that the effect of focusing is a manifestation of the so-called stickiness, i.e., the presence of such parts of the chaotic trajectory where the latter exhibits an almost regular behavior. © *2005 Acoustical Society of America.*
[DOI: 10.1121/1.1854751]

PACS numbers: 43.30.Cq, 43.20.Dk [AIT] Pages: 1595–1606

I. INTRODUCTION

In the past two decades it has been realized that ray chaos—analog to dynamical chaos in mechanics—is an important factor determining properties of sound fields at long-range propagation in the ocean. This has been demonstrated for idealized propagation models with mainly periodic range dependences,[1–9] as well as for more realistic models including mesoscale inhomogeneities or that induced by internal waves.[10–15] Special attention has been given to studying chaotic properties of the ray travel times representing arrival times of sound pulses coming to the receiver through individual rays.[11,12,16–21] The ray travel time has been the main signal parameter in the underwater acoustics experiments from which inversions have been performed to reconstruct ocean temperature field.[22,23]

An environmental model widely used in studies of the chaotic ray motion represents a superposition of a smooth range-independent (or slowly varying with range) sound-speed field and a weak range-dependent perturbation induced by internal waves.[11,12] The perturbation gives rise to ray chaos. Numerical ray tracing has demonstrated that there exists a remarkable connection between travel times and shapes (topologies) of ray trajectories. It turns out that travel times of chaotic eigenrays, that is, rays arriving at a fixed observation point, usually form rather compact clusters.[16,21] Each cluster is formed by eigenrays with equal identifiers $\pm M$, where M is the number of ray turning points and $\pm$ is the sign of the launch angle. Clusters formed by steep rays may be resolved even at very long ranges (up to a few thousand kilometers). This phenomenon sheds new light on the surprising stability of the early portion of the arrival pattern observed in both numerical simulations and field experiments.[24–29] The point is that the center of each cluster is close to the travel time of an unperturbed eigenray whose identifier is equal to that associated with the cluster. It is important to note that, having the same topology and close travel times, trajectories of rays forming the cluster may significantly deviate from each other.

In a series of recent papers[19,30,31] (see also Ref. 32) we further investigated connections between travel times of chaotic rays and topologies of their trajectories. This was done for a waveguide with a periodic range dependence, and the emphasis was given to effects caused by the coexistence of chaotic and regular rays. Although an environmental model exploited in Refs. 19, 30, and 31 (this model was borrowed from Ref. 3) is somewhat artificial, it provides a good starting point for a ray-based analysis of an irregular structure of the wave field. The fact is, the properties of ray chaos in a periodic waveguide are identical to extensively studied properties of chaotic dynamics of a nonlinear oscillator driven by an external periodic force.[33,34] Therefore, in order to investigate the chaotic ray motion one can apply methods derived in mechanics. In particular, the use of the Poincaré map allows one to find areas in the phase space visited by predominantly chaotic (''chaotic sea'') or predominantly regular (''stable islands'') rays.[1–3] On the other hand, a counterpart of the ray travel time in classical mechanics is the Hamilton principal function (the action integral), i.e., such a characteristic of dynamical systems that is, typically, not measured

experimentally and for this reason has not received much attention in theory of dynamical chaos.

In this paper we present a brief review of results obtained in Refs. 19, 30, and 31 that concern properties of ray travel times at long-range sound propagation. For simplicity it is assumed that the wave propagation can be described in the parabolic equation approximation.[11,35,36] The analysis of the ray motion will be performed in the scope of the Hamiltonian formalism.[1,11,12] One of our objectives is to investigate the dependence of the ray travel time T on the starting momentum $p_0=\tan\chi_0$ (χ_0 is the launch angle). It is shown that in the unperturbed (range-independent) waveguide the evolution of this dependence with range is governed by a scaling law. It turns out that some features of function $T(p_0)$ established in the unperturbed waveguide remain valid in the presence of perturbation giving rise to ray chaos.

We also investigate a chaotic structure of the timefront, representing ray arrivals in the time–depth plane. Our attention is restricted to two phenomena caused by the coexistence of chaotic and regular rays. One of them is the appearance of a gap in the timefront with a very low density of arrivals. Another phenomenon consists of focusing of ray travel times within some area of the time–depth plane near the gap. It is shown that this effect is associated with the so-called stickiness, i.e., the presence of such parts of the chaotic trajectory where the latter exhibit an almost regular behavior.[37] This occurs when, after wandering in the phase space, the trajectory approaches a stable island and ''sticks'' to its border for some time that may be fairly long. We demonstrate that the effect of focusing clearly manifests itself at finite wavelength in the appearance of a bright spot in the distribution of acoustic energy in the time–depth plane.

In numerical simulation we have applied standard methods routinely used for modeling wave fields in inhomogeneous underwater sound channels.[11,36] Ray equations have been solved using the fourth-order Runge–Kutta method. In order to compute the wave field at finite wavelength, we have solved the parabolic equation with the code MMPE.[38] This code, created by Smith and Tappert for solving a wide-angle parabolic equation, was slightly modified for treating the standard parabolic equation considered in our work. All numerical computations were done in double precision to insure minimal effects of roundoff.

The paper is organized as follows. In Sec. II we give a brief description of ray theory in the small-angle approximation. Properties of ray travel times in the range-independent (unperturbed) waveguide are examined in Sec. III. Here, we analyze $T(p_0)$ curves representing the travel time, T, versus starting momentum, p_0. Simple relations characterizing individual segments of this curve are derived. A scaling law describing the evolution of the segment with range is established.

In Sec. IV we study ray travel times in the perturbed waveguide. This is done with an example of a propagation model representing a deep ocean acoustic waveguide with a range-dependent perturbation giving rise to ray chaos. A phase portrait of our system is constructed using the Poincaré map. By means of numerical ray tracing we investigate stable features of the travel time dependence on starting momentum. In this section we demonstrate the presence of areas in the time–depth plane with abnormally low (gap) and abnormally high (focusing of ray travel times) densities of arrivals. The role of stickiness and chaotic jets in focusing of ray travel times is discussed. The manifestation of this phenomenon at finite wavelength is illustrated with a numerical example. In Sec. V the results of this work are summarized.

II. PARABOLIC EQUATION APPROXIMATION

Consider a two-dimensional underwater acoustic waveguide with the sound speed c being a function of depth, z, and range, r. The sound wave field u as a function of r, z, and time, t, may be represented as

$$u(r,z,t)=\int d\omega\tilde{u}(r,z,\omega)e^{-i\omega t}, \tag{1}$$

with $\tilde{u}$ governed by the Helmholtz equation[35,36]

$$\frac{\partial^2\tilde{u}}{\partial r^2}+\frac{\partial^2\tilde{u}}{\partial z^2}+k^2n^2\tilde{u}=0, \tag{2}$$

where $k=\omega/c_0$, $n=c_0/c(r,z)$, and c_0 is a reference sound speed. In the small-angle approximation, when sound waves propagate at small grazing angles with respect to the horizontal, the Helmholtz equation may be approximated by the standard parabolic equation.[11,35,36] Present $\tilde{u}$ as

$$\tilde{u}(r,z,\omega)=\tilde{v}(r,z,\omega)\,e^{ikr}, \tag{3}$$

and substitute this expression into Eq. (2). On neglecting the second derivative of $\tilde{v}$ with respect to r, we arrive at the parabolic equation

$$2ik\frac{\partial\tilde{v}}{\partial r}+\frac{\partial^2\tilde{v}}{\partial z^2}+k^2(n^2-1)\tilde{v}=0, \tag{4}$$

which formally coincides with the time-dependent Schrödinger equation. In this analogy r plays a role of time and k^{-1} is associated with the Planck constant.

The solution to Eq. (4) in the geometrical optics (semiclassical) approximation is expressed through parameters of ray trajectories governed by the Hamilton equation[1,11,39]

$$\frac{dz}{dr}=\frac{\partial H}{\partial p},\quad \frac{dp}{dr}=-\frac{\partial H}{\partial z}, \tag{5}$$

with the Hamiltonian

$$H=\frac{p^2}{2}+U(r,z), \tag{6}$$

where

$$U(r,z)=\frac{1-n^2(r,z)}{2}. \tag{7}$$

The variable p is an analog to the mechanical momentum. It relates to the ray-grazing angle χ by $p=\tan\chi$.

In deep-water acoustic waveguides the sound-speed profile at any cross section ($r=\text{const}$) usually has a minimum at a depth of about 1 km.[35] Correspondingly, the ''potential'' U in Eq. (7) represents a potential well whose parameters may vary with range r. Since r is a time-like variable, our Hamiltonian system formally coincides with that describing a mechanical particle oscillating in a time-dependent potential well.

For the point source located at $r=0$ and $z=z_0$

$$\tilde{v}=\sum_{\nu} A^{\nu}(z_r,z_0,r,\omega)e^{ikS^{\nu}(z_r,z_0,r)}, \tag{8}$$

where each term represents a contribution to the total wave field from an eigenray. The sum goes over all the eigenrays contributing at a particular receiver position (r,z_r). In the above formula the superscript ν numbers the eigenrays, A^{ν} and S^{ν} are the amplitude and the eikonal of the νth eigenray, respectively. The eikonal—analog to the Hamilton principal function in mechanics—is defined by the integral[11,39]

$$S=\int (p\,dz-Hdr)\equiv\int p\,dz-\int Hdr, \tag{9}$$

over the ray trajectory from $(0,z_0)$ to (r,z_r). It satisfies the well-known relations[39]

$$\frac{\partial S(z_r,z_0,r)}{\partial z_r}=p_r, \quad \frac{\partial S(z_r,z_0,r)}{\partial z_0}=-p_0, \tag{10}$$

where p_0 and p_r are starting and final momenta, respectively.

The amplitude $A(z_r,z_0,r)$ is given by[40]

$$A=C(\omega)\sqrt{\left|\frac{\partial^2 S}{\partial z_r \partial z_0}\right|}=C(\omega)\sqrt{\frac{1}{|\partial z_r/\partial p_0|}}, \tag{11}$$

where $C(\omega)$ is a function determined by the time dependence of the radiated signal, and the derivative $\partial z_r/\partial p_0$ is taken at the range r.

Substituting Eqs. (3) and (8) into Eq. (1) yields

$$u(r,z_r,t)=\sum_{\nu}\int d\omega A^{\nu}(z_r,z_0,r,\omega) \times\exp\left(i\omega\left(\frac{r}{c_0}+\frac{1}{c_0}S^{\nu}(z_r,z_0,r)-t\right)\right). \tag{12}$$

Each term in this sum represents a sound pulse coming to the observation point through a particular eigenray. The quantity

$$T=\frac{r}{c_0}+\frac{1}{c_0}S(z_r,z_0,r), \tag{13}$$

determines a delay of the pulse relative to an initially radiated signal and is called the ray travel time. Note that for the point source Eq. (12) translates to

$$u(r,z_r,t)=\sum_{\nu}\sqrt{\frac{k_0}{2\pi i|\partial z_r/\partial p_0|_{p_0=p_{0,\nu}}}}\,s(T^{\nu}(r,z_r)-t), \tag{14}$$

where $k_0=\omega_0/c_0$ with ω_0 being a central frequency, and $s(t)$ represents the shape of an emitted pulse.

Our main task in this paper is to study general properties of ray travel times in acoustic waveguides. This problem is equivalent to studying properties of the principal Hamilton function S of a mechanical particle oscillating in a potential well.

III. RAY TRAVEL TIMES IN A RANGE-INDEPENDENT WAVEGUIDE

A. Environmental model

In what follows we consider the sound-speed field

$$c(r,z)=\overline{c}(z)+\delta c(r,z), \tag{15}$$

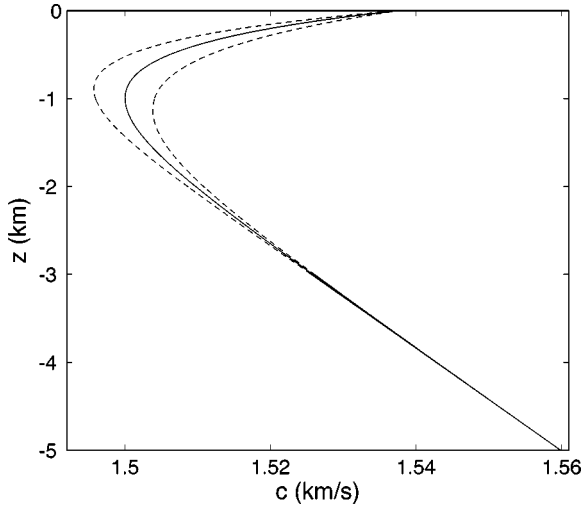


FIG. 1. Solid curve represents the unperturbed sound-speed profile. Two dashed curves show maximum deviations due to the perturbation.

representing a superposition of a smooth range-independent profile $\overline{c}(z)$ and a weak perturbation $\delta c(r,z)$ giving rise to ray chaos. In numerical simulation $\overline{c}(z)$ has been taken in the form of the Munk profile[35,41]

$$\overline{c}(z)=c_0(1+\varepsilon(e^{\eta}-\eta-1)), \quad \eta=2(z-z_a)/B, \tag{16}$$

where $c_0=1.5$ km/s, $z_a=-1$ km is the sound-channel axis (depth corresponding to the minimum of sound speed), $\varepsilon=0.0057$, and $B=1$ km. It is assumed that the z axis is directed upward and $z=0$ corresponds to a water surface.

Following Refs. 3, 17, 19, and 30, we choose a perturbation

$$\delta c(r,z)=2\gamma c_0\frac{z}{B}e^{2z/B}\cos(2\pi r/\lambda), \tag{17}$$

where $\gamma=0.01$ and the spatial period $\lambda=10$ km. The amplitude of δc takes its maximum value of about 5 m/s at 500-m depth. The total sound-speed profile and an amplitude of its variations are shown in Fig. 1. Although this perturbation is highly idealized, it gives rise to a chaotic ray motion whose properties resemble that observed in more realistic environmental models.[11,12,21] For example, travel times of chaotic rays with equal identifiers in a waveguide determined by Eqs. (15)–(17) form rather compact clusters.[30] It also should be mentioned that the Lyapunov exponents characterizing stochastic instability of chaotic ray paths in our model are approximately the same [about 0.01 km^{-1} (Ref. 19)] as in a typical deep-water waveguide with internal wave-induced inhomogeneities.[11,21] Besides, the expression (17) is so simple that it can be used not only in numerical simulations but even in obtaining analytical estimates (e.g., as it was done in Ref. 31). An important advantage of our perturbation is its periodic range dependence, which allows one to apply the Poincaré map for investigating the structure of the phase space. This option will be used in Sec. IV A.

B. Description of a ray trajectory

Let us begin with a simple case of a range-independent sound-speed profile $c(r,z)=\overline{c}(z)$. Then, Snell's law[35] has the form analogous to the energy conservation law in classical mechanics.[39] Along the ray path

$$H(p,z)=E, \tag{18}$$

with E being a constant.[39] The ray trajectory is a periodic curve whose momentum as a function of depth z and "energy" E is

$$p(z,E)=\pm\sqrt{2[E-U(z)]}. \tag{19}$$

Let us introduce the action variable I connected to E via[1,39]

$$I=\frac{1}{\pi}\int_{z_{\min}}^{z_{\max}} dz\sqrt{2[E-U(z)]}, \tag{20}$$

where $z_{\min}$ and $z_{\max}$ are the lower and upper turning points satisfying an equation

$$U(z)=E. \tag{21}$$

Equations (20) and (21) define the function $E(I)$. They also define $z_{\min}$ and $z_{\max}$ as functions of I.

According to the first of Eqs. (5) and Eq. (6), $dz/dr=p$. This relation, combined with Eq. (19), gives a simple formula for a range between a minimum of the ray trajectory and a closest point at which the ray takes coordinate z

$$g(z,I)=\int_{z_{\min}}^{z} dz/p(z,E(I)). \tag{22}$$

Notice that $g(z_{\max},I)=L(I)/2$, where L is a period (cycle length) of the ray path. Differentiating Eq. (20) with respect to E yields a well-known relation for the spatial frequency of ray path oscillations[1,39]

$$\Omega(I)=\frac{2\pi}{L(I)}=\frac{dE(I)}{dI}. \tag{23}$$

Note that, in the vicinity of the lower turning point, i.e., at z close to $z_{\min}$

$$g(I,z)=\frac{2}{Q}(z-z_{\min})^{1/2}, \tag{24}$$

where

$$Q=\sqrt{-2U'(z_{\min})}, \quad U'\equiv dU/dz. \tag{25}$$

By definition

$$g\leqslant L/2. \tag{26}$$

The general expression determining the ray trajectory connecting points $(0,z_0)$ and (r,z) is given by

$$r=R(z_0,z,I), \tag{27}$$

with

$$R(z_0,z,I)=\alpha g(z_0,I)+N\,L(I)+\beta g(z,I). \tag{28}$$

Here, N is the number of trajectory cycles, and coefficients α and β are defined by signs of starting and final momenta, respectively: $\alpha=-\operatorname{sgn} p_0$, $\beta=\operatorname{sgn} p$. For brevity, we do not indicate explicitly the arguments α, β, and N of the function R.

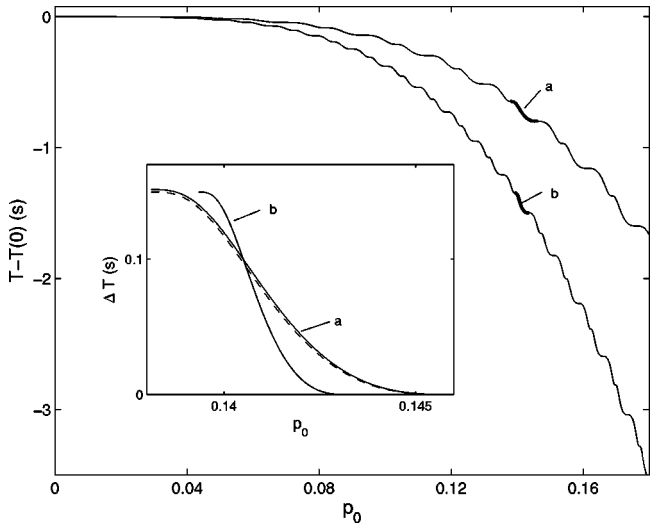


FIG. 2. Ray travel time as a function of starting momentum at ranges of 1500 km (upper curve) and 3000 km (lower curve). Rays escape a point source set at a depth of 1 km. Travel times are figured from arrival times of axial rays with starting momenta $p_0=0$. Segments *a* and *b* marked by thick lines are formed by rays with identifiers $+62$ (1500-km curve) and $+124$ (3000-km curve), respectively. Starting momenta of both segments are close to $p_0=0.14$. Inset: Segments *a* and *b* (solid lines) shown on the same plot with travel times figured from their earliest points. A dashed line represents segment *b* extended along the p_0 axis by a factor of 2 and shifted along this axis.

C. Ray travel time as a function of starting momentum

Two curves shown in Fig. 2 graph dependencies of the ray travel time T on the starting momentum p_0 at ranges of 1500 and 3000 km (in modern experiments on long-range sound propagation, acoustic paths typically are on the order of a few thousand kilometers[24–26]). Here and in what follows, we present numerical results obtained for a case when the wave field is excited by a point source set at a depth of 1 km, i.e., on the axis of the sound channel. Only positive p_0 are considered in Fig. 2. For negative p_0 the results are similar. For each curve the ray travel time is figured from $T(0)$, arrival time of a pulse coming through a ray starting horizontally (with $p_0=0$).

Note that both $T(p_0)$ curves in Fig. 2 have an important common feature: each curve has "shelves" where its slopes with respect to the horizontal are small. At intervals of starting momenta corresponding to these shelves, the ray travel time T is most stable (least sensitive) with respect to small variations in p_0. Using the first of Eqs. (10) and Eq. (13), it is easy to see that

$$\frac{\partial T}{\partial p_0}=\frac{1}{c_0}p\frac{\partial z}{\partial p_0}. \tag{29}$$

The derivative $\partial T/\partial p_0$ vanishes at ranges where the ray passes through caustics ($\partial z/\partial p_0=0$) and at ranges corresponding to ray turning points ($p=0$). Although the ray escaping a point source does not encounter caustics at turning points, it is well-known that at long ranges each caustic is located in the vicinity of a turning point.[42] So, zeros of $\partial T/\partial p_0$ should come in closely spaced pairs, determining positions of centers of the shelves.

This result means that the $T(p_0)$ curve may be divided into elementary segments so that each segment connects cen-

ters of neighboring shelves and is formed by a ray with: (i) the same number of cycles; (ii) starting momenta of the same sign; and (iii) final momenta of the same sign. All such rays have the same values of constants α, β, and N [see Eqs. (27) and (28)]. In other words, all rays forming the segment have the same identifier $\pm M$. At long ranges, where the condition

$$N \gg 1 \tag{30}$$

is met, the shape of the segment can be analytically described. Let us consider this issue in detail.

Take a segment of the $T(p_0)$ curve at the range r formed by rays starting from the point $(0,z_0)$ with initial momenta p_0 belonging to an interval

$$p_0^* \leqslant p_0 \leqslant p_0^{**}. \tag{31}$$

Assume that the limiting values p_0^* and p_0^{**} correspond to rays that have at the range r the lower and upper turning points, respectively. The difference

$$\Delta T = T(p_0) - T(p_0^*), \tag{32}$$

with p_0 satisfying the condition (31), determines the shape of the segment. Our task is to derive an approximate analytical expression for the function $\Delta T(\Delta p)$, where

$$\Delta p = p_0 - p_0^*. \tag{33}$$

If the condition (30) is met, then an approximate relation following from Eqs. (27) and (28)

$$g(z,I) = -NL_{p_0}\Delta p, \tag{34}$$

with $L_{p_0} \equiv \partial L/\partial p_0$, is valid. Since the value of g is restricted [see (26)] Δp decreases with range as $L/r = O(1/N)$. This means that the interval of p_0, and, hence, the interval of action I, corresponding to the segment decreases with range. Neglecting the difference between actions of rays forming the segment, we can approximately integrate the first of Eqs. (10) arriving at

$$\Delta T = \frac{1}{c_0} G(z,I), \tag{35}$$

where

$$G(z,I) = \int_{z_{\min}}^{z} dz \sqrt{2(E - U(z))}. \tag{36}$$

To find $\Delta T(\Delta p)$ one should substitute in Eq. (35) the coordinate z and the action I of the ray with starting momentum $p_0 = p_0^* + \Delta p$. Note that the total time spread corresponding to the segment (given by the maximum value of ΔT) is

$$\Delta T_{\max} = \pi \overline{I}/c_0, \tag{37}$$

where $\overline{I}$ is the mean action of the segment.

An explicit dependence of ΔT on Δp can be easily derived for rays with small final momenta. Consider rays whose coordinates z at the range r are close to their lower turning points. For such rays Eq. (24) gives a simple connection between g and z. A similar result for $G(z,I)$ is given by

$$G(z,I) = Q\,\tfrac{2}{3}(z - z_{\min})^{3/2}. \tag{38}$$

Substituting Eq. (34) into Eq. (24) and exploiting Eqs. (35) and (38) yields that, in the vicinity of the lower turning point

$$\Delta T = -\frac{1}{12c_0} Q^4 L_{p_0}^3 (N\Delta p)^3, \tag{39}$$

which explicitly describes the shape of the shelf corresponding to the lower turning point. A similar expression is valid in the neighborhood of the upper turning point.

Equation (39) may be used for estimation of the "shelf" width, $\overline{\Delta p}$. Defining $\overline{\Delta p}$ as a value of Δp corresponding to $\Delta T = \Delta T_{\max}/12$, we get from Eq. (39)

$$\overline{\Delta p} = \frac{1}{N} \frac{(\pi I)^{1/3}}{L_{p_0} Q^{4/3}}. \tag{40}$$

It should be pointed out that Eqs. (35) and (36) as well as Eq. (39) do not take into account the presence of a caustic near the ray turning point. These equations predict only one zero of the derivative $\partial T/\partial p_0$ on the shelf, namely, that at the turning point where $p = 0$. Making use of Eqs. (24), (28), and (34), we find that the value of $\Delta p = \Delta p_c$ corresponding to a ray passing through a caustic at the range r can be estimated as

$$\Delta p_c = \frac{1}{N^2} \frac{8\pi p_0^*}{Q^4 L_{p_0}^2 L}. \tag{41}$$

Thus, we have $\overline{\Delta p} = O(1/N)$ and $\Delta p_c = O(1/N^2)$, i.e., at long ranges

$$\Delta p_c \ll \overline{\Delta p}.$$

This means that our approach based on Eqs. (35), (36), and (39) provides an analytical description of the whole segment, excluding only small parts formed by rays with p_0 close to p_0^* and p_0^{**}. At these parts of the segment the value of ΔT is either very small ($\Delta T \ll \Delta T_{\max}$) or close to $T_{\max}$ ($\Delta T_{\max} - \Delta T \ll \Delta T_{\max}$).

D. Scaling law

Let us apply the above analytical results for investigation of general properties of the $T(p_0)$ curve in a range-independent waveguide. Notice that the rays reflecting off the surface have a nonzero momentum at its upper turning point, and the derivative $\partial T/\partial p_0$ does not tend to zero when the ray approaches the upper turning point. Therefore, surface-bounce rays do not form shelves corresponding to upper turning points unless they strike the surface at small grazing angles. In what follows we restrict our attention to purely refracted rays.

Both curves shown in Fig. 2 are formed by rays that do not reflect off any surface. Each segment of these curves connects two shelves corresponding to the lower and upper turning points. The temporal interval between neighboring shelves, $\Delta T_{\max}$, is given by Eq. (37).

Equations (35)–(37) describing the shape of the segment can be applied for studying how the $T(p_0)$ curve evolves with range. As earlier, we consider only long enough ranges and assume that the condition (30) is met. Note that Eq. (37) does not depend explicitly on the range r. It means that if we select some starting momentum $\overline{p_0}$ and take segments of $T(p_0)$ curves embracing this $\overline{p_0}$ at different ranges, then all such segments have approximately the *same* time spread.

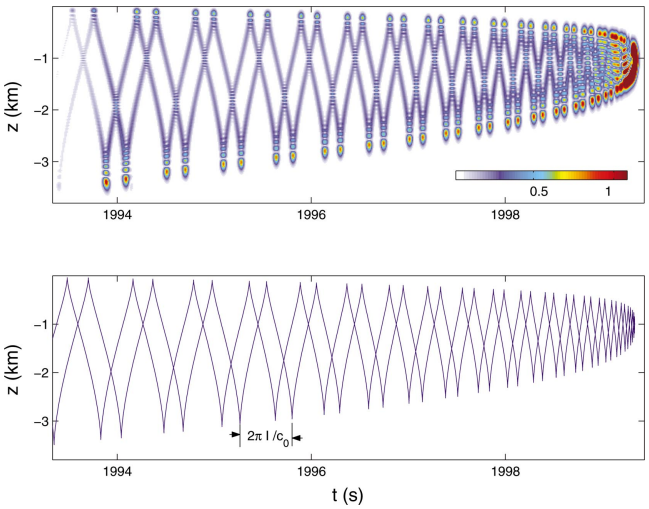


FIG. 3. Upper panel: Pulse signal in the unperturbed waveguide at range 3000 km. Amplitude of the acoustic wave field is shown as a function of travel time and depth. Lower panel: Timefront in the unperturbed waveguide at the range 3000 km.

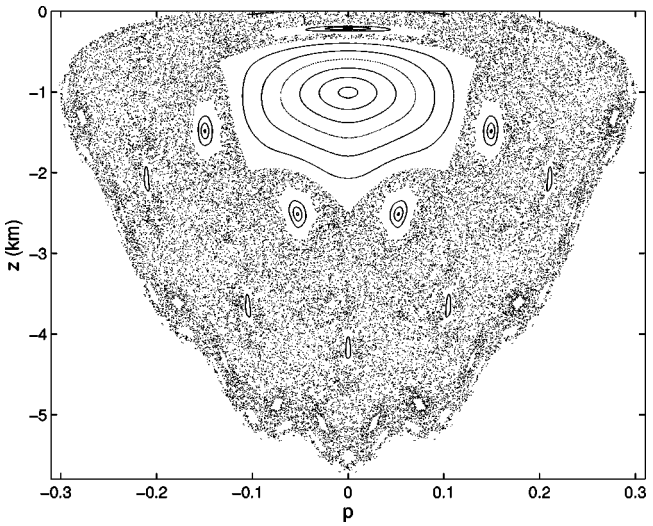


FIG. 4. Phase portrait of the ray system constructed using the Poincaré map.

This conservation law may be generalized if we take into account that, according to Eq. (34), the coordinate z is a function of the product $N\Delta p$. Since N for a ray with some fixed starting momentum is proportional to r, Eq. (35) leads to a scaling law determining the range dependence of the segment shape. Consider ΔT defined by Eq. (32) as a function of three arguments: Δp, r, and $\overline{p_0}$. The latter defines a particular segment: it determines the interval (31). Then, the range variation of the segment is given by

$$\Delta T(\Delta p, r, \overline{p_0}) = \Delta T(\Delta p/\mu, \mu r, \overline{p_0}), \tag{42}$$

where μ is a positive constant. In particular, this result means that a segment determined by a fixed starting momentum $\overline{p_0}$ has a fixed (range-independent) time spread, and an interval of starting momenta corresponding to this segment is inversely proportional to range, i.e.,

$$p_0^{**} - p_0^{*} = \text{const}/r. \tag{43}$$

To illustrate the scaling law (42), take two segments in Fig. 2 marked with symbols a (at 1500 km) and b (at 3000 km). Both are formed by rays with starting momenta close to $p_0 = 0.14$. In agreement with Eq. (43), the number of segments on the lower curve is two times greater than that on the upper one. We use segments a and b to check Eq. (42). This is done in the inset. The segments are shown here on the same time scale with travel times ΔT figured from their minima. In agreement with Eq. (42), both segments span almost the same temporal interval, while the interval of starting momentum p_0 for b is two times smaller compared to that for a. Furthermore, the segment b extended by a factor of 2 along the p_0 axis (dashed curve) after an appropriate shift along this axis almost coincides with the segment a.

E. Timefront

In the lower panel of Fig. 3 the distribution of ray travel times is presented in an alternative form. Each point depicts a ray arrival in a time–depth plane. This is the so-called timefront consisting of segments of smooth curves. Every segment is formed by rays with the same identifier $\pm M$. Therefore, each segment of the timefront has a counterpart in the corresponding $T(p_0)$ curve (see Sec. III C). Using Eq. (37) a period of the broken line—time delay between two consecutive minima—at long ranges can be estimated as

$$\tau = 2\pi \overline{I}/c_0. \tag{44}$$

Here, we have taken into account that at long ranges rays forming neighboring segments have close actions, and denote the mean value of these actions by $\overline{I}$. This formula representing a particular case of Eq. (35) is a known result. It was derived and discussed in Refs. 43–46.

The upper panel in Fig. 3 shows the amplitude of a short pulse as a function of time and depth. This function has been computed at a range of 3000 km for a case when an initial pulse emitted by the point source has a shape

$$s(t) = \exp\left(-\frac{t^2}{2\tau^2}\right)\cos\omega_0 t, \tag{45}$$

determined by two parameters: a central frequency $\omega_0/2\pi = 75$ Hz and an effective pulse duration $\tau = 0.02$ s. It is seen that the timefront properly predicts distinctive features in the distribution of acoustic energy in the time–depth plane.

IV. TRAVEL TIMES OF CHAOTIC RAYS

A. Phase portrait of the ray system

The ray motion in the environmental model defined by Eqs. (15)–(17) was studied in Refs. 3, 17, 19, and 30. The phase portrait of the ray system is shown in Fig. 4. It has been plotted using the Poincaré map[1–5]

$$(p_{n+1}, z_{n+1}) = \hat{T}(p_n, z_n), \tag{46}$$

where a symbol $\hat{T}$ denotes transformation of the momentum and coordinate of a ray trajectory taken at a range $n\lambda$ to that at a range $(n+1)\lambda$. Thus, Fig. 4 shows ray positions in phase space at ranges spaced apart by λ. Only rays that do not reflect off the bottom (set at a 6-km depth) have been considered. We see a typical picture of stable islands filled with regular curves surrounded by a chaotic sea filled with randomly scattered points.

Rays leaving our point source have initial conditions at the straight line $z=-1$ km. In Fig. 4 it is seen that the central part of this line corresponding to interval of starting momenta

$$-0.131<p_0<0.131, \tag{47}$$

belongs to the central stable island. This means that flat rays are predominantly regular while steep rays with $|p_0|>0.131$ are predominantly chaotic. Note that in more realistic environmental models steep rays are, on the contrary, more stable than flat ones.[12,21]

B. Stable features in the travel-time dependence on the starting momentum

Ray dynamics in the environmental model defined by Eqs. (16) and (17) has been studied in Refs. 3, 17, 19, and 31. Numerical simulations demonstrate that the ray chaos in this model at ranges of a few thousand km is well developed and ray paths at such ranges practically ''forget'' their starting parameters. Therefore, it might be expected that travel times of chaotic rays also become practically random and their dependences on starting momenta are completely different from that in the range-independent waveguide. However, it turns out that travel times of chaotic rays retain some properties observed in range-independent waveguides. In Sec. III C we have noticed the following features of the unperturbed $T(p_0)$ curve.

(i) The curve has shelves where the derivative $\partial T/\partial p_0$ vanishes. Correspondingly, the density of points depicting ray arrivals at shelves is larger than at other parts of the $T(p_0)$ curve.
(ii) Each segment of the curves connecting neighboring shelves is formed by rays with the same identifiers.
(iii) A range variation of the $T(p_0)$ curve exhibits a scaling law.

We argue that these features have analogs for chaotic rays. In the dependence of the ray travel time on the starting momentum shown in Fig. 5, we see that in the perturbed waveguide the density of arrivals has local maxima at times close to that of unperturbed shelves. Figure 6 demonstrates that arrival times of chaotic rays with the given identifier occupy approximately the same temporal interval as that spanned by unperturbed rays with this identifier. It can be said that in the presence of perturbation, the segment of $T(p)$ curve turns into a strip that covers approximately the same interval of times. Widths of intervals of starting momenta corresponding to such strips grow inversely proportional to range, and in this sense the scaling law remains valid in the perturbed waveguide.

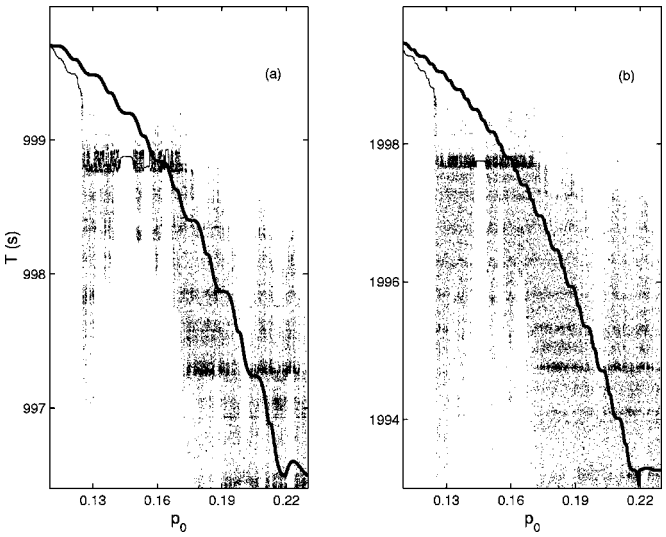


FIG. 5. Ray travel time versus starting momentum in the unperturbed (thick solid lines) and perturbed (points) waveguides at ranges of 1500 km (a) and 3000 km (b).

C. Gap between travel times of chaotic and regular rays

In a similar way the stability of ray travel times reveals itself in the perturbed timefront shown in Fig. 7. A remarkable property of the ray travel time is that even under conditions of ray chaos this characteristic, to a considerable extent, is determined by the ray identifier, i.e., by the topology of a ray trajectory. Ray tracing demonstrates that travel times of chaotic rays with the same identifier form surprisingly compact clusters.[21] In the presence of perturbation giving rise to ray chaos, perturbed rays with the given identifier produce a fuzzy version of the unperturbed segment.[19,21,46] Some of these fuzzy segments do not overlap and the corresponding part of the timefront looks similar to its counterpart in the unperturbed waveguide. Six fuzzy segments formed by rays with identifiers ± 105, ± 106, and ± 107 are indicated. These are the same arrivals that form strips shown in two lower plots in Fig. 6 (each strip corresponds to two fuzzy segments with identifiers $+M$ and $-M$). Consistent with our expectation, most stable are segments formed by flat rays belonging to the central stable island in Fig. 4. These segments are well resolved in the rear part of the arrival pattern in Fig. 7.

In Refs. 19 and 30 it has been shown that in our environmental model the coexistence of chaotic and regular rays causes an appearance of the gap in the timefront. In the upper panel of Fig. 7 this gap is observed in the interval of times between 1998 and 1998.5 s. In order to explain this phenomenon, we turn our attention to Fig. 8, where a distribution of numbers of turning points M at range 3000 km is shown for rays with negative starting momenta. The distribution has a gap at $M=121$, 122, and 123. For rays with positive starting momenta there is a similar gap for the same values of M. So, the percentage of rays with identifiers ± 121, ± 122, and ± 123 is comparatively small.

In Ref. 30 we argue that this is due to the chain of five stable islands within a chaotic sea (see Fig. 2). The point is that chaotic and regular (unperturbed) rays with equal identifiers have equal mean cycle lengths and, as a consequence, close values of mean actions. Therefore, trajectories of these rays in the phase space should be more or less close. A numerical ray tracing shows that unperturbed rays with identifiers ± 121, ± 122, and ± 123 intersect areas occupied by the five stable islands seen in Fig. 4. Since chaotic rays cannot ''penetrate'' into the islands, this fact qualitatively explains

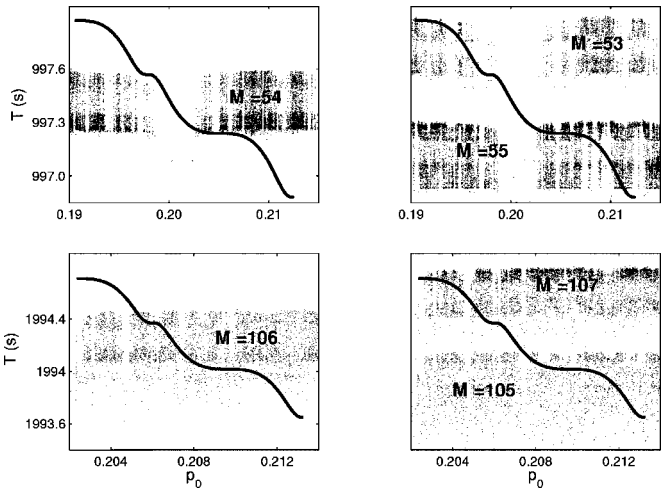


FIG. 6. Ray travel time versus starting momentum in the unperturbed (thick solid lines) and perturbed (points) waveguides at ranges of 1500 km (upper row of plots) and 3000 km (lower row of plots). In the upper plots randomly scattered points depict arrivals of rays with $M=54$ turning points (left plot) and with $M=53$ and 55 (right plot). Similarly, the lower row of plots depict arrivals of rays with $M=106$ turning (left plot) and $M=105$ and 107 (right plot).

the lack of rays with the aforementioned identifiers in the presence of perturbation.

Even in the presence of perturbation the spread of travel times for rays with the same identifier is comparatively small. Therefore, a gap in the distribution of ray turning points leads to an appearance of a gap in the timefront. Moreover, since travel times of perturbed and unperturbed rays with the same identifier are rather close,[46] the width of the gap in travel times can be roughly estimated using Eq. (44). According to this formula the difference in travel times of rays with identifiers $\pm M$ and $\pm(M+\Delta M)$ can be roughly estimated as $\Delta t \approx \pi I \Delta M / c_0$. A width of an interval of missing identifiers—a width of the gap in Fig. 8—can be estimated as $\Delta M=4$. Mean values of actions for rays with M close to 120 are about 0.06 km. This yields $\Delta t=0.5$ s, a reasonable prediction for a width of the gap in the perturbed timefront seen in the lower panel of Fig. 7.

The density of dots in the timefront is approximately

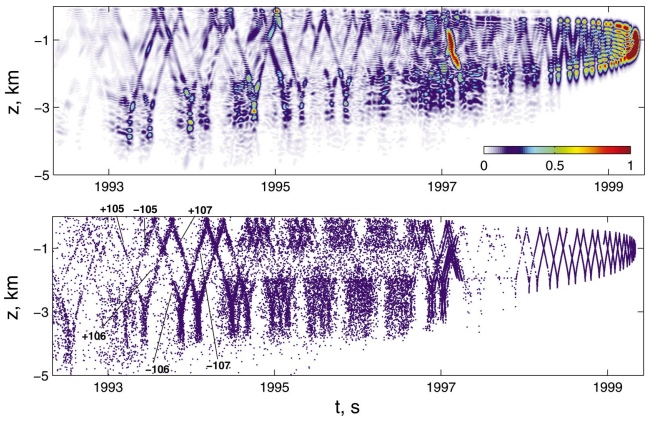


FIG. 7. The same as in Fig. 3 but for the perturbed waveguide. Identifiers ± 105, ± 106, and ± 107 are shown next to corresponding fuzzy segments. These segments are formed by the same rays that make up the strips present in two lower plots in Fig. 6.

proportional to the acoustic field intensity. Comparing the upper and lower panels in Fig. 7 we see that, due to effects of diffraction, the gap, so clearly seen in the timefront, is to a considerable extent "filled" (in both space and time) in the arrival pattern computed at a finite wavelength.

D. "Focusing" of ray travel times

Numerical simulation demonstrates another interesting feature in the distribution of sound energy in the time–depth plane. We mean a bright spot clearly seen in the lower panel of Fig. 7 at times between 1997 and 1997.3 s. Its appearance is another manifestation of the coexistence of chaotic and regular rays. In order to investigate this phenomenon, we return to Fig. 8 and note the presence of a large maximum at $M=120$. A similar maximum exists in a similar plot (not shown) constructed for rays with positive starting momenta. Thus, a percentage of rays with identifiers ± 120 is much greater than that for any other identifier. Figure 9 shows that the bright spot lies at the intersection of two fuzzy segments of the perturbed timefront formed by contributions from rays with identifiers ± 120. Comparison with unperturbed segments with the same identifiers shown in the figure demonstrates that the perturbation causes the shift of ray arrivals toward early times.

Another important thing about the rays with $M=120$ is that they are relatively stable. In order to characterize quantitatively the sensitivity of ray paths to small variations in initial conditions, we have followed Ref. 11 and computed the absolute difference, $\delta z=|z_1-z_2|$, in arrival depths, z_1 and z_2, of two rays with starting momenta that differ by $\delta p_0=10^{-7}$. In Fig. 10(a) this difference is shown as a function of starting momentum p_0 at the range 3000 km. The plot has been constructed by tracing 80 000 pairs of fan rays escaping the point source. A segment of this line within the interval $|p_0|<0.131$ belongs to the large central island. Since rays with such initial conditions are predominantly regular, the smallness of δz in this interval is not surprising. Much less expected are intervals of small δz corresponding to values of p_0 belonging to the chaotic sea. Two (most discernible) of these intervals are marked with symbols **A** and **B**. A detailed analysis shows that they are formed by rays with identifiers ± 120. In Fig. 10(b) we present values of δz for all fan rays with identifiers -120. Figure 10(c) zooms in on a small fraction of Fig. 10(b), and Fig. 10(d) in turn provides a magnified view of a small fraction of Fig. 10(c). Results presented in Figs. 10(b), (c), and (d) look similar, suggesting that the dependence of δz on p_0 has a fractal structure. All plots consist of intervals filled with randomly scattered points interspersed with intervals of very small δz. At such intervals the dependences of δz on p_0 look like segments of straight lines. Similar results have been obtained for rays with identifiers $+120$.

E. Role of stickiness and chaotic jets in focusing of ray travel times

In Fig. 4 we can see that the straight line $z=-1$ km that determines initial conditions for rays escaping our point source intersects none of the five islands forming the chain.

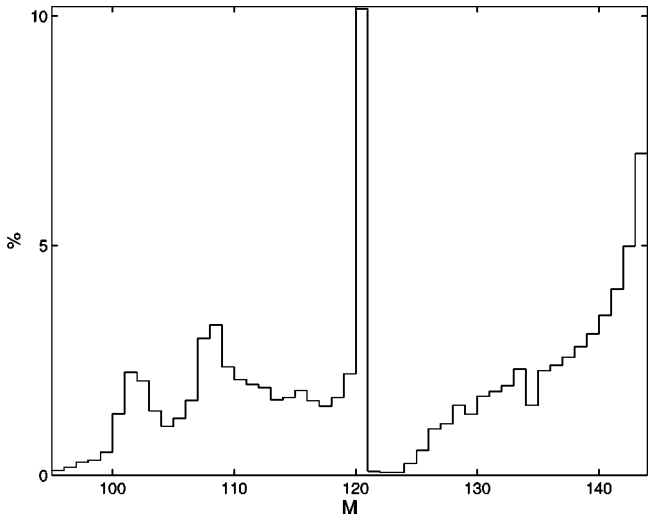


FIG. 8. Percentage of rays as a function of M, the number of turning points, at the range of 3000 km. Plot is constructed for rays with negative starting momenta.

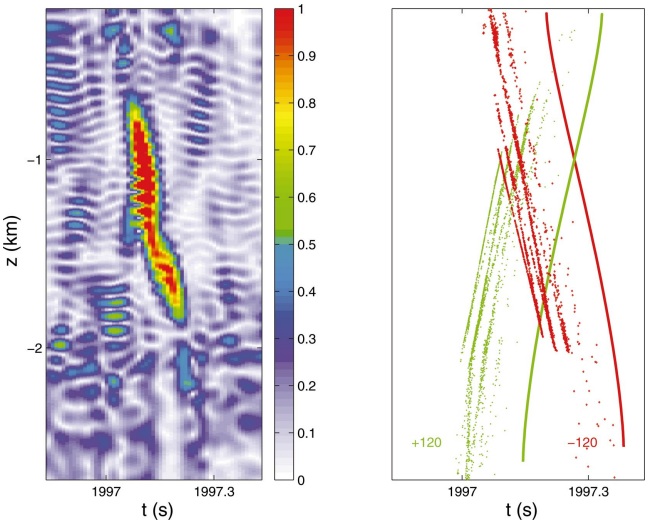


FIG. 9. Left panel: A magnified view of the portion of the arrival pattern containing the bright spot. Right panel: A part of the timefront in the perturbed waveguide formed by rays with identifiers $+120$ (red points) and -120 (green points). Two intersecting smooth curves represent segments of the unperturbed timefront formed by rays with identifiers $+120$ (red) and -120 (green).

However, this does not mean that all fan rays with starting momenta outside the central island are chaotic. A more detailed analysis show that each of the five islands is surrounded by six smaller islands graphed in Fig. 11. Two of these islands are intersected by the line $z=-1$ km. Segments of this line within the two islands form central parts of intervals **A** and **B** shown in Fig. 10(a) [interval **A** is also seen in Fig. 10(b)].

Our ray tracing has demonstrated that most parts of rays with identifiers ± 120, including those that form segments with small δz in Fig. 10, are chaotic. Their relative stability is caused by the so-called stickiness, i.e., the presence of such parts in chaotic trajectories where the latter exhibit an almost regular behavior. This occurs when, after wandering

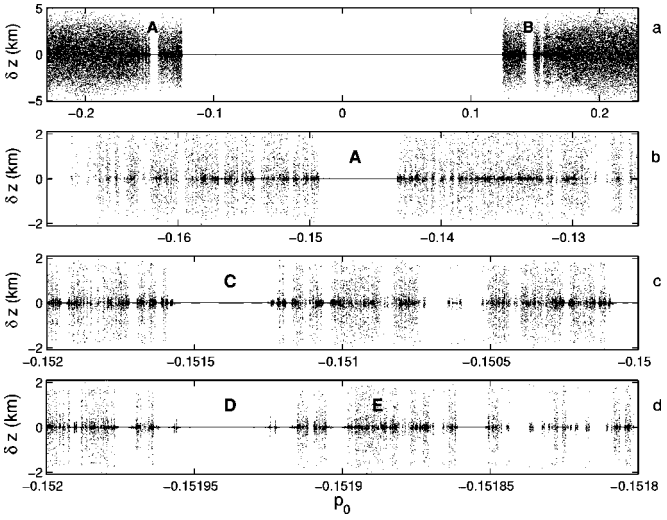


FIG. 10. Depth differences at 3000-km range for pairs of rays with infinitesimally close starting momenta. Each point represents the absolute difference in arrival depth for two rays: a ray with original momentum and one with a slightly perturbed momentum (which is the original plus 10^{-7}). The differences are shown for all fan rays whose starting momenta belong to interval $-0.24<p_0<0.24$ (a) and for rays with identifier -120 and starting momenta from intervals $-0.17<p_0<-0.125$ (b), $-0.152<p_0<-0.15$ (c), and $-0.152<p_0<-0.1518$ (d).

in the phase space, the trajectory approaches a stable island and ''sticks'' to its border for some time that may be fairly long.[37] The point is that regular rays belonging to any of the islands shown in Fig. 11 at the range 3000 km have identifiers $+120$ or -120. A significant part of rays starting from points in the phase plane close to these islands sticks to their borders and stays near the borders at long enough range interval. At this interval the sticky rays exhibit almost regular behavior, mimicking that of regular rays forming the islands. Therefore, the sticky rays have the same identifiers as their regular counterparts from the islands.

This statement is illustrated in Fig. 12. In the left column we see two Poincaré maps for a typical ray with starting momentum from interval **C** in Fig. 10(c). The trajectory has been computed up to a range of 30 000 km. In the upper plot we use only the first 3000 km of this trajectory. The lower plot has been constructed using the whole ray path. By comparing the upper and lower panels we conclude that while up to 3000 km the trajectory sticks to the islands, at longer ranges it begins to break away. In the central column of Fig. 12 we see two analogous Poincaré maps computed for a ray with the starting momentum from an interval **D** in Fig. 10(d). Although up to 3000 km the ray adheres to the islands, the remaining part of its 30 000-km trajectory passes through different portions of the chaotic sea. In the right column we have presented Poincaré maps constructed for a ray from an interval **E** [see Fig. 10(d)] filled with randomly scattered points. Even this ray, which can hardly be identified as a sticky one, does not deviate significantly from the stable islands at the first 3000 km.

The phenomenon of stickiness is a consequence of an extremely complicated structure of the phase space in the neighborhood of the island. Besides comparatively big islands seen in Figs. 4 and 11, there exist an infinite number of smaller islands indiscernible in our plots. Imbedding of such islands in the stochastic sea creates a fractal pattern of the

phase space (its manifestation we see in Fig. 10). A fractal system of islands generates a corresponding system of traps. A ray trajectory approaching this system will be trapped for some time. For detailed discussion of this issue, see Refs. 37 and 47.

Beside the stickiness, we encounter another important phenomenon of the chaotic motion. This is the existence of chaotic jets, i.e., bunches of chaotic trajectories that do not diverge at a long enough range interval. As it has been predicted in Ref. 47, the chaotic jet is formed by rays with initial conditions localized at a special area of the phase space near the border of a stable island. This is exactly what we see in our numerical simulation. Each interval of the starting momentum p_0 at which the dependence of δz on p_0 looks like a segment of a straight line $\delta z = 0$ can be associated with a chaotic jet. Typical examples are intervals **C** [in Fig. 10(c)] and **D** [in Fig. 10(d)].

F. Smoothed intensity of pulse signal

It is clear that extreme sensitivity of chaotic trajectories to initial conditions and variations of environmental parameters should impose severe limitations on the predictability of underwater sound fields in the scope of the geometrical optics approximation. On the other hand, it is known that even under conditions of ray chaos the geometrical optics can properly predict intensities of some large-scale features of the wave field.[15] We have seen that the large bright spot in the arrival pattern is formed by rays with relatively stable trajectories. This suggests that the ray theory may predict a coarse-grained intensity in a corresponding part of the received signal.

In order to confirm this conjecture, we consider a smoothed intensity defined as

$$J(r,t) = \frac{1}{\sqrt{2\pi}\Delta_t} \int dz' dt' |u(r,z',t')|^2 \times \exp\left(- \frac{(t-t')^2}{2\Delta_t^2} \right), \tag{48}$$

where Δ_t is a smoothing scale and the integration over the vertical coordinate z' goes along the whole cross section of the waveguide.

Assuming that the smoothing scale is greater than the duration of an initial pulse, i.e.,

$$\Delta_t > \tau, \tag{49}$$

substituting Eq. (14) into Eq. (48) and neglecting small contributions from rapidly oscillating terms representing products of pulses coming to the observation point through different eigenrays, we arrive at

$$J(r,z,t) = \frac{k_0 \tau}{(2\pi)^{3/2}\Delta_t} \times \sum_\nu \int dz' \frac{1}{|\partial z/\partial p_0|_{p_0 = p_{0,\nu}}} \times \exp\left(- \frac{(t - t^\nu(r,z'))^2}{2\Delta_t^2} \right)$$

$$= \frac{k_0 \tau}{2\pi\Delta_t} \int dp_0 \times \exp\left(- \frac{(t - t(r,p_0,z_0))^2}{2\Delta_t^2} \right), \tag{50}$$

where the symbol $t(r,p_0,z_0)$ denotes the travel time of a ray with starting momentum and coordinate equal to p_0 and z_0, respectively, at range r. The last integral can be evaluated as a sum over a dense enough fan of rays escaping a point source. In our calculation we have considered a fan of 80 000 rays with initial grazing angles within the interval ± 12.5 deg. In Fig. 13 we compare a ray-based prediction of the smoothed intensity predicted by Eq. (50) with the smoothing scale $\Delta_t = 0.04$ s to a result obtained by direct solving of the parabolic equation. The comparison demonstrates that the ray-based approach provides a good prediction for both maxima of the smoothed intensity formed by contributions from predominantly stable rays. On the other hand, there is a significant discrepancy between the ray-based and PE-based results in the temporal interval from 1998 to 1998.5 s. As it was already noticed at the end of Sec. IV C, the ray-based approach is unable to describe ''filling'' the gap in the timefront due to effects of diffraction.

V. CONCLUSION

The main objective of the present study has been to investigate how the coexistence of chaotic and regular rays reveals itself in the distribution of ray travel times. We have considered a deep ocean propagation model with a smooth background sound speed profile perturbed by an inhomogeneity with a periodic range dependence.

When studying ray travel times in the unperturbed waveguide, we focus mainly on the function $T(p_0)$ representing the arrival time of a pulse coming through a given ray path as a function of the starting momentum. It is shown that the $T(p_0)$ curve consists of segments formed by rays with the same identifier. A most striking feature of this curve

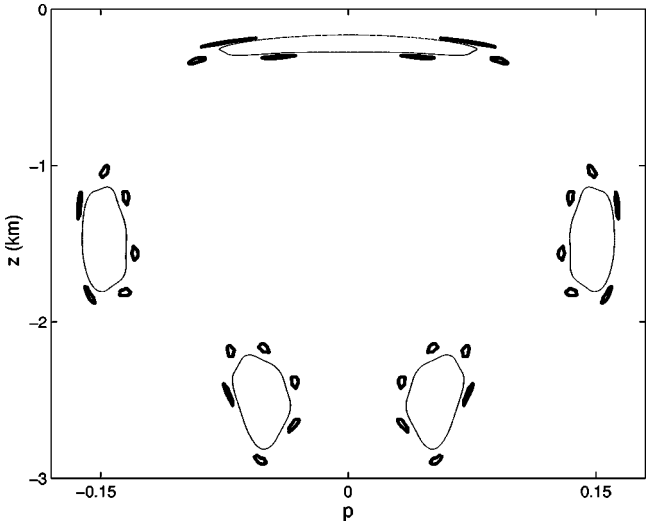


FIG. 11. Small islands (depicted by thick points) that surround a chain of five bigger islands (thin points). The small islands shown here are indiscernible in Fig. 5.

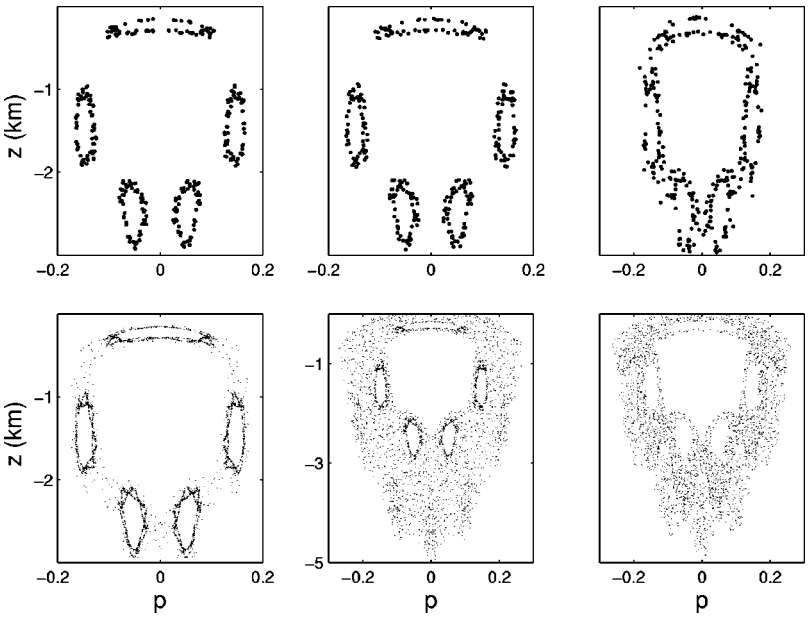


FIG. 12. Poincaré maps constructed using ray trajectories with 3000-km lengths (upper row of plots) and 30 000-km lengths (lower row of plots).

is the presence of shelves centered at p_0 corresponding to zeros of the derivative dT/dp_0. A surprisingly simple scaling law determining how the segment evolves with range is established. In the presence of a weak perturbation giving rise to ray chaos, the ray trajectory becomes extremely sensitive to the starting momentum. Nevertheless, the dependence of the ray travel time on the identifier remains close to that in the unperturbed waveguide. In particular, it has been shown that temporal intervals covered by travel times of rays with the same identifier in the perturbed and unperturbed waveguides almost coincide. Moreover, the shelves in the unperturbed waveguide have analogs in the perturbed one: the density of points depicting arrivals of chaotic rays in the $T-p_0$ plane has maxima close to travel times corresponding to the shelves in the unperturbed function $T(p_0)$.

Another topic addressed in our study is the presence of areas in the time–depth plane where the density of points depicting ray arrivals is (i) abnormally low or (ii) abnormally high. The first effect means the presence of a gap between travel times of chaotic and regular rays. It has been found that this gap is associated with the gap in ray distribution over identifiers, i.e., with the lack of rays with particular identifiers. The latter phenomenon, in turn, is associated with a chain of ''stable islands'' seen in the phase portrait of our Hamiltonian system that blocks (for chaotic rays) a part of phase space.

The second effect corresponds to focusing of ray travel times within a comparatively small temporal interval preceding the gap. Numerical simulation shows that this phenomenon manifests itself in the appearance of a bright spot in the distribution of acoustic energy in the time–depth plane. We have found this bright spot when modeling a long-range propagation of a short sound pulse. A ray tracing has shown that the bright spot is formed by sticky rays whose trajectories have identifiers ± 120. Since the ray travel time to a significant extent is determined by the ray identifier,[21,30] ray paths with similar topologies have close times. The sticky rays mimic regular ray paths belonging to the chain of five stable islands in the phase portrait. Thus, a large percentage of rays with the identifiers $+120$ and -120 is caused by the stickiness. In turn, the presence of a large number of rays with similar topologies leads to a high concentration of ray travel times within a small temporal interval.

The ray tracing has also demonstrated another important phenomenon caused by the stickiness. It turns out that most stable rays from those that contribute to the bright spot come in chaotic jets first predicted in Ref. 47.

For now it is unclear whether the effects of gap and focusing may have analogs in more realistic environmental models. This issue requires a further investigation.

ACKNOWLEDGMENTS

This work was supported by the ONR Grant N00014-97-1-0426 and by the Russian Foundation for Basic Research under Grant No. 03-02-17246.

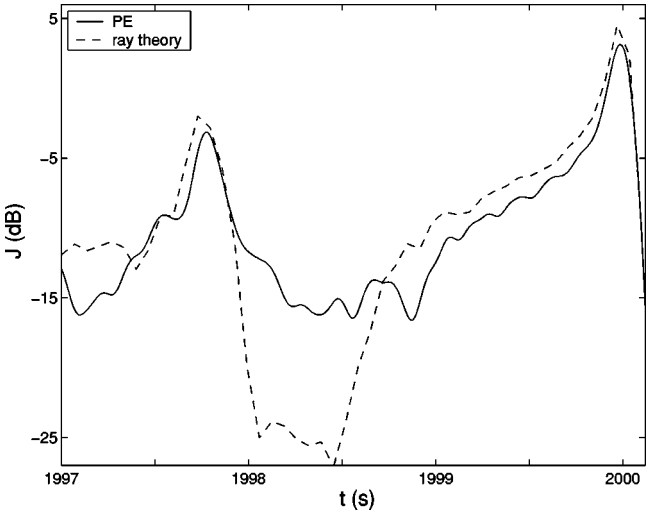


FIG. 13. Smoothed pulse intensity $J(r,t)$ defined by Eq. (48). Solid line: result obtained by direct solution of the parabolic equation. Dashed line: ray-based prediction provided by Eq. (50).

[1] S. S. Abdullaev and G. M. Zaslavsky, ''Classical nonlinear dynamics and chaos of rays in wave propagation problems in inhomogeneous media,'' Usp. Fiz. Nauk **161**, 1–43 (1991).
[2] D. R. Palmer, M. G. Brown, F. D. Tappert, and H. F. Bezdek, ''Classical chaos in nonseparable wave propagation problems,'' Geophys. Res. Lett. **15**, 569–572 (1988).
[3] K. B. Smith, M. G. Brown, and F. D. Tapper, ''Ray chaos in underwater acoustics,'' J. Acoust. Soc. Am. **91**, 1939–1949 (1992).
[4] S. S. Abdullaev, *Chaos and Dynamics of Rays in Waveguide Media*, edited by G. Zaslavsky (Gordon and Breach, New York, 1993).
[5] M. G. Brown, F. D. Tappert, and G. Goñi, ''An investigation of sound ray dynamics in the ocean volume using an area preserving mapping,'' Wave Motion **14**, 93–99 (1991).
[6] F. D. Tappert, M. G. Brown, and G. Goñi, ''Weak chaos in an area-preserving mapping for sound ray propagation,'' Phys. Lett. A **153**, 181–185 (1991).
[7] M. G. Brown, F. D. Tappert, G. Goñi, and K. B. Smith, ''Chaos in underwater acoustics,'' in *Ocean Variability and Acoustic Propagation*, edited by J. Potter and A. Warn-Varnas (Kluwer Academic, Dordrecht, 1991), pp. 139–160.
[8] D. R. Palmer, T. M. Georges, and R. M. Jones, ''Classical chaos and the sensitivity of the acoustic field to small-scale ocean structure,'' Comput. Phys. Commun. **65**, 219–223 (1991).
[9] M. G. Brown, ''Phase space structure and fractal trajectories in 11/2 degree of freedom Hamiltonian system whose time dependence is quasiperiodic,'' Nonlinear Processes Geophys. **5**, 69–74 (1998).
[10] K. B. Smith, M. G. Brown, and F. D. Tappert, ''Acoustic ray chaos induced by mesoscale ocean structure,'' J. Acoust. Soc. Am. **91**, 1950–1959 (1992).
[11] J. Simmen, S. M. Flatté, and G.-Y. Wan, ''Wavefront folding, chaos, and diffraction for sound propagation through ocean internal waves,'' J. Acoust. Soc. Am. **102**, 239–255 (1997).
[12] M. G. Brown and J. Viechnicki, ''Stochastic ray theory for long-range sound propagation in deep ocean environment,'' J. Acoust. Soc. Am. **104**, 2090–2104 (1998).
[13] M. A. Wolfson and F. D. Tappert, ''Study of horizontal multipaths and ray chaos in underwater acoustics,'' J. Acoust. Soc. Am. **107**, 154–162 (2000).
[14] M. A. Wolfson and S. Tomsovic, ''On the stability of long-range sound propagation through a structured ocean,'' J. Acoust. Soc. Am. **109**, 2694–2703 (2001).
[15] A. L. Virovlyansky and G. M. Zaslavsky, ''Evaluation of the smoothed interference pattern under conditions of ray chaos,'' Chaos **10**, 211–223 (2000).
[16] F. D. Tappert and X. Tang, ''Ray chaos and eigenrays,'' J. Acoust. Soc. Am. **99**, 185–195 (1996).
[17] G. M. Zaslavsky and S. S. Abdullaev, ''Chaotic transmission of waves and 'cooling' of signals,'' Chaos **7**, 182–186 (1997).
[18] M. Wiercigroch, M. Badiey, J. Simmen, and A. H.-D. Cheng, ''Nonlinear dynamics of underwater acoustics,'' J. Sound Vib. **220**, 771–786 (1999).
[19] I. P. Smirnov, A. L. Virovlyansky, and G. M. Zaslavsky, ''Theory and applications of ray chaos to underwater acoustics,'' Phys. Rev. E **64**, 366221 (2001).
[20] M. G. Brown, J. A. Colosi, S. Tomsovic, A. L. Virovlyansky, M. A. Wolfson, and G. M. Zaslavsky, ''Ray dynamics in long-range deep ocean sound propagation,'' J. Acoust. Soc. Am. **113**, 2533–2547 (2003).
[21] F. J. Beron-Vera, M. G. Brown, J. A. Colosi, S. Tomsovic, A. L. Virovlyansky, M. A. Wolfson, and G. M. Zaslavsky, ''Ray dynamics in a long-range acoustic propagation experiment,'' J. Acoust. Soc. Am. **114**, 1226–1242 (2003).
[22] W. Munk and C. Wunsch, ''Ocean acoustic tomography: A scheme for large scale monitoring,'' Deep-Sea Res. **26**, 123–161 (1979).
[23] J. Spiesberger and K. Metzger, ''Basin-scale tomography: A new tool for studying weather and climate,'' J. Geophys. Res. **96**, 4869–4889 (1991).
[24] T. F. Duda, S. M. Flatté, J. A. Colosi, B. D. Cornuelle, J. A. Hildebrand, Jr. W. S. Hodgkiss, P. F. Worcester, B. M. Howe, J. A. Mercer, and R. C. Spindel, ''Measured wavefront fluctuations in 1000-km pulse propagation in the Pacific Ocean,'' J. Acoust. Soc. Am. **92**, 939–955 (1992).
[25] P. F. Worcester, B. D. Cornuelle, J. A. Hildebrand, Jr. W. S. Hodgkiss, T. F. Duda, J. Boyd, B. M. Howe, J. A. Mercer, and R. C. Spindel, ''A comparison of measured and predicted broadband acoustic arrival patterns in travel time–depth coordinates at 1000 km range,'' J. Acoust. Soc. Am. **95**, 3118–3128 (1994).
[26] P. F. Worcester, B. D. Cornuelle, M. A. Dzieciuch, W. H. Munk, M. Howe, A. Mercer, R. C. Spindel, J. A. Colosi, Metzger, T. Birdsall, and A. B. Baggeroer, ''A test of basin-scale acoustic thermometry using a large-aperture vertical array at 3250-km range in the eastern north pacific ocean,'' J. Acoust. Soc. Am. **105**, 3185–3201 (1999).
[27] J. A. Colosi, E. K. Scheer, S. M. Flatté, B. D. Cornuelle, M. A. Dzieciuch, W. H. Munk, P. F. Worcester, B. M. Howe, J. A. Mercer, R. C. Spindel, K. Metzger, T. Birdsall, and A. B. Baggeroer, ''Comparisons of measured and predicted acoustic fluctuations for a 3250-km propagation experiment in the eastern north Pacific Ocean,'' J. Acoust. Soc. Am. **105**, 3202–3218 (1999).
[28] J. A. Colosi, S. M. Flatté, and C. Bracher, ''Internal-wave simulation of the forward scattering of sound in a structured ocean: Simulation and comparison with experiment,'' J. Acoust. Soc. Am. **96**, 452–468 (1994).
[29] J. L. Spiesberger, R. C. Spindel, and K. Metzger, ''Stability and identification of ocean acoustic multipath,'' J. Acoust. Soc. Am. **67**, 2011–2917 (1980).
[30] I. P. Smirnov, A. L. Virovlyansky, and G. M. Zaslavsky, ''Sensitivity of ray travel times,'' Chaos **12**, 617–635 (2002).
[31] I. P. Smirnov, A. L. Virovlyansky, and G. M. Zaslavsky, ''Wave chaos and mode-medium resonances at long range sound propagation in the ocean,'' Chaos **14**, 317–322 (2004).
[32] A. Iomin and G. Z. Zaslavsky, ''Sensitivity of ray paths to initial conditions,'' Commun. Nonlinear Sci. Numer. Simul. **14**, 401–413 (2003). http://www.sciencedirect.com/science/journal/10075704.
[33] A. J. Lichtenberg and M. A. Lieberman, *Regular and Chaotic Dynamics* (Springer, New York, 1992).
[34] G. M. Zaslavsky, *Physics of Chaos in Hamiltonian Systems* (Imperial College Press, Singapore, 1998).
[35] L. M. Brekhovskikh and Yu. P. Lysanov, *Fundamentals of Ocean Acoustics* (Springer, Berlin, 1991).
[36] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (AIP, Woodbury, NY, 1994).
[37] G. M. Zaslavsky, M. Edelman, and B. A. Niyazov, ''Self similarity, renormalization and phase nonuniformity of Hamiltonian chaotic dynamics,'' Chaos **7**, 159–181 (1997).
[38] K. B. Smith, ''Convergence, stability, and variability of shallow water acoustic predictions using the split-step Fourier parabolic equation model,'' J. Comput. Acoust. **9**, 243–285 (2001).
[39] L. D. Landau and E. M. Lifshitz, *Mechanics*, 3rd ed. (Pergamon, Oxford, 1976).
[40] M. C. Gutzwiller, ''Phase-integral approximation in momentum space and the bound states of an atom,'' J. Math. Phys. **8**, 1979–2000 (1967).
[41] S. M. Flatté, R. Dashen, W. M. Munk, K. M. Watson, and F. Zakhariasen, *Sound Transmission through a Fluctuating Ocean* (Cambridge University Press, London, 1979).
[42] L. M. Brekhovskikh, *Waves in Layered Media* (Academic, New York, 1980).
[43] W. Munk and C. Wunsch, ''Ocean acoustic tomography: Rays and modes,'' Rev. Geophys. Space Phys. **21**, 1–37 (1983).
[44] A. L. Virovlyanskii, ''Travel times of acoustic pulses in the ocean,'' Sov. Phys. Acoust. **31**, 399–401 (1985).
[45] A. L. Virovlyansky, ''On general properties of ray arrival sequences in oceanic acoustic waveguides,'' J. Acoust. Soc. Am. **97**, 3180–3183 (1995).
[46] A. L. Virovlyansky, ''Ray travel times at long ranges in acoustic waveguides,'' J. Acoust. Soc. Am. **113**, 2523–2532 (2003).
[47] V. V. Afanasiev, R. Z. Sagdeev, and G. M. Zaslavsky, ''Chaotic jets with multifractal space-time random walk,'' Chaos **1**, 143–159 (1991).

Rays, modes, wavefield structure, and wavefield stability

Michael G. Brown,[a] Francisco J. Beron-Vera, Irina Rypina, and Ilya A. Udovydchenkov
RSMAS/AMP, University of Miami, Miami, Florida 33149

(Received 19 March 2004; revised 15 July 2004; accepted 19 July 2004)

Sound propagation is considered in range-independent environments and environments consisting of a range-independent background on which a weak range-dependent perturbation is superimposed. Recent work on propagation in both types of environments, involving both ray- and mode-based wavefield descriptions, have focused on the importance of α, a ray-based ''stability parameter,'' and β, a mode-based ''waveguide invariant.'' It is shown that, when β is evaluated using asymptotic mode theory, $\beta=\alpha$. Using both ray and mode concepts, known results relating to the manner by which α (or β) controls both the unperturbed wavefield structure and the stability of the perturbed wavefield are briefly reviewed. © *2005 Acoustical Society of America.*
[DOI: 10.1121/1.1854749]

PACS numbers: 43.30.Bp, 43.30.Cq, 43.30.Ft [AIT] Pages: 1607–1610

I. INTRODUCTION

In recent work[1–6] on sound propagation in the ray limit in weakly range-dependent ocean environments it has been shown that both ray amplitude and phase (travel time) distributions are largely controlled by a property—described by the *stability parameter* α—of the background sound speed profile. Other investigators[7–18] making use of a modal description of the wavefield in both range-independent environments and range-independent environments on which weak perturbations are superimposed have shown that many wavefield properties are controlled by a property—the *waveguide invariant* β—which is defined using mode-based quantities in the background environment. In this Letter it is shown that when β is evaluated using asymptotic mode-theoretical results, $\beta=\alpha$.

In the two sections that follow α and β, respectively, are defined and briefly discussed. In the final section those properties of wavefields that are known to be controlled by α or β are briefly reviewed. To keep our presentation brief we show only equations required to derive the result $\beta\sim\alpha$ or those that provide insight into wavefield properties discussed in the final section.

II. RAYS: α

We consider underwater acoustic wavefields in three space dimensions that are excited by a point source. We assume that the environment consists of a depth-dependent background on which a weak perturbation is superimposed, so the sound speed is written as $c(z,r)=C(z)+\delta c(z,r)$. Here $-z$ is depth and the range r is the horizontal distance from the source. When $\delta c\neq 0$ it is assumed that azimuthal coupling of the wavefield is negligible, so that one need only consider propagation in the (r,z) plane. To make our discussion concrete we will assume that $C(z)$ has a single minimum, but this assumption is not necessary.

It is well known (see e.g., Refs. 5, 6) that substitution of the geometric ansatz,

$$\overline{u}(z,r,\sigma)=a(z,r)\mathrm{e}^{\mathrm{i}\sigma T(z,r)}, \tag{1}$$

where $\overline{u}(z,r,\sigma)$ is the Fourier transform of the acoustic pressure $u(z,r,t)$ and $\sigma=2\pi f$ is the acoustic frequency, into the Helmholtz equation and collecting terms in descending powers of σ yields the eikonal and transport equations. The eikonal equation can be solved for the travel time T by integrating the ray equations,

$$\frac{\mathrm{d}p_z}{\mathrm{d}r}=-\frac{\partial H}{\partial z},\quad \frac{\mathrm{d}z}{\mathrm{d}r}=\frac{\partial H}{\partial p_z},\quad \frac{\mathrm{d}T}{\mathrm{d}r}=p_z\frac{\mathrm{d}z}{\mathrm{d}r}-H \tag{2}$$

where

$$H(p_z,z,r)=-\sqrt{c^{-2}(z,r)-p_z^2}. \tag{3}$$

It follows from these equations and the relationship $\mathrm{d}z/\mathrm{d}r=\tan\varphi$, where φ is the ray angle with respect to the horizontal, that $cp_z=\sin\varphi$, so the vertical slowness p_z can be thought of as a scaled angle variable. For a point source it is convenient to label rays by their p_z value at $(z,r)=(z_0,0)$, $p_{z,0}$. The *Hamiltonian* $H=-p_r$, where p_z and p_r are the vertical and horizontal components of the slowness vector $\mathbf{p}$ with $\|\mathbf{p}\|=c^{-1}(z,r)$. In a range-independent environment $c=C(z)$, p_r is constant following a ray. The transport equation can be reduced to a statement of the constancy of energy flux in ray tubes; its solution, accounting for azimuthal spreading, can be written

$$a^2=a_0^2\frac{r_0^2}{r}|H(\partial z/\partial p_{z,0})_r|^{-1}. \tag{4}$$

Here a_0^2 is the value of a^2 at the small distance (1 m by convention) r_0 from the source. The partial derivative in (4) is evaluated keeping r fixed: $z(p_{z,0},r)$ is ray depth. In a range-independent environment an alternative form of (4) is

$$a^2=a_0^2\frac{r_0^2}{r}\frac{|p_r/p_{z,0}|}{|p_z(\partial R/\partial p_r)_z|}, \tag{5}$$

where $R(p_r,z)$ is the range of a ray, and ray depth is held constant in the partial derivative.

[a]Author to whom correspondence should be addressed. Electronic mail: mbrown@rsmas.miami.edu

For the range-independent problem $z(r)$ and $p_z(r)$ (following rays) are periodic functions. This periodic motion is most naturally described using *action-angle variables* (I,ϑ). The transformed ray equations (see, e.g., Refs. 2, 3 for details of the transformation) are

$$\frac{\mathrm{d}I}{\mathrm{d}r}=-\frac{\partial\overline{H}}{\partial\vartheta}=0, \tag{6}$$

$$\frac{\mathrm{d}\vartheta}{\mathrm{d}r}=\frac{\partial\overline{H}}{\partial I}=\omega(I), \tag{7}$$

$$\frac{\mathrm{d}T}{\mathrm{d}r}=I\omega(I)-\overline{H}(I), \tag{8}$$

where $\overline{H}(I)$ is the transformed Hamiltonian. The action,

$$I=\frac{1}{2\pi}\int_{z_-}^{z_+}\mathrm{d}z\,p_z(z)=\frac{1}{\pi}\int_{z_-}^{z_+}\mathrm{d}z\sqrt{C^{-2}(z)-p_r^2}, \tag{9}$$

where $z_\pm$ correspond to the ray upper $(+)$ and lower $(-)$ turning depths, where $C^{-1}(z_-)=C^{-1}(z_+)=p_r$. The angle variable ϑ increases by 2π each time a ray completes a cycle, and $\omega(I)=2\pi/R_l(I)$ where R_l is the range of a ray cycle (double loop),

$$R_l(p_r)=-2\pi\frac{\mathrm{d}I}{\mathrm{d}p_r}=2p_r\int_{z_-}^{z_+}\frac{\mathrm{d}z}{\sqrt{C^{-2}(z)-p_r^2}}, \tag{10}$$

where $I(p_r)$ is defined in (9). The action-angle form of the ray equations can trivially be integrated:

$$\vartheta(r)=\vartheta_0+\omega(I)r \bmod 2\pi, \tag{11}$$

$$I(r)=I_0, \tag{12}$$

$$T(r)=[I\omega(I)-\overline{H}(I)]r. \tag{13}$$

[More correctly, a term $\mathrm{d}(G-I\vartheta)/\mathrm{d}r$ should be added to the rhs of (8) where G is the generating function of the canonical transformation $(p_z,z)\rightarrow(I,\vartheta)$. The corresponding end-point corrections to (13) is both numerically insignificant and not relevant to the discussion that follows.]

The stability parameter $\alpha(I)$, whose relationship to wavefield properties will be reviewed in the final section, is defined as

$$\alpha(I)=\frac{I}{\omega(I)}\frac{\mathrm{d}\omega}{\mathrm{d}I}. \tag{14}$$

It follows from this definition, (10) and the relationship $\omega(I)=2\pi/R_l(I)$, that $\alpha(I)$ can be expressed in the form

$$\alpha(p_r)=2\pi\frac{I(p_r)}{R_l^2(I)}\frac{\mathrm{d}R_l}{\mathrm{d}p_r}. \tag{15}$$

III. MODES: β

In a stratified environment $c=C(z)$ in three space dimensions the modal decomposition of the wavefield has the form

$$\overline{u}(z,r,\sigma)=\frac{\mathrm{i}}{4}\sum_{m=0}^{\infty}\frac{\phi_m(z_0,\sigma)\phi_m(z,\sigma)H_0^{(1)}(\sigma p_{rm}r)}{\int\mathrm{d}z\,\phi_m^2(z,\sigma)}. \tag{16}$$

Here $H_0^{(1)}$ is the zeroth-order Hankel function of the first kind and the normal modes $\phi_m(z;\sigma)$ satisfy

$$\frac{\mathrm{d}^2\phi_m}{\mathrm{d}z^2}+\sigma^2[C^{-2}(z)-p_{rm}^2]\phi_m=0, \tag{17}$$

together with a pair of boundary conditions. Here $(\sigma p_{rm})^2$ is a separation constant. For sound speed profiles with a single minimum it is well known (see, e.g., Ref. 19 for the outline of a uniform asymptotic derivation) that an asymptotic analysis of (17) for modes with turning depths within the water column reveals that each ϕ_m is associated with a discrete value of the action I,

$$\sigma I(p_{rm})=m+\frac{1}{2},\quad m=0,1,2,..., \tag{18}$$

where $I(p_r)$ is defined in (9).

We now introduce, following the argument given in Ref. 19, the important concepts of modal *group* and *phase slowness*. Owing to the orthogonality of the modes, the quantity $\overline{u}_m(r,\sigma)=\int\mathrm{d}z\,\overline{u}(z,r,\sigma)\phi_m(z,\sigma)/\int\mathrm{d}z\,\phi_m^2(z,\sigma)$ isolates the contribution to the wavefield from the mode with frequency σ and mode number m. The inverse Fourier transformation of $\overline{u}_m(r,\sigma)$, weighted by $\overline{s}(\sigma)$, the Fourier transformation of the source time history $s(t)$, is denoted $u_m(r,t)$. If $\overline{s}(\sigma)$ has a narrow bandwidth, centered at σ_0, then a Taylor series expansion of $k_{rm}=\sigma p_{rm}$, with $p_{rm}=p_{rm}(\sigma)$ via Eq. (18), about σ_0 yields the result

$$u_m(r,t)=\mathrm{e}^{\mathrm{i}[k_{rm}(\sigma_0)r-\sigma_0 t]}\Psi_m(r,t), \tag{19}$$

where the envelope function $\Psi_m(r,t)$ travels at the group speed, $(\partial k_r/\partial\sigma)^{-1}$, evaluated at the center frequency and mode number m. The group slowness is defined as

$$S_g=\frac{\partial k_r}{\partial\sigma}. \tag{20}$$

Note that (19) represents a slowly varying dispersive wave train whose envelope moves at the group slowness, but within which surfaces of constant phase move at the phase slowness $k_r/\sigma=p_r$.

Consistent with the asymptotic analysis presented here,

$$S_g(p_r)=\frac{T_l(p_r)}{R_l(p_r)} \tag{21}$$

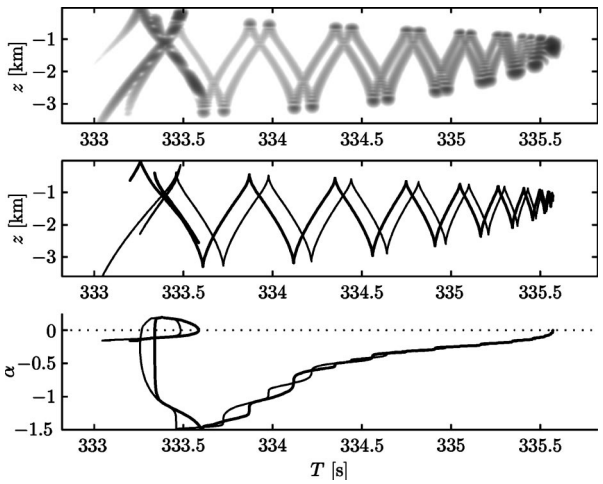


FIG. 1. Upper panel: Intensity of a transient acoustic wavefield in a range-independent deep-ocean environment as a function of depth and time at a fixed range of 500 km. The wavefield was excited by a compact source near the sound channel axis with $f_0=75$ Hz and $\Delta f\approx 30$ Hz. The dynamic range of intensity is 40 dB. Middle panel: Corresponding ray travel times vs depth. The light (heavy) line corresponds to rays with an upward (downward) inclination at the source. Lower panel: Corresponding plots of α vs T, computed by eliminating the I dependence from $\alpha(I)$ and $T(I;r)$.

(see, e.g., Refs. 20 or 19; the latter reference also includes, with additional references, the exact expression for S_g). Here $R_l(p_r)$ is given in (10) and $T_l(p_r)$ is the corresponding expression for the single-cycle travel time,

$$T_l(p_r)=2\pi I(p_r)+p_r R_l(p_r)=2\int_{z_-}^{z_+}\mathrm{d}z\,\frac{C^{-2}(z)}{\sqrt{C^{-2}(z)-p_r^2}}. \quad (22)$$

Although, according to the asymptotic expression (21), S_g depends only on the turning depths of a mode, this dictates via the quantization condition (18) that S_g is, in general, a function of both frequency σ and mode number m.

The waveguide invariant β, whose relationship to wavefield properties will be reviewed in the next section, is defined as

$$\beta=-\frac{\partial S_g}{\partial p_r}, \quad (23)$$

where p_r is the modal phase slowness (often written as S_p). Like S_g, in general β is a function of both σ and m. It follows from Eqs. (10), (21)–(23) that β can be expressed as

$$\beta(p_r)=2\pi\frac{I(p_r)}{R_l^2(I)}\frac{\mathrm{d}R_l}{\mathrm{d}p_r}. \quad (24)$$

This and other equivalent expressions for β in terms of ray-based quantities are discussed in Refs. 21, 22. A comparison of (15) and (24) reveals that $\alpha=\beta$; this is the main result of this Letter. It should be emphasized that our modal analysis is based on asymptotic results, so that we have only demonstrated the asymptotic equivalence of β and α.

IV. WAVEFIELD STRUCTURE AND STABILITY

In this section we briefly review those properties of wavefield structure and wavefield stability (to a small range-dependent perturbation) that are known to be controlled by α or β. Figure 1 shows wavefield intensity $|u(z,r,t)|^2$ in a range-independent sound channel in the depth–time plane at a fixed range, $r=500$ km, for waves excited by a transient compact source with center frequency $f_0=75$ Hz and bandwidth $\Delta f\approx 30$ Hz where $\sigma=2\pi f$. The wavefield shown was produced by solving the Thomson–Chapman[23] parabolic equation. Use of this parabolic approximation introduces some very minor distortion to the corresponding Helmholtz equation wavefield, on which our analysis is based.

Perhaps the simplest interpretation of α is that it is a measure of the rate at which small elements of the extended phase space (z,p,T) are deformed by the background flow (with $\mathrm{d}z/\mathrm{d}r$ treated as the z coordinate of a fluid motion, etc.) by shearing motion.[1,2] Although this property does not correspond directly to any observable wavefield feature, this property helps to understand other observable wavefield features. A very simple observable wavefield property controlled by α is travel time dispersion. It follows from Eq. (13) that $\mathrm{d}T/\mathrm{d}I=I(\mathrm{d}\omega/\mathrm{d}I)r=\alpha(I)\omega(I)r$. In a range-independent environment I is a ray label that increases monotonically with increasing axial ray angle, so this equation describes the rate of change of ray travel time with increasing axial ray angle. This is illustrated in the ray simulation in Fig. 1, where $\alpha<0$ for most rays; note that the ray-based prediction of travel time dispersion is clearly reproduced in the corresponding finite frequency wavefield. Note also that zeros of α correspond approximately to cusps in the (z,T) plane. At such cusps geometric amplitudes diverge. This can be seen from Eq. (5). In that expression R is the total range of a ray, which can be written as nR_l (n complete ray loops) plus end-segment corrections that depend on source and receiver depths and ray inclination (positive or negative) at the source and receiver. At long range the dominant contribution to R is nR_l. Then it follows from (5) and (15) that at caustics either $p_z=0$ or $\alpha=0$. [We emphasize that is true only in the large r asymptotic limit. At short range $(\partial R/\partial p_r)_z$ has a complicated structure, usually including a singularity when $p_z=0$, so that the application of (5) generally requires great care near such points.] The same argument reveals that at long range geometric intensities are inversely proportional to $|\alpha|$; this property is also seen (away from caustics and interference fringes) in the finite frequency wavefield shown in Fig. 1.

In the presence of a weak range-dependent perturbation, α has been shown to control both ray stability,[1] quantified by geometric amplitudes or inverse Lyapunov exponents, and several measures of travel time spreads.[2,3] The latter property is linked to the travel time dispersion property in the background environment noted above. Also, in the presence of a weak perturbation, α plays a critical role in ideas relating to ray dynamics.[4–6,24,25] References 1–4 show that unperturbed rays for which $|\alpha|$ is large are generally more sensitive to environmental perturbations than those for which $|\alpha|$ is small.

In a range-independent environment β controls modal group delay dispersion: modal group delays satisfy $T_g=S_g r$, where $T_g=T_g(m,\sigma)$ and $S_g=S_g(m,\sigma)$, so $(\partial T_g/\partial p_r)_r=-\beta r$. A different measure of dispersion is provided by the temporal duration of a narrow band pulse consisting of modes with a fixed mode number m:[19] $\Delta t=2[(\Delta f)^{-2}+r^2(\Delta f)^2(2\pi\beta I/\sigma R_l)^2]^{1/2}$. The second term in this expression describes dispersive pulse broadening, which is seen to be controlled by β (evaluated at the center frequency σ_0).

A property that is commonly used[9] to motivate the introduction of the waveguide-invariant β is that, following a surface of constant wavefield intensity, a relationship $\delta r/r=\beta\delta\sigma/\sigma$ is satisfied. This relationship describes, for instance, the perturbation in range of an intensity peak due to a perturbation to the acoustic frequency. This relationship leads naturally to the defining equation (23). (Note that in some references, including the original work by Chuprov,[7] $-\partial S_g/\partial p_r$ is defined to be $1/\beta$ rather than β.) This defining equation (23) implies a power-law dependence of the difference in eigenvalues of the Helmholtz equation on the acoustic frequency, $k_{rn}-k_{rm}=a_{mn}\sigma^{-\beta}$, where m and n are mode numbers and a_{mn} is a constant. More generally, Grachev[10] argued that $k_{rn}-k_{rm}=b_{mn}\eta^{\beta/\gamma}\sigma^{-\beta}$, where η is a parameter, such as water depth, describing the environment, and γ is a constant. Consistent with this dependence is the relationship $\delta r/r=\beta\,\delta\sigma/\sigma+(\beta/\gamma)\delta\eta/\eta$. It is important to note (see also Ref. 11) that because $\beta=\beta(\sigma,m)$, when modes corresponding to different values of β are not temporally

resolved,[9] application of the above results may be difficult. Much work relating to β has focused on the homogeneous, constant depth waveguide for which β is independent of σ and m for modes corresponding to small values of the axial ray angle; for that problem β can be approximated as a constant.

In an environment consisting of a range-independent background on which a range-dependent perturbation is superimposed, β controls the spread of modal group delays owing to mode coupling. This is seen by noting that the total group delay of modal energy that has been scattered among mode numbers m_i, $i=1,...,N$, is $T_g=\Sigma_i S_g(m_i,\sigma)r_i$ where $r=\Sigma_i r_i$ is the range. It follows that

$$T_g = S_g(\overline{m},\sigma)r + \frac{\partial S_g}{\partial p_r}(\overline{m},\sigma)\frac{\partial p_r}{\partial I}\sum_i (I_i-\overline{I})r_i$$

$$= S_g(\overline{m},\sigma)r + \beta(\overline{m},\sigma)\frac{2\pi}{R_l(\overline{p_r})}\sum_i (I_i-\overline{I})r_i, \qquad (25)$$

where $\overline{I}=(\overline{m}+1/2)/\sigma=I(\overline{p_r})$ and $\overline{m}$ is a suitably chosen mode number (e.g., the mode number at $r=0$ or the average mode number). Also in range-dependent environments, the relationship $\delta r/r=\beta\delta\sigma/\sigma+(\beta/\gamma)\delta\eta/\eta$ has been exploited in the interpretation of measurements in shallow water waveguides[15–18] and time reversal applications.[11,12–14,26] Note that large $|\beta|$ is associated with a high degree of wavefield sensitivity, e.g., high sensitivity of modal group delays to an environmental perturbation, or high sensitivity of the location of a time reversal focus to an environmental perturbation $\delta\eta$.

In retrospect, it is not surprising that there should be a simple connection between α and β inasmuch as, whether one adopts a ray or modal wavefield description, the object of study, the wavefield, is the same. The result $\beta\sim\alpha$ is an aspect of *ray-mode duality* (see, e.g., Ref. 20). We expect that there are many wavefield properties that we have overlooked that are controlled by this parameter. We are unaware of any parameter that characterizes an acoustic environment whose important rivals that of α (or β).

ACKNOWLEDGMENTS

This research was supported by Code No. 321 OA of the Office of Naval Research and the National Science Foundation under Grant No. CMG-0417425.

[1] F. J. Beron-Vera and M. G. Brown, "Ray stability in weakly range-dependent sound channels," J. Acoust. Soc. Am. **144**, 123–130 (2003).
[2] F. J. Beron-Vera and M. G. Brown, "Travel time stability in weakly range-dependent sound channels," J. Acoust. Soc. Am. **115**, 1068–1077 (2004).
[3] A. L. Virovlyansky, "Ray travel times at long range in acoustic waveguides," J. Acoust. Soc. Am. **113**, 2523–2532 (2003).
[4] F. J. Beron-Vera, M. G. Brown, J. A. Colosi, S. Tomsovic, A. L. Virovlyansky, M. A. Wolfson, and G. M. Zaslavsky, "Ray dynamics in a long-range acoustic propagation experiment," J. Acoust. Soc. Am. **114**, 1226–1242 (2003).
[5] I. P. Smirnov, A. L. Virovlyansky, and G. M. Zaslavsky, "Theory and applications of ray chaos to underwater acoustics," Phys. Rev. E **64**, 036221 (2001).
[6] M. G. Brown, J. A. Colosi, S. Tomsovic, A. L. Virovlyansky, M. Wolfson, and G. M. Zaslavsky, "Ray dynamics in long-range deep ocean sound propagation," J. Acoust. Soc. Am. **113**, 2533–2547 (2003).
[7] S. D. Chuprov, "Interference structure of an acoustic field in a layered waveguide," in *Acoustics of the Ocean*, edited by L. M. Brekhovskikh and I. B. Andreevoi (Nauka, Moscow, 1982), pp. 71–91.
[8] G. L. D'Spain and W. A. Kuperman, "Application of waveguide invariants to analysis of spectrograms from shallow water environments that vary in range and azimuth," J. Acoust. Soc. Am. **106**, 2454–2468 (1999).
[9] L. M. Brehovsikh and Y. P. Lysanov, *Fundamentals of Ocean Acoustics*, 2nd ed. (Springer-Verlag, Berlin, 1991).
[10] G. A. Grachev, "Theory of acoustic field invariants in layered waveguides," Acoust. Phys. **39**, 33–35 (1993).
[11] S. Kim, W. A. Kuperman, W. S. Hodgkiss, M. C. Song, G. F. Edelmann, and T. Akal, "Robust time reversal focusing in the ocean," J. Acoust. Soc. Am. **114**, 145–157 (2003).
[12] W. S. Hodgkiss, H. C. Song, W. A. Kuperman, T. Akal, C. Ferla, and D. R. Jackson, "A long range and variable focus phase conjugation experiment in shallow water," J. Acoust. Soc. Am. **105**, 1597–1604 (1999).
[13] G. L. D'Spain, J. J. Murray, W. S. Hodgskiss, N. O. Booth, and P. W. Schey, "Mirages in shallow water matched field processing," J. Acoust. Soc. Am. **105**, 3245–3265 (1999).
[14] H. C. Song, W. A. Kuperman, and W. S. Hodgkiss, "A time-reversal mirror with variable range focusing," J. Acoust. Soc. Am. **103**, 3234–3240 (1998).
[15] D. E. Weston and K. J. Stevens, "Interference of wide-band sound in shallow water," J. Sound Vib. **21**, 57–64 (1972).
[16] V. M. Kuz'kin, "The effect of variability of ocean stratification on a sound field interference structure," Acoust. Phys. **41**, 300–301 (1995).
[17] V. M. Kuz'kin, A. V. Ogurtsov, and V. G. Petnikov, "The effect of hydrodynamic variability on frequency shifts of the interference pattern of a sound field in shallow water sea," Acoust. Phys. **44**, 77–82 (1998).
[18] V. M. Kuz'kin, "Frequency shifts of the interference pattern of a sound field in shallow water," Acoust. Phys. **45**, 224–229 (1998).
[19] M. G. Brown, I. Viechnicki, and F. D. Tappert, "On the measurement of modal group time delays in deep ocean," J. Acoust. Soc. Am. **100**, 2093–2102 (1995).
[20] W. Munk and C. Wunsch, "Ocean acoustic tomography: rays and modes," Rev. Geophys. Space Phys. **21**, 777–793 (1983).
[21] G. L. D'Spain, G. L. Rovner, P. Gerstoft, W. A. Kuperman, and W. S. Hodgkiss, "Determination of the waveguide invariant in general environments," US Navy J. Underwater Acoust. **51**, 123–142 (2002).
[22] P. Gerstoft, G. L. D'Spain, W. A. Kuperman, G. L. Rovner, and W. S. Hodgkiss, "Calculating the waveguide invariant β by ray-theoretic approaches," Marine Physical Laboratory TM-468, University of California, San Diego, 2002.
[23] 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).
[24] V. I. Arnold, *Mathematical Methods of Classical Mechanics*, 2nd ed. (Springer-Verlag, 1989).
[25] M. Tabor, *Chaos and Integrability in Nonlinear Dynamics* (Wiley, New York, 1989).
[26] W. A. Kuperman, S. Kim, G. F. Edelman, W. S. Hodgkiss, H. C. Song, and T. Akal, "Group and phase speed analysis for predicting and mitigating the effects of fluctuations," in *Impact of Litoral Environmental Variability on Acoustic Predictions and Sonar Performance*, edited by N. G. Pace and F. B. Jensen (Kluwer, Dordrecht, 2002), pp. 279–286.

Entropy and scintillation analysis of acoustical beam propagation through ocean internal waves

Andrey K. Morozov and John A. Colosi[a)]
Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

(Received 9 January 2004; revised 5 August 2004; accepted 5 August 2004)

Parabolic equation numerical simulations of waveguide acoustical beam propagation in an ocean of Garrett–Munk internal waves are used to examine the range evolution of beam properties such as beamwidth (both spectral and spatial), Shannon entropy, and scintillation index, as a function of beam angle. Simulations are carried out at 250- and 125-Hz acoustic frequencies. The ray trajectories associated with these beams are predominantly chaotic or exponentially sensitive to initial conditions and/or medium perturbations. At long range near saturation, the finite-frequency beams show a constant rate of change of Shannon entropy with range, independent of acoustic frequency. This full-wave rate of entropy is of the same order of magnitude as the average rate of entropy for the ray trajectories associated with this beam. Finite-range Lyapunov exponents provide the estimates of ray entropy rate or Kolmogorov–Siani entropy. The correspondence between full-wave and ray entropies suggests a full-wave manifestation of ray chaos, but only once statistical saturation is obtained. In spite of this correspondence, the simulated acoustical beams expand diffusively not exponentially (or explosively).
[DOI: 10.1121/1.1854571]

PACS numbers: 43.60.Cg, 43.30.Ft, 43.30.Cq [AIT] Pages: 1611–1623

I. INTRODUCTION

Scattering by small-scale ocean processes like internal waves imposes the ultimate limitations on large-scale ocean acoustic remote sensing and matched-field processing. Further, the stochastic properties of the sound field scattered by internal waves are poorly understood in long-range propagation as conditions of saturation are established (Colosi *et al.*, 1999a,b). Recent progress in utilizing ray methods to understand scattering processes in long-range acoustic propagation suggests there is an exponential sensitivity to initial conditions and a rapid growth of acoustic field complexity with a scale of a few hundred kilometers (Brown *et al.*, 2003; Beron-Vera *et al.*, 2003). However, as is well known in other fields of wave propagation, finite-frequency effects can slow, suppress, or mask this increasing complexity (Casati, 1996). The objective of this paper is to examine the phenomenon of ocean acoustic ray chaos and its potential manifestations in finite-frequency wave fields, by the method of numerical simulation. This same question has received much attention in the quantum chaos field (Gianonni *et al.*, 1989; Casati and Chirikov, 1995) and is often termed wave chaos.

From an ocean acoustic perspective and motivated by the work of Wolfson and Tappert (2000), this problem was initially considered by Tappert (2003), who hypothesized that the width of a narrow-angle beam propagating through a strong mesoscale sound-speed field would increase exponentially (explosively) with range at a rate that is predicted by the geometric Lyapunov exponent. Tappert (2003) examined this problem in detail for the case of no waveguide and a single-scale random medium where he showed theoretically the possibility of observing explosive beam growth between the limits of short-range dominance by diffraction and long-range dominance of uncorrelated multipath (or saturation). Indeed, some recent results for an equivalent problem involving a 2D electron gas have demonstrated this exponential growth (Topinka and Westervelt, 2003).

Inspired by Tapperts' work, this problem of ocean acoustic wave chaos is examined in the presence of a waveguide and realistic ocean sound-speed fluctuations for which the problem of finding a full-wave manifestation of ray chaos can be different. Using parabolic equation (PE) and ray theory numerical simulations in the presence of ocean sound-speed fluctuations obeying the Garrett–Munk (GM) internal wave model, PE beam simulations are done to mimic a small-angle bundle of rays. To test the frequency dependence of the results, simulations at the two frequencies of 125 and 250 Hz are done. An example of unperturbed and perturbed 250-Hz PE beams for an axial angle of about 5 deg is shown in Fig. 1 (bottom panels). Unlike the previous analysis (Tappert, 2003), more acoustic observables than simply beamwidth are considered. Importantly, we consider the Shannon entropy of vertical profiles of the beam complex envelope as a function of range, $E(r)$, because of the close connection between the instability of the ray equations (i.e., Lyapunov exponent) and the Kolmogorov–Sinai (KS) entropy (actually a rate). Recent results have shown that the rate of change of Shannon entropy, $E(t)$, for some idealized dynamical systems is closely related to the KS entropy over an intermediate regime of times (Latora and Baranger, 1999).[1] Thus, the rate of Shannon entropy for our finite-frequency numerical simulations can be directly compared to ray simulation results of average finite-range Lyapunov exponent (a measure of KS entropy).

In this study, it is found that acoustical beams in an

[a)]Author to whom correspondence should be addressed. Electronic mail: jcolosi@whoi.edu

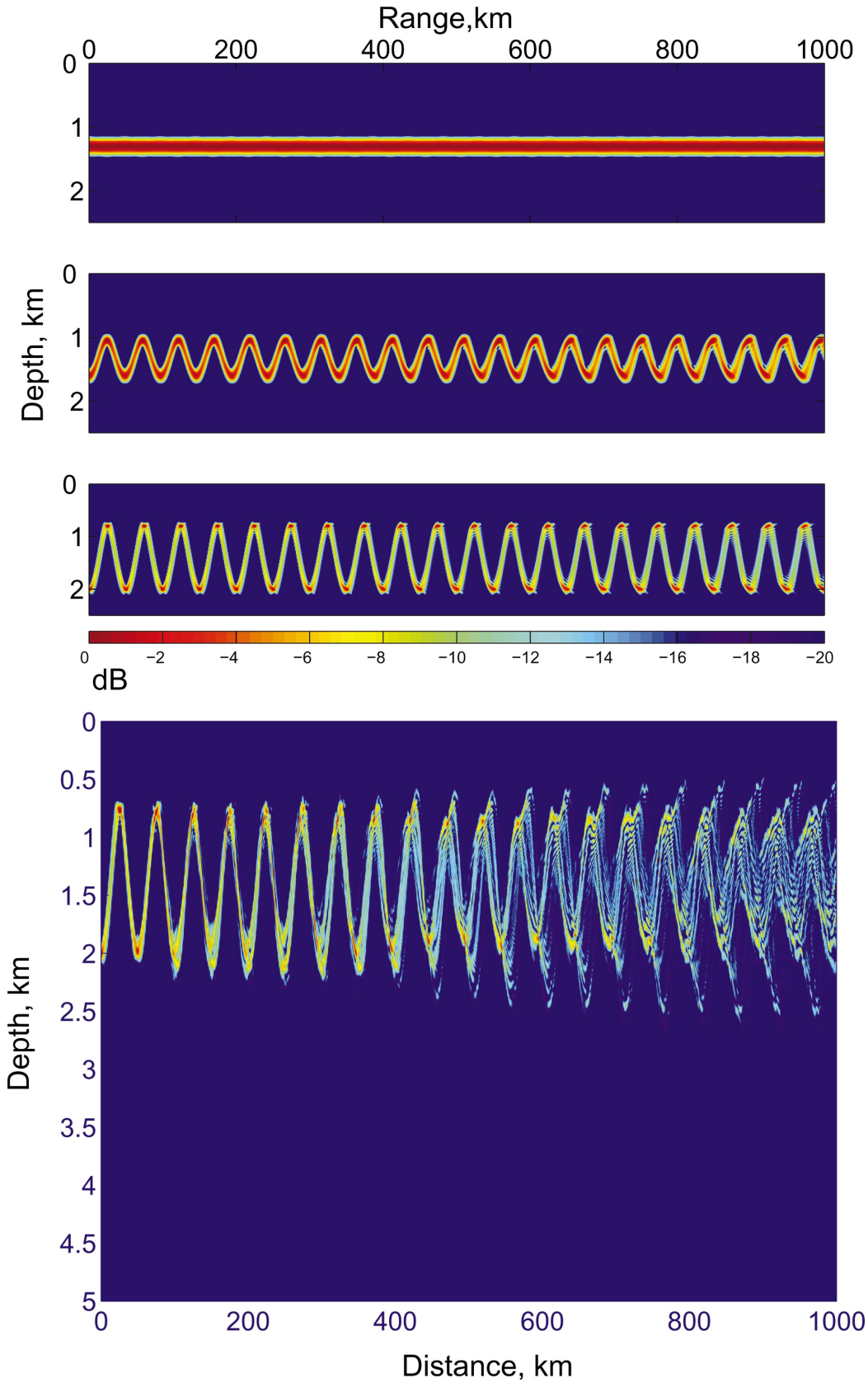


FIG. 1. Examples of parabolic equation calculations of four different weakly divergent acoustical beams. In the upper three panels the simulations were done without internal wave sound-speed perturbations and the source was placed at depths of 1300 m (axial depth; upper), 1600 m (second upper), and 2000 m (third upper). In the lowest panel is an example of the 2000-m source depth acoustical beam with internal-wave sound-speed perturbations. These simulations were carried out at 250 Hz using the Munk canonical profile. Intensity is in dB referenced to the maximum intensity.

ocean waveguide with realistic sound-speed perturbations from internal waves do not expand exponentially (or explosively), but they expand diffusively (Fig. 1). Moreover, it is also found that as the beam fills the waveguide and the wave function behaves like Gaussian random noise (i.e., full saturation), $E(r)$ varies linearly with range. The rate of Shannon entropy in this linear regime is computed to be between 0.06 and 0.11 bit/km independent of frequency, and the smaller values are associated with the higher angle beam.[2] These values are to be compared to at least twice the KS entropy (which lacks phase information) and are computed to be between 0.02 and 0.04 bit/km, with again smaller entropy rates associated with higher angle rays. Clearly, the wave field entropy cannot grow indefinitely as this system is bounded by the ocean surface and bottom. We surmise that the entropy will grow linearly until acoustic energy fills the entire water column, at which time the entropy should stabilize. The diffusive beam spreading and the linear rate of Shannon entropy are the same in both the 125- and 250-Hz simulations. These results suggest a finite-frequency manifestation of ray chaos, but only after the establishment of saturation, where the scintillation index is close to 1. Thus, Tappert's conjecture that there should be a full-wave manifestation of ray chaos is confirmed but in a different way than anticipated from the homogeneous background case. That is, in the case of a waveguide, the full-wave manifestation of ray chaos is seen in the entropy, not in the beam spread.

The organization of the paper is as follows. In Sec. II the acoustic propagation and internal wave models which are the basis of the Monte Carlo numerical simulation method are discussed. The narrow angle beam is described in Sec. III. In Secs. IV and V the range evolution of beam intensity statistics and beamwidth are described. Section VI presents results on beam entropy, while Sec. VII presents results on the connection between wave and ray entropies. Summary and conclusions are given in Sec. VIII.

II. SOUND PROPAGATION SIMULATION

The sound speed as a function of range (r), and depth (z) is defined by the sum

$$c(r,z)=c_m(z)+\delta c(r,z), \tag{1}$$

where $c_m(z)$ is the mean sound-speed profile and $\delta c(r,z)$ is a random perturbation induced by internal waves. For simplicity, a canonical model is used for the mean profile (Munk, 1976) and, the internal wave contribution is modeled using the Garrett–Munk spectrum which has been described in detail by Colosi and Brown (1998). For the buoyancy frequency, $N(z)$, a canonical form is also used such that, $N=6\exp(z/1300)$ cph. For the internal wave simulations 50 vertical modes and 2^{12} horizontal wave numbers in each of the horizontal directions are used, with a horizontal step of 25 m. The GM spectrum is cut off at a maximum horizontal wave number of 4.0 cpkm, and the internal wave energy was normalized so that at the depth where $N=3$ cph the rms internal wave displacement is equal to 7.3 m. The potential gradient of sound speed is proportional to $N^2(z)$ as in Colosi and Brown (1998). Two hundred realizations of the random sound-speed fields were calculated and used to compute 200 realizations of the acoustic pressure field for the various beam geometries out to 2000-km range. The acoustic frequencies were 125- and 250 Hz. The sound-pressure propagation modeling was carried out using a wide-angle parabolic equation (PE) approach (Jensen *et al.*, 2000). The PE solution provides the complex envelope $\psi(r,z)$, which is assumed to vary slowly in range. The full-wave field can be written as

$$p(r,z)=\psi(r,z)H_0^1(k_0r), \tag{2}$$

where H_0^1 is the Hankel function, which satisfies the well-known Bessel equation. The Hankel function can be replaced by its asymptotic form for $k_0r\gg 1$, giving

$$H_0^1(k_0r)=\left(\frac{2}{\pi k_0r}\right)^{1/2}\exp(i(k_0r-\pi/4)). \tag{3}$$

In the subsequent analysis all cylindrical spreading factors were removed. In this paper a numerical implementation of the PE solution is used which is based on the Greene approximation as described in Jensen *et al.* (2000). For the starting field an analytical Gaussian source function was used of the form

$$\psi(0,z)=\left(\frac{2k_0^2}{\pi\Delta^2}\right)^{1/4}\exp(-k_0^2(z-z_s)^2/\Delta^2), \tag{4}$$

where z_s is the source depth, and the width factor Δ was adjusted to values between 60 and 100 to achieve weakly diverging beams, as will be discussed in the next section.

Before we leave this subject, it must be noted that a few simulations were carried out using the standard parabolic equation (SPE) for the envelope which has the form

$$\frac{i}{k_0}\frac{\partial\psi}{\partial r}=-\frac{1}{2k_0^2}\frac{\partial^2\psi}{\partial z^2}-U(r,z)\psi=0, \tag{5}$$

$$U(r,z)=\frac{1}{2}\left[\left(\frac{c_0}{c(r,z)}\right)^2-1\right]\simeq-\left[\frac{c_m-c_0}{c_0}+\frac{\delta c}{c_0}\right], \tag{6}$$

where $k_0=\omega/c_0$ is a reference wave number. It was found that the statistics of the beams was insensitive to whether the SPE or the larger-angle PE were used in the computation.

III. WEAKLY DIVERGENT BEAMS

The existence of weakly divergent bundles of rays was discovered and described by Brekhovskih *et al.* (1995), Goncharov and Kurtepov (1994), and Petukhov (1994). Long-range propagation experiments often showed strong increases of sound intensity at different depths far from the axis of the underwater channel. This phenomenon was explained by focusing of acoustical energy along weakly divergent bundles of rays. Rays comprising this bundle are grouped around a central ray with the turning point at the source depth, i.e., the beam emerges horizontally from the source. When a sound source is near the sound-channel axis,

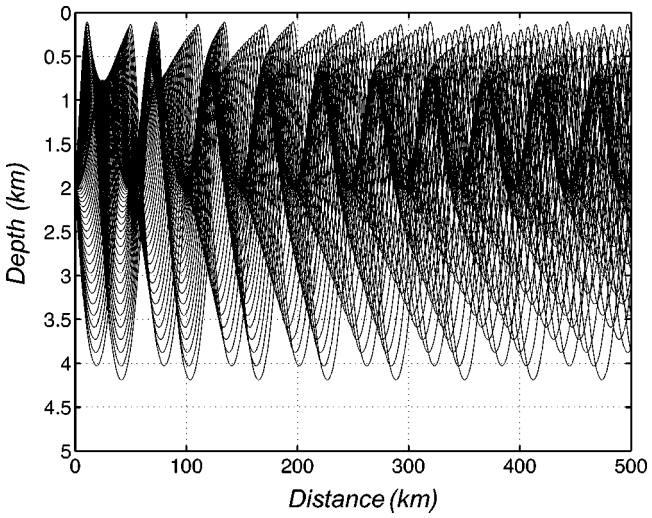


FIG. 2. Concentration of ray trajectories along the weakly divergent bundle of rays. The numerical simulation was performed for a Munk canonical sound-speed profile.

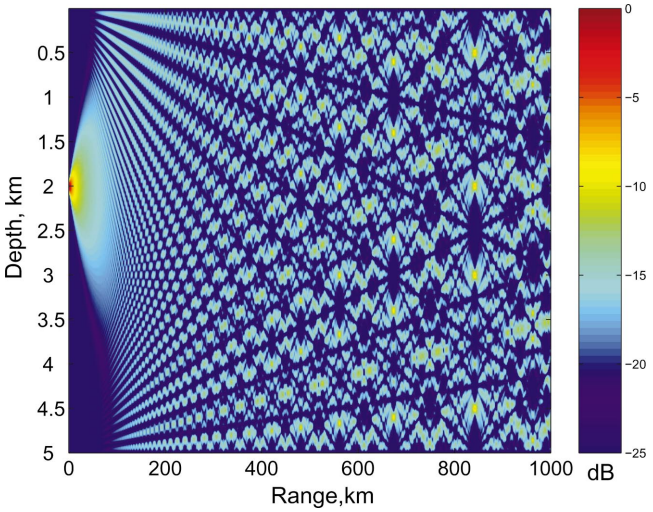


FIG. 4. Sound emitted by the same aperture as the 2000-m source beam in Fig. 1 (lower), but here there is a constant background sound speed.

this bundle corresponds to the usual energy concentration along the channel axis. The angles of weakly divergent bundle of rays are close to points of smooth extremum of the angular dependence of a ray cycle length (Brekhovskih *et al.*, 1990). It must be noted, however, that for the adopted canonical profile in this study, the Brekhovskih condition does not apply since the ray cycle distance is a monotonic function of ray grazing angle. Nevertheless, in Fig. 2 one can see how energy and ray trajectories are concentrating near the direction of weakly divergent bundle of rays for this canonical waveguide (Munk, 1976). Such concentration can be found in any horizontally homogeneous ocean profile and this property remains (in the adiabatic approach) for small horizontal gradients of sound speed. A further geometrical acoustics interpretation of a narrow weakly divergent bundle of rays is demonstrated in Fig. 3. Suppose that the source radiation width is confined by crossing trajectories of two acoustic rays shifted in the horizontal direction by a small distance roughly equal to the horizontal width of bundle; then, the sound energy will spread in the narrow path bounded by those rays. A sound source with such directivity will form a narrow weakly divergent beam radiated at zero

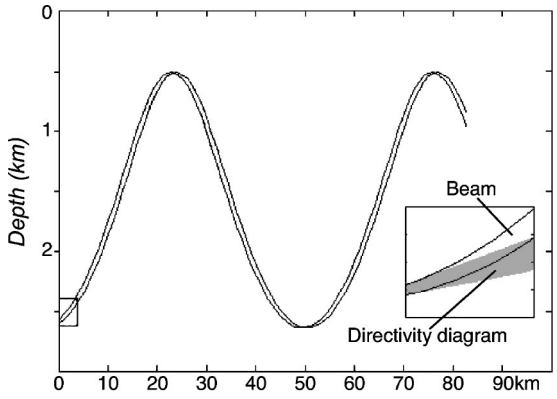


FIG. 3. Ray diagram of two paths slightly displaced in range to demonstrate generation of a weakly divergent beam.

initial angle. This property can be used to simulate a very narrow-angle beam emitting from a rather small, finite vertical aperture. It can be achieved by variation of the initial Gaussian beamwidth, Δ, until the condition of a weakly divergent beam is satisfied. Each turning point acts like a local focusing lens.

The focusing effect partially compensates the diffractive expansion, thereby forming the weakly divergent beam. An example of a 250-Hz weakly divergent beam is shown in Fig. 1. The vertical dimension of the 2000-m-depth sound source aperture is roughly 150 m ($\Delta=60$). The sound from a source with the same aperture as Fig. 1 was numerically simulated in a homogeneous medium, where H is the ocean depth (Fig. 4). A comparison between Figs. 1 and 4 shows that in a waveguide even a relatively small aperture sound source can concentrate energy in a very narrow beam. The same property can be used in receiving arrays for forming super-resolution directivity by taking into account the sound-speed profile $c(z)$ measured near the receiver (Dzieciuch, Munk, and Worcester, 2001).

The question naturally arises: Can these narrow beams be constructed in such a way as to be absolutely nondivergent? The previous example (Fig. 2) shows that a nondivergent beam is possible in the high frequency or ray limit, but the realization of such a beam at finite frequencies is an interesting problem in underwater acoustics. Nevertheless, on a practical note experimentally, it appears that these beams would be rather easy to generate, as the vertical aperture required is relatively small. An acoustic source array with such an aperture could be constructed for high frequencies, while for low frequencies the beam could be generated using a synthetic vertical aperture in which a source is moved vertically during the signal transmission.

The aforementioned concentrated narrow-angle beams are a simple and interesting subject for numerical simulation of acoustic energy scattering by internal waves, because these beams can mimic the propagation of a narrow bundle of rays, while retaining the computational efficiency of a

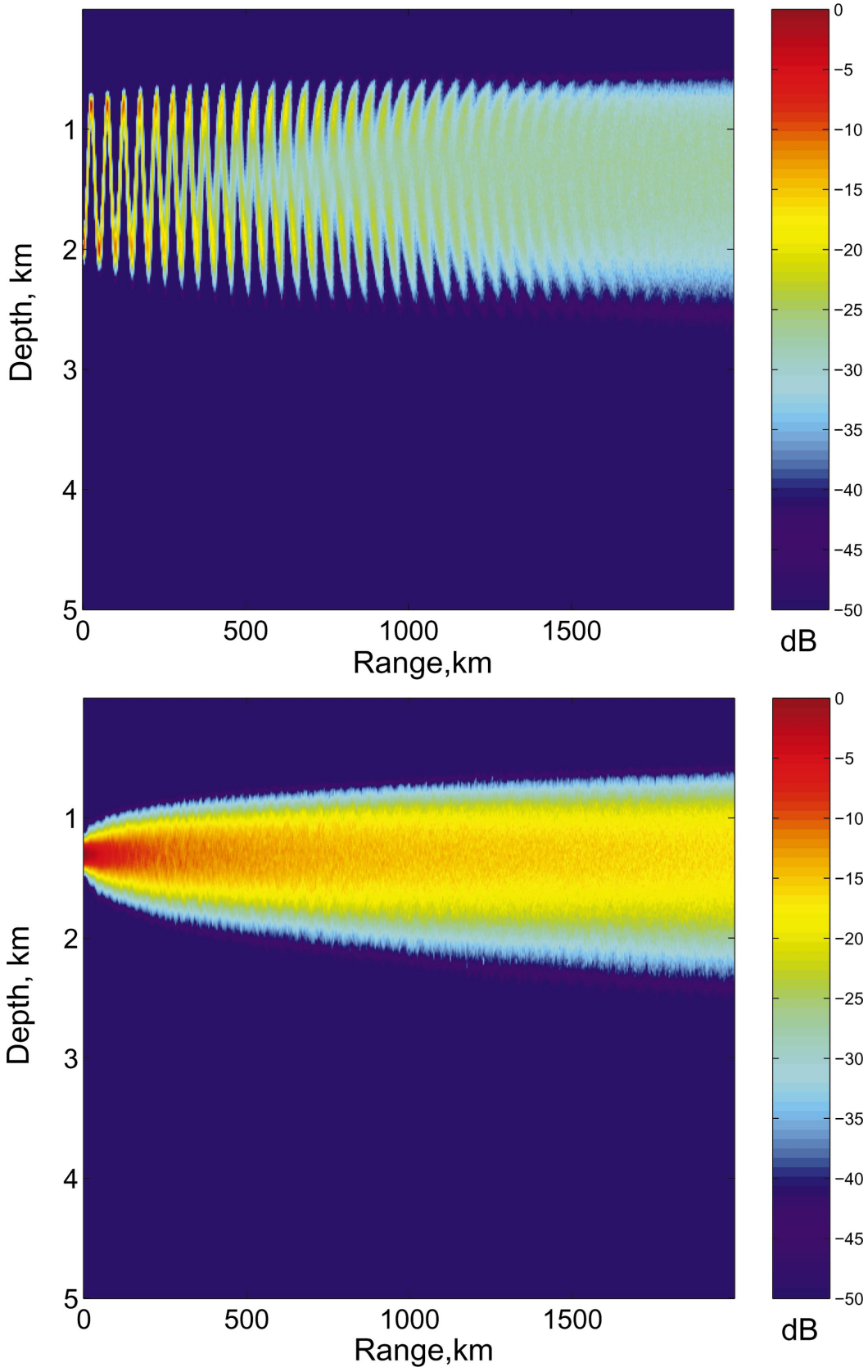


FIG. 5. Range evolution of mean intensity for the 250-Hz beam with source depths of 2000 m (upper) and 1300 m (lower).

single-frequency calculation. Clearly, broadband calculations can more narrowly define this bundle of rays by separating multipaths in time, but these calculations are beyond the scope of the present analysis.

IV. INTENSITY FLUCTUATIONS

In this section intensity fluctuations are analyzed to uncover how the real and imaginary parts of ψ approach uncor-

related Gaussian random variables or full saturation (Flatté *et al.*, 1979). In full saturation, the scintillation index

$$\mathrm{SI} = \frac{\langle I^2 \rangle}{\langle I \rangle^2} - 1, \tag{7}$$

approaches the value of 1 (Flatté *et al.*, 1976), and this approach can be from either above or below (Colosi and Baggeroer, 2004). Furthermore, the Gaussian approximation allows a simplified calculation of the Shannon entropy, which will be discussed in Sec. VI. Figure 5 shows the mean intensity, ensemble averaged over 200 internal wave realizations, for a narrow 250-Hz beam with source depths of 1300 and 2000 m. For the 2000-m beam acoustical energy is seen to scatter and diffuse from the narrow path, in-filling the axial depth region with energy and slowly spreading the energy towards the ocean boundaries. The axial beam, of course, only shows the slower outward spreading towards the boundaries. Figure 6 shows the spatial distribution of SI for the same beams. To avoid problems of division by zero, the SI is only calculated for regions of space where there is significant average intensity. At short ranges SI is increasing, while at the longest ranges near saturation values of 1 are observed more-or-less uniformly over the depth of the beam between 750 and 2000 m. In the midranges SI has significant spatial variability, and values that are much larger than 1. As the beam is breaking up, strong focuses leading to SI values significantly above 1 are observed near the caustics or vertical turning points of the beams, while near the axis the SI values are only slightly higher than 1. As the vertical caustic structure of the beam disintegrates around 1500-m range, the SI values become more uniform in depth, leading to an approach to saturation from above 1 for the 2000-m beam, but an on-axis approach from below 1 for the 1300-m beam.

A different way of examining the scintillation curves for several source depths simultaneously is to plot the SI values as a function of range, but along a curve defined by the energy centroid of the unperturbed beam. These curves are shown in Fig. 7 for the 250-Hz beams. As previously shown in Fig. 6, the 2000-m source depth beam approaches saturation from above 1, and a similar trajectory is found for a 4000-m source depth. However, for near-axial beams the approach to saturation is quite different. Here, SI approaches 1 from below, and these beams reach saturation at a much earlier range.

The 125-Hz calculation shows very similar behavior as compared to the 250-Hz calculation, except in the following respects. The high-angle beams do not show as large SI values above 1, and the initial growth of SI at short range is slower by about one-half as expected from weak fluctuation theory. For the near-axial beams the same approach to saturation from below 1 is seen, and as in the high-angle case the initial growth of SI is slower by about one-half.

V. BEAM SPREAD

Next, the issue of vertical beam spread is addressed, and it proves useful to provide the following definitions for the quantitative description of the acoustical beam evolution. First, the wave function is normalized at each range r so that

$$\psi_0(r,z) = \psi(r,z) \left(\int_{-H}^{0} |\psi(r,z)|^2 dz \right)^{-1/2}, \tag{8}$$

where H is the ocean depth. Thus, the normalized complex envelope has unit power. The central depth $\hat{z}(r)$ and rms width $w(r)$ of a narrow beam at the range r are defined as

$$\hat{z}(r) = \int_{-H}^{0} z |\psi_0(r,z)|^2 dz, \tag{9}$$

$$w(r) = 2 \left(\int_{-H}^{0} (z - \hat{z}(r))^2 |\psi_0(r,z)|^2 dz \right)^{1/2}. \tag{10}$$

The spatial spectrum of the beam can be defined using the Fourier transformation, so that

$$S(r,k_z) = \int_{-H}^{0} \psi(r,z) \exp(-ik_z z) dz. \tag{11}$$

In complete analogy to the spatial domain analysis, in wave number space there is the normalized spatial spectrum S_0, the center wave number $\hat{k}_z$, and the rms spectrum width w_s, all defined as

$$S_0(r,k_z) = S(r,k_z) \left(\int_{-(k_z)_{\max}}^{(k_z)_{\max}} |S(r,k_z)|^2 dk_z \right)^{-1/2}, \tag{12}$$

$$\hat{k}_z(r) = \int_{-(k_z)_{\max}}^{(k_z)_{\max}} k_z |S_0(r,k_z)|^2 dk_z, \tag{13}$$

$$w_s(r) = 2 \left(\int_{-(k_z)_{\max}}^{(k_z)_{\max}} (k_z - \hat{k}_z(r))^2 |S_0(r,k_z)|^2 dk_z \right)^{1/2}. \tag{14}$$

In this representation the spatial and spectral widths roughly obey an ''uncertainty'' relation, $w(r)w_s(r) \simeq 1$, so that the spatial width, $w(r)$, gives information about the large-scale beam variability, while $w_s(r)$ gives information about the high wave number, small-scale beam variability. Figures 5 and 8 show the average spatial and spectral patterns of the 250-Hz beams with source depths at 1300 and 2000 m. These figures show that the 2000-m beams are spreading in such a way as to in-fill the axial region, to make the energy density more-or-less uniform in depth between the unperturbed beam turning points, and to ''whiten'' the spatial spectrum between the maximum/minimum wave numbers of the unperturbed beam. In addition to the in-filling and spectral whitening, there is a slow expansion of the beam towards the ocean boundaries (spatial) and toward high vertical wave number (spectral), which is the dominant mechanism of change for the 1300-m beam.

Using the spatial and spectra widths of the beams, defined from Eqs. (10) and (14), the range evolution of several 250-Hz beams is shown in Fig. 9. In both cases the short ranges show deterministic variability of the beams over a beam cycle, which gradually damps out to more smooth variability with range as saturation becomes established. In all cases the beamwidth defined either spatially or spectrally increases much more slowly than exponentially. The range variations of $w(r)$ and $w_s(r)$ in Fig. 9 is very close to $r^{1/2}$,

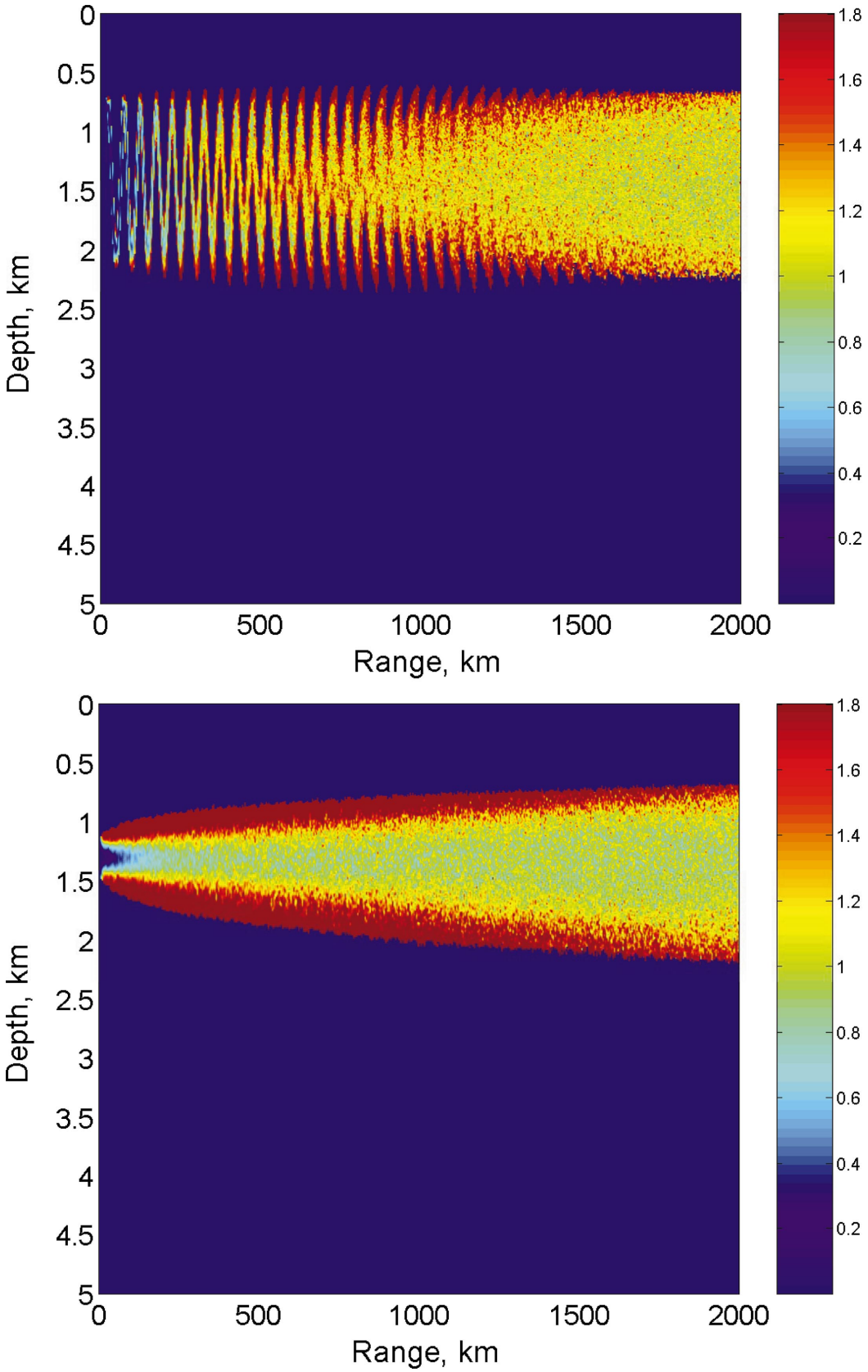


FIG. 6. Spatial distribution of scintillation index for the 250-Hz beam with source depths of 1300 m (lower) and 2000 m (upper). Scintillation index is only computed where there is significant mean intensity (see Fig. 6); thus, regions of dark blue represent areas where the scintillation index is not well defined.

implying a diffusive process. The simulations at 125 Hz, not shown here, have essentially identical beam-spreading curves, which suggests that the scattering is close to geometrical.

VI. WAVE FIELD COMPLEXITY AND ENTROPY

A useful, though nonunique, representation of the simulated wave field complexity is sought. The complexity of a

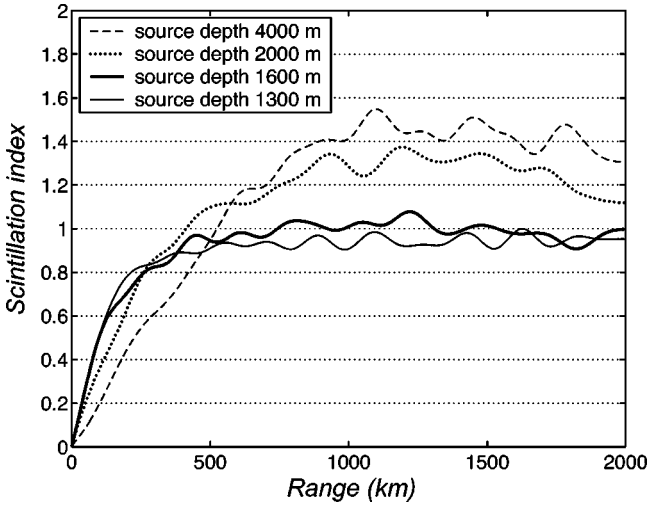


FIG. 7. Scintillation index of 250-Hz beams for four different source depths, and computed only along a line defined by the centroid of the unperturbed narrow beams.

random process depends on the number of independent components, and one such representation has already been presented, namely the spatial spectrum of the beams, $S_0(r,k_z)$, where clearly a white spectrum has more complexity than a simple line spectrum. However, it proves useful to apply the Karhunen–Loeve (KL) expansion (Van-Trees, 1968),[3] instead of the Fourier expansion because the KL expansion yields uncorrelated components in this nonstationary environment. The complex envelope of the sound-pressure field at each range step can be represented by a stochastic series with $M(r)$ significant contributions. $M(r)$ will be called the dimension of the random process, with

$$\psi(r,z)=\sum_{j=1}^{M(r)}\alpha_j(r)\Phi_j(r,z) \tag{15}$$

$$\alpha_j(r)=\int_{-H}^{0}\psi(r,z)\Phi_j(r,z)dz, \tag{16}$$

$$\lambda_j(r)\Phi_j(r,z)=\int_{-H}^{0}K(z,\tilde{z},r)\Phi_j(r,\tilde{z})d\tilde{z}, \tag{17}$$

$$K(z,\tilde{z},r)=\langle\psi(r,z)\psi^*(r,\tilde{z})\rangle, \tag{18}$$

where $K(z,\tilde{z},r)$ is the transverse correlation function at range r, and $\lambda_j(r)$ and $\Phi_j(r,z)$ are the eigenvalues and eigenfunctions of the KL expansion at range r, respectively. The functions $\lambda_j(r)$ and $M(r)$ are shown in Figs. 10 and 11. Note that curves for the angle spectrum width and number of significant KL expansion components both can be used as an approximate measure of random process.

Precisely, for a stochastic variable Γ, a logarithmic measure of the complexity of that variable is the entropy E

$$E=C-I, \tag{19}$$

where C is a constant and I is the Shannon information (Shannon, 1948) measured in bits

$$I=-\int P(\Gamma)\log_2[P(\Gamma)]d\Gamma. \tag{20}$$

Here, $P(\Gamma)$ is the probability density function (PDF). In this analysis, the complexity of the complex envelope as a function of depth z at a given range r is of interest; therefore, the Shannon information at a range r is written

$$\begin{aligned}I(r)&=-\int P(\psi(r,z_1),\psi(r,z_2),\ldots)\\&\quad\times\log_2[P(\psi(r,z_1),\psi(r,z_2),\ldots)]d\psi(r,z_1)d\psi(r,z_2)\ldots\\&=-\int\log_2[P(\psi(r,z)]dP(\psi(r,z)),\end{aligned} \tag{21}$$

where $P(\psi(r,z))$ is the probability density function for obtaining the vertical profile of ψ at range r. The calculation of sound-pressure entropy directly by Monte Carlo stochastic simulation needs a large number of realizations to define $P(\psi)$ and is therefore computationally intensive. A simplified approach based on the Gaussian approximation of the acoustical fluctuations will be used, that is, the entropy will be calculated as if a process probability density function (PDF) is spatially Gaussian. Figure 7 shows that this is a correct assumption for the ranges where SI is close to 1, that is, $r>200$ km. Since there is a one-to-one mapping between the coefficients of the KL decomposition and the complex envelope ψ, the entropy can be calculated as an entropy of a random vector of these coefficients. Generally, the probability density of a vector $X=(x_1,x_2,\ldots,x_M)$ of M Gaussian complex random values with nonsingular correlation matrix K_X has the following form:

$$P=\frac{1}{\pi^M\det(K_X)}\exp(-(X-m_x)^TK_X^{-1}(X-x_m)), \tag{22}$$

where $K_X=\langle(X-m_x)(X-m_x)^T\rangle$ is the complex correlation matrix of the complex random vector X, $m_x=\langle X\rangle$, and the superscript ^{T} means complex conjugate and matrix transpose. The entropy, E, of a Gaussian random vector can be calculated analytically, yielding (Van-Trees, 1968; Cover and Thomas, 1991)

$$E(r)=C-I(r)=C+\log_2(\pi^M\det(K_X))+M\log_2(e). \tag{23}$$

In the case of a random vector of uncorrelated components of a KL expansion, the entropy is simply the sum of the entropies of the individual KL modes, namely

$$E(r)=E(\lambda_1(r),\lambda_2(r),\ldots,\lambda_{M(r)}(r)) \tag{24}$$

$$=C+\sum_{j=1}^{M(r)}\log_2(e\pi\lambda_j(r)) \tag{25}$$

$$=C+M(r)\log_2(e\pi)+\sum_{j=1}^{M(r)}\log_2\lambda_j(r). \tag{26}$$

For equipartition of energy, $\lambda_j=\lambda$ (eigenvalue spectrum "white") the entropy is maximum, but for any other eigenvalue spectrum the entropy is less than this maximum value. The eigenvalue spectrum from the 250-Hz simulation with source depth at 2000 m is shown in Fig. 10. The spectrum expands with range, but even at the distance 2000 km it remains nonuniform ("not white"). Note that this simple

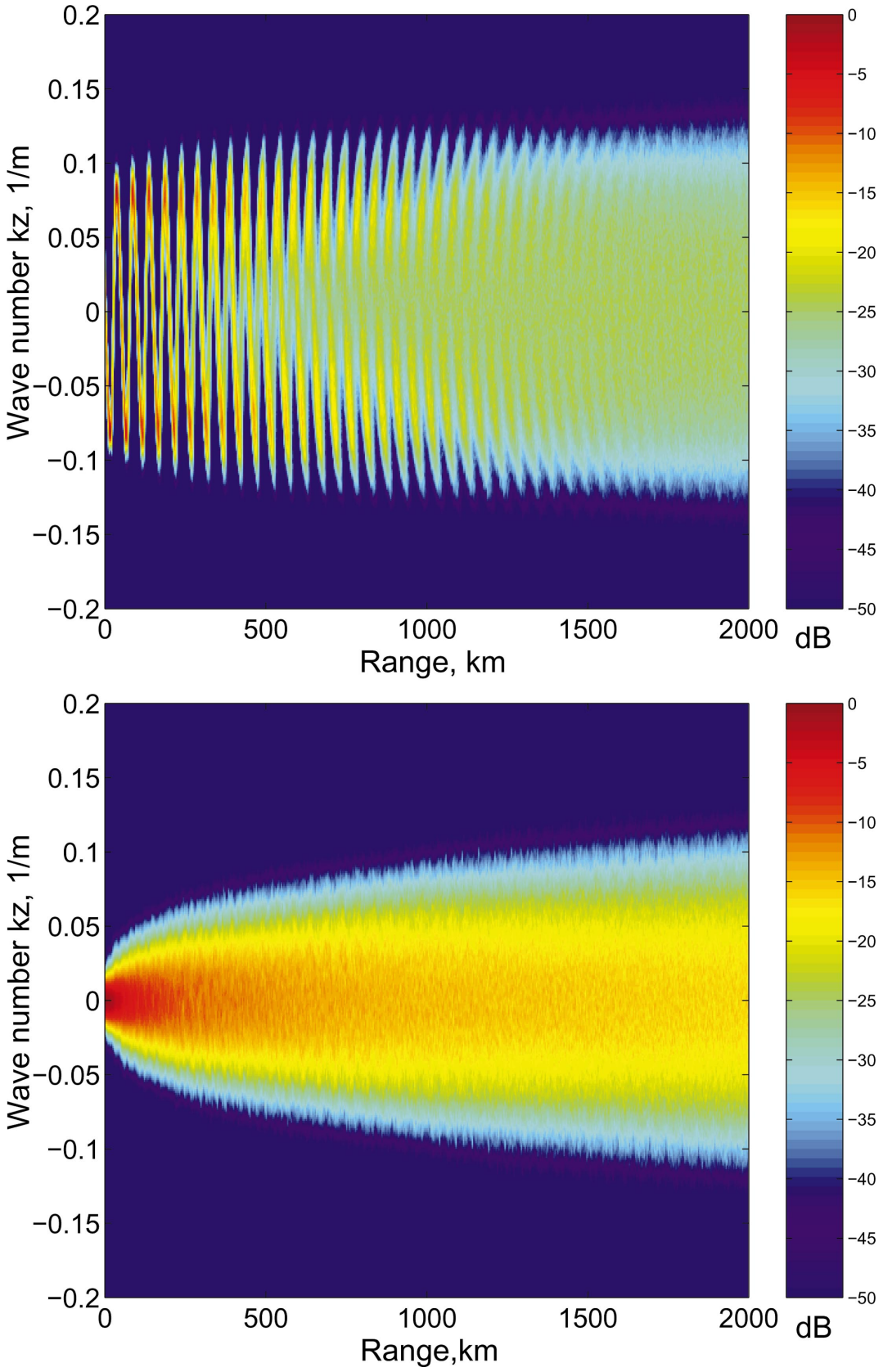


FIG. 8. Range evolution of the beam wave number k_z spectrum of the 250-Hz beam with source depths of 2000 m (upper) and 1300 m (lower).

definition of entropy is based on the assumption that the Gaussian random process is nonsingular; otherwise, a more rigorous definition must be applied in terms of absolute continuity of probability measures and Radon–Nikodim derivative (Halmos, 1950).

The entropy of the 250-Hz weakly divergent sound beams is shown in Fig. 12 for beams with different sound source depths. In all cases $C=0$ and $M(r)=100$, which is larger than the maximum number of significant KL eigenvalues at any range and for any of the acoustical beams. The

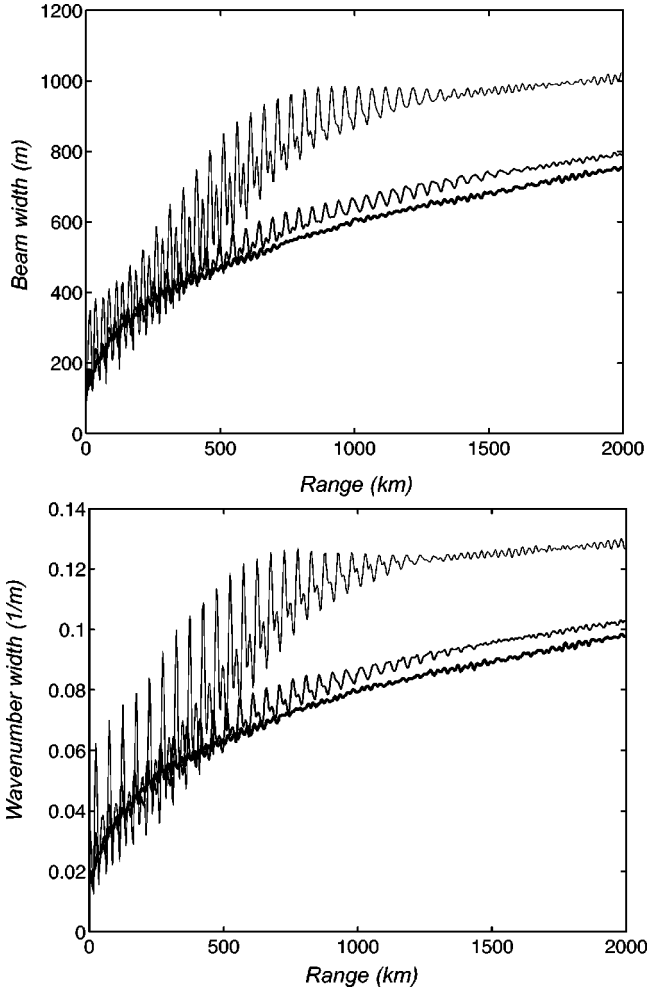


FIG. 9. Acoustical beamwidth, $w(r)$ (upper) and spectral beamwidth $w_s(r)$ (lower) as a function of propagation range for the 250-Hz beams at source depths of 1300 m (thick line), 1600 m (medium line), and 2000 m (thin line).

entropy was calculated for the ranges where the unperturbed beam crosses the sound-channel axis (i.e., 1300-m depth). This approach eliminates the deterministic oscillations of entropy at the shorter ranges, an effect seen in Fig. 9. At ranges

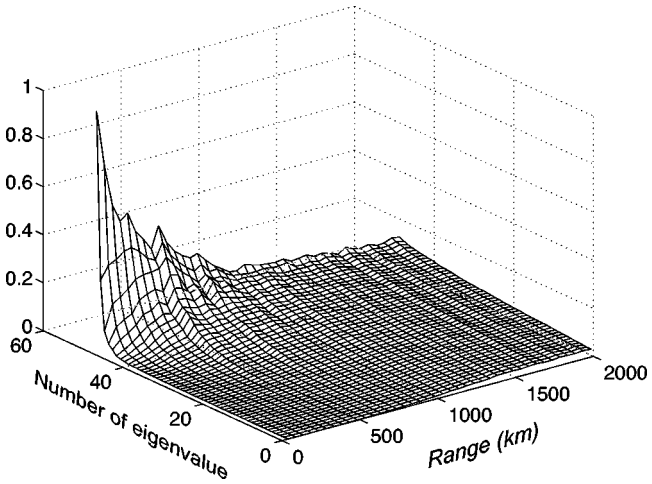


FIG. 10. Karhunen–Loeve (KL) eigenvalue spectrum as a function of range for the 250-Hz beam with source depth at 2000 m.

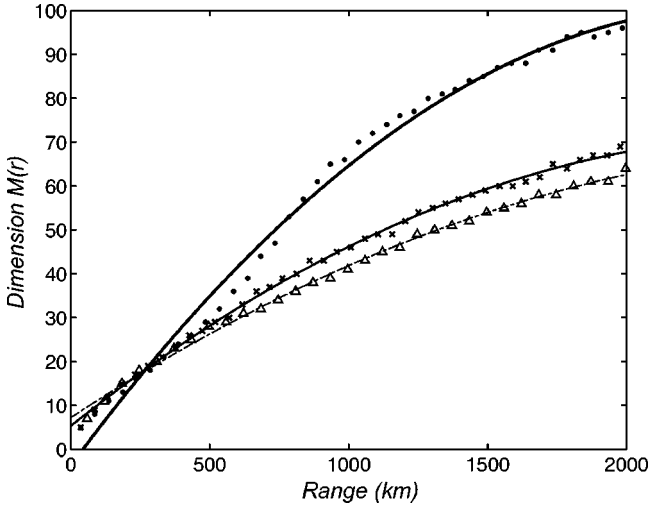


FIG. 11. Dimension $M(r)$ of random beams as a function of range. Beams at different source depths are 1300 m (thick line:axial), 1600 m (medium line), and 2000 m (thin line).

larger than roughly 1000 km, the entropy is increasing monotonically and approximately linearly for all beams, and the entropy rates in the linear regime are 0.11, 0.11, and 0.06 bit/km for the 1300-, 1600-, and 2000-m source depths, respectively. The scintillation index curves (Fig. 7) show that the linear behavior of entropy occurs very close to full saturation. Further, the linear rates are about the same for all beams, so that the anisotropy associated with beam angle is not strong, though primarily small-angle beams were considered. It must be noted that the initial rapid nonlinear rise of the entropy is an artifact of the Gaussian approximation in the entropy calculation; in the linear entropy region the Gaussian approximation is very good (see Fig. 7). Figure 12 also shows the entropy of a 125-Hz beam with the sound source at 2000 m, and there is virtually no difference be-

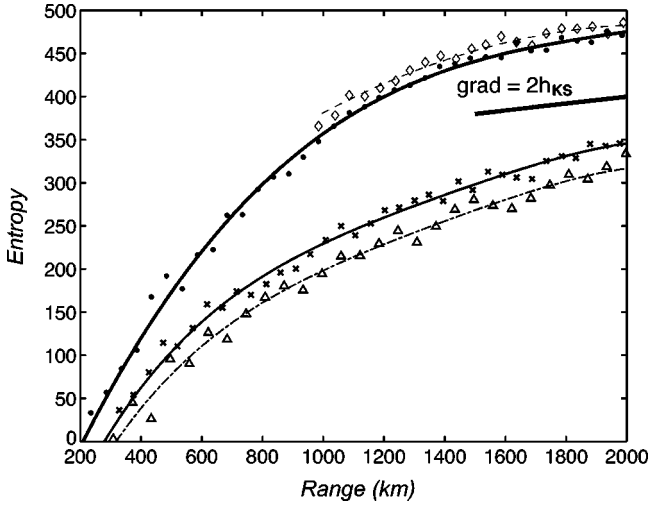


FIG. 12. Entropy of 250-Hz acoustical beams for source depths of 1300 m (thick line circles), 1600 m (medium line crosses), and 2000 m (thin line triangles). Fits in the linear region of the curves yield entropy rates of 0.11 (1300 m), 0.11 (1600 m), and 0.06 (2000 m) bit/km. The 2000-m source depth, 125-Hz entropy curve (arbitrarily vertical offset) is shown with a dash curve and diamond symbols. Also shown in solid is twice the KS entropy, $2h_{\mathrm{KS}}=0.04$ bit/km (arbitrary vertical offset) computed from the ray equations for the 1300-m axial ray.

tween the long-range entropy rate at the two frequencies. The linear increase in $E(r)$ means that the number of equiprobable states of the dynamical system (complexity) is increasing exponentially and independent of acoustic frequency, in complete accordance with ray chaos predictions (Brown *et al.*, 2003, Beron-Vera *et al.*, 2003). The frequency independence of the entropy rate not only reinforces the geometric result but also suggests that these results carry over to broadband signals. In the next section it is determined if the rate of exponential increase in complexity computed from the full-wave simulations is comparable to the rate computed from ray theory.

VII. RAY ENTROPY

It has been well established that ray trajectories in ocean acoustic propagation through internal waves are chaotic or unstable to perturbations in the medium and in the initial conditions (Brown *et al.*, 2003; Beron-Vera *et al.*, 2003). The chaotic nature of ray propagation in the ocean is quantified using the Lyapunov exponent ν_L, which is derived from a stability analysis of the ray path (Brown *et al.*, 2003), but it is often useful to express this instability using information theory (Latora and Baranger, 1999; Casati, 1996) such that the information $I(s)$ associated with a segment of ray trajectory of length s is equal asymptotically to

$$\lim_{|s|\to\infty} \frac{I(s)}{s} = h_{\mathrm{KS}}, \tag{27}$$

where h_{KS} is the Kolmogorov–Sinai (KS) entropy. Note here that KS entropy is actually an entropy rate. For bounded dynamical systems like underwater sound $\nu_L \simeq h_{\mathrm{KS}}$ (Gaspard, 1990). A consequence of the chaotic nature of ray paths is that the number of eigenrays connecting a source and a receiver will grow exponentially with range, leading to an exponential increase in wave field complexity. Thus, the exponential increase in wave field complexity is examined using the Lyapunov exponent (a measure of the KS entropy), which is to be compared to the computed rate of Shannon entropy. Another comparison between Shannon entropy and KS entropy for different dynamical systems was done by Latora and Baranger (1999).

The ray numerics are carried out as follows. Using the SPE Hamiltonian

$$H(z,p;r) = \frac{p^2}{2} + U(r,z), \tag{28}$$

where $p = \tan\theta_{\mathrm{ray}}$, the following well-known ray and stability equations are solved:

$$\frac{dz}{dr} = \frac{\partial H}{\partial p}, \quad \frac{dp}{dr} = -\frac{\partial H}{\partial z}, \tag{29}$$

$$\frac{dJ}{dr} = KJ, \tag{30}$$

where

$$J = \begin{pmatrix} \partial p/\partial p_0|_{z_0} & \partial p/\partial z_0|_{p_0} \\ \partial z/\partial p_0|_{z_0} & \partial z/\partial z_0|_{p_0} \end{pmatrix}, \tag{31}$$

$$K = \begin{pmatrix} -\partial_{zp}H & -\partial_{zz}H \\ \partial_{pp}H & -\partial_{zp}H \end{pmatrix} = \begin{pmatrix} 0 & -\partial_{zz}U \\ 1 & 0 \end{pmatrix}. \tag{32}$$

The initial condition for J at $r=0$ is the identity matrix. The Lyapunov exponent is derived from the stability equations, and is given by

$$\nu_L = \lim_{r\to\infty} \frac{\log|\mathrm{Tr}(J)|}{r}, \tag{33}$$

where $\mathrm{Tr}(J)$ is the trace or sum over the diagonal elements of the stability matrix J (Brown *et al.*, 2003; Wolfson and Tomsovic, 2001). In the present calculation the asymptotic range result is not computed, however, and a finite range stability exponent defined by

$$\nu = \frac{\log|\mathrm{Tr}(J)|}{r}, \tag{34}$$

is evaluated. This finite range stability exponent has the property that (Wolfson and Tomsovic, 2001)

$$\nu_L = \langle \nu \rangle \sim h_{\mathrm{KS}}, \tag{35}$$

where the expectation values imply averaging over realizations of the ocean internal wave field. Last, the ray initial conditions are $z(0) = 1300\,\mathrm{m}$ and $p(0) = 0$, an axial ray.

An explicit adaptive Runge–Kutta algorithm was used to calculate 200 realization of ray trajectories with the above initial conditions. The second derivatives of $U(r,z)$ were calculated using cubic spline interpolation, and the base 2 logarithm was used in Eq. (35) to get result measured in bit/km so as to be comparable with the Shannon entropy results.[2] The calculation yields an estimate of h_{KS} of 0.020 bit/km for the axial ray, decreasing linearly to 0.012 bit/km for the 2000-m source ray. These values are comparable to other calculations of Lyapunov exponent for ray propagation through ocean internal waves (Beron-Vera *et al.*, 2003; Beron-Vera and Brown, 2003). In comparing this value to the computed gradients of Shannon entropy for the complex envelope, it must be noted that the complex envelope includes information about both phase and amplitude, while h_{KS} is only a measure of the sound-field energy characteristics, such as intensity and angular power density. So, the information gradient of the full-field entropy must be at least twice the KS entropy. Thus, $2h_{\mathrm{KS}} = 0.04\,\mathrm{bit/km}$ is placed in Fig. 12 for comparison with the full-wave Shannon entropy, and order of magnitude agreement with the axial beam is seen. It must be emphasized here that the comparison between KS entropy and Shannon entropy rate is crude as we have only compared the KS entropy of one ray, while the Shannon entropy involves a bundle of rays with a range (albeit small) of initial conditions. It is known, for example, that there can be significant variability of Lyapunov exponent as a function of initial conditions due to the structure of the background sound-speed profile (Beron-Vera *et al.*, 2003; Beron-Vera and Brown, 2003).

The comparison of Shannon or physical entropies curves $E(r)$ of the complex envelope of the full-field simulation with $2h_{\mathrm{KS}}$ shows that the evolution of narrow-beam entropy has two stages. In the first stage the beamwidth and spatial spectrum are growing rapidly with oscillations due to diffrac-

tive effects. These effects are deemed diffractive because they are not described by the geometrical h_{KS} rate. The second stage occurs approximately near saturation; the entropy gradient slows down and tends to a level which is approximately twice the KS entropy of the corresponding chaotic ray trajectories. This second stage can be called chaotic from both a full-wave and ray perspective, because entropy is monotonically growing with the gradient determined by the average Lyapunov exponent. Simultaneously, beamwidth, spatial spectrum width, and dimension of the random process are all approximately linearly increasing. The stabilization of entropy gradient shows that, when fluctuations approach saturation, ray chaos phenomenon becomes the main factor determining the properties of scattering processes.

VIII. SUMMARY AND DISCUSSION

Monte Carlo stochastic simulation of acoustic scattering by internal waves allows analysis of the variability of weakly divergent beam as a function of range and beam angle. A realistic Garrett–Munk internal wave model for the induced sound-speed fluctuation was applied. The analysis was performed in terms of information theory using the physical or Shannon entropy of the complex acoustical field envelope as a logarithmic measure of wave chaos complexity and Kolmogorov–Sinai entropy as a logarithmic measure of ray chaos complexity. The relative intensity variance or scintillation index was used to quantify the approach to saturation, and for steep-angle beams SI approaches 1 from above, while for small-angle beams the approach is from below 1. At the range near saturation the rate of Shannon entropy for the complex envelope of the finite-frequency acoustical beams decreases and converges to a constant value which is approximately twice as large as the Kolmogorov–Sinai entropy or average Lyapunov exponent associated with the central ray of that beam. This correspondence suggests a full-wave manifestation of ray chaos, but only after saturation of the scintillation index has occurred. In spite of this correspondence, the simulated acoustical beams are not seen to expand exponentially (or explosively), but they expand diffusively.

The results obtained in this analysis can be compared to the work of Wolfson and Tappert (2000), and Tappert (2003), who treat the problem of wave and ray chaos for constant background sound speed (no waveguide) and a ocean mesoscale-like single-scale random medium (Gaussian correlation function). Using the theory of Gaussian beams, they predict that the width of a narrow-angle beam will increase exponentially with range at a rate that is predicted by the geometric Lyapunov exponent, which they can compute analytically. The observation of this explosive growth of the beam is masked at short range by diffractive beam spreading, and at long range by a surmised breakdown of geometrical theory near saturation. The breakdown at long range has proved very difficult to estimate, while the short-range masking by diffraction is relatively straightforward.

In the present analysis, which involves the ocean waveguide and realistic internal-wave sound speed perturbations, a very different picture emerges. Because of the restoring force of the waveguide, a narrow-angle beam cannot spread exponentially, and thus the signature of chaos must be found in some other observable. It has been shown that such a signature is found in the wave field entropy. Like the no-waveguide case, at short ranges the signature of ray chaos is not manifest in the entropy due to deterministic and diffractive effects. However, once the wave field comes very close to saturation, the variation of entropy becomes linear with range roughly as predicted by the KS entropy. Thus, while the no-waveguide case suggests an arrest of ray chaos manifestations at saturation, the waveguide case examined here shows that near-saturation is an essential element. Clearly, the wave field entropy cannot grow indefinitely as this system is bounded by the ocean surface and bottom. We surmise that the entropy will grow linearly until acoustic energy fills the entire water column, at which time the entropy should stabilize.

It also should be emphasized that our results do not preclude the use of geometric methods for describing scattering phenomena at shorter ranges where the Shannon entropy rate and KS entropy do not agree. In fact, the $\sqrt{r}$ scaling of the beamwidth curves and their insensitivity to acoustic frequency (Fig. 9) suggest strongly that the scattering is geometric.

The numerical results of this paper require a theoretical foundation which we hope to formulate in future work.

ACKNOWLEDGMENTS

This work was inspired by Fred Tappert, and we have benefited from many useful discussions with Mike Wolfson. This is Woods Hole Oceanographic Institution contribution number 11208.

[1]In ocean acoustic propagation the range variables, r, is equivalent to the time variable, t, in dynamical systems theory.

[2]The conversion from bits/km to km^{-1} simply involves a change in the base of the logarithms. Thus, an entropy rate of 0.02 bits/km would be $0.02(\log_2 e)^{-1}=0.0139\ km^{-1}$.

[3]In geophysics the Karhunen–Loeve (KL) expansion is often referred to as an empirical orthogonal function (EOF) expansion.

Beron-Vera, F. J., and Brown, M. G. (**2003**). ''Ray stability in wekaly range dependent sound channels,'' J. Acoust. Soc. Am. **114**(1), 123–130.

Beron-Vera, F. J., Brown, M. G., Colosi, J. A., Tomsovic, S., Virovlyansky, A., Wolfson, M. A., and Zaslavsky, G. M. (**2003**). ''Ray dynamics in a long range acoustic propagation experiment,'' J. Acoust. Soc. Am. **114**(3), 1226–1242.

Brekhovskih, L. M., Goncharov, V. V., and Kurtepov, V. M. (**1995**). ''Weakly divergent bundles of sound rays in the Arctic,'' Atmos. Oceanic Phys. **31**(3), 441–446.

Brown, M. G., Colosi, J. A., Tomsovic, S., Virovlyansky, A., Wolfson, M. A., and Zaslavsky, G. M. (**2003**). ''Ray dynamics in a long range deep ocean sound propagation,'' J. Acoust. Soc. Am. **113**(5), 2533–2547.

Casati, G. (**1996**). ''Quantum chaos,'' Chaos **6**(3), 391–398.

Casati, G., and Chirikov, B. V. (**1995**). *Quantum Chaos: Between Order and Disorder: A Selection of Papers* (Cambridge University Press, New York).

Colosi, J. A., and Brown, M. G. (**1998**). ''Efficient numerical simulation of stochastic internal-wave induced sound-speed perturbation fields,'' J. Acoust. Soc. Am. **103**(4), 2232–2235.

Colosi, J. A., and the ATOC Group (**1999b**). ''A review of recent results on ocean acoustic wave propagation in random media: Basin scales,'' IEEE J. Ocean. Eng. **24**(2), 138–155.

Colosi, J. A., Scheer, E. K., Flatte, S. M., Cornuelle, B. D., Dzieciuch, M. A., Munk, W. H., Worcester, P. F., Howe, B. M., Mercer, J. A., Spindel, R. C., Metzger, K., Birdsall, T. G., and Baggeroer, A. B. (**1999a**). ''Compari-

sons of measured and predicted acoustic fluctuations for a 3250-km propagation experiment in the eastern North Pacific Ocean,'' J. Acoust. Soc. Am. **105**(6), 3202–3218.
Colosi, J. A., and Baggeroer, A. B. (**2003**). ''On the kinematics of broadband multipath scintillation and the approach to saturation,'' J. Acoust. Soc. Am. **116**(6), 3515–3522.
Cover, T. M., and Thomas, J. A. (**1991**). *Elements of Information Theory* (Wiley Science, New York).
Dzieciuch, M. A., Munk, W. H., and Worcester, P. F. (**2001**). ''Turning point filters: Analysis of sound propagation on a gyre-scale,'' J. Acoust. Soc. Am. **110**(1), 135–149.
Flatté, S. M., Dashen, R., Munk, W., Watson, K., and Zachariasen, F. (**1979**). *Sound Transmission through a Fluctuating Ocean* (Cambridge University Press, Cambridge).
Gaspard, P., and Nicolis, G. (**1990**). ''Transport properties, Lyapunov exponents, and entropy per unit time,'' Phys. Rev. Lett. **65**, 1693–1696.
Giannoni, M. J., Voros, A., and Zinn-Justin, J. (**1989**). *Chaos and Quantum Physics* (Elsevier Science, Amsterdam).
Goncharov, V. V., and Kurtepov, V. M. (**1994**). ''Formation and propagation of weakly diverging bundles of rays in a horizontally inhomogeneous ocean,'' Acoust. Phys. **40**(5), 685–692.
Halmos, P. R. (**1950**). *Measure Theory* (Springer-Verlag, New York).
Jensen, F. B., Kuperman, W. A., Porter, M. B., and Schmidt, H. (**2000**). *Computation Ocean* Acoustics, AIP Series in Modern Acoustics and Signal Processing (Springer, New York).
Latora, V., and Baranger, M. (**1999**). ''Kolmogorov–Sinai entropy rate versus physical entropy,'' Phys. Rev. Lett. **82**(3), 520–523.
Munk, W. H. (**1976**). ''Sound channel in an exponentially stratified ocean, with application to SOFAR,'' J. Acoust. Soc. Am. **55**(2), 220–226.
Petukhov, Y. V. (**1994**). ''A sound beam with minimal wave front divergence in a stratified ocean waveguide,'' Acoust. Phys. **40**(1), 97–105.
Shannon, C. E. (**1948**). ''A mathematical theory of communication,'' Bell Syst. Tech. J. **27**, 379–423.
Tappert, F. (**2003**). ''Theory of explosive beam spreading due to ray chaos,'' J. Acoust. Soc. Am. **114**(5), 2775–2781.
Topinka, M. A., and Westervelt, R. M. (**2003**). ''Imaging electron flow,'' Phys. Today **56**(12), 47–52.
Van-Trees, H. (**1968**). *Detection, Estimation, and Modulation Theory* (Wiley, New York).
Wolfson, M. A., and Tappert, F. D. (**2000**). ''Study of horizontal multipaths and ray chaos due to ocean mesoscale structure,'' J. Acoust. Soc. Am. **107**, 154–162.
Wolfson, M. A., and Tomsovic, S. (**2001**). ''On the stability of long-range sound propagation through a structured ocean,'' J. Acoust. Soc. Am. **109**, 2693–2703.

The effect of bottom interaction on transmissions from the North Pacific Acoustic Laboratory Kauai source

Michael D. Vera[a)]
Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093-0225

Kevin D. Heaney
OASIS Incorporated, Falls Church, Virginia 22044

The NPAL Group (J. A. Colosi, B. D. Cornuelle, B. D. Dushaw, M. A. Dzieciuch, B. M. Howe, J. A. Mercer, W. H. Munk, R. C. Spindel, and P. F. Worcester)[b)]

(Received 23 March 2004; revised 19 July 2004; accepted 20 July 2004)

Acoustic signals transmitted from a 75-Hz broadband source near Kauai as part of the North Pacific Acoustic Laboratory (NPAL) experiment were recorded on an array of receivers near California at a range of 3890 km, and on a vertical line array at a range of 3336 km in the Gulf of Alaska. Because the source is approximately 2 m above the seafloor, and the bottom depth at the receivers near California is approximately 1800 m, acoustic interaction with the bathymetry complicates the identification of the recorded arrivals with those present in numerical simulations of the experiment. Ray methods were used to categorize acoustic energy according to interactions with the sea bottom and surface and to examine the significance of seafloor geometry. A modal decomposition was also used to examine the role of range-dependent bathymetry and to associate the effects on the acoustic field with seafloor features at specific ranges. Parabolic-equation simulations were performed in order to investigate the sensitivity of the received signal to geoacoustic parameters; shear excitations within the seafloor were modeled using a complex-density, equivalent-fluid technique. Incorporation of bottom interaction into models of the propagation enables an identification between experimental and simulated arrivals. [DOI: 10.1121/1.1854491]

PACS numbers: 43.30.Zk, 43.30.Pc, 43.30.Qd, 43.30.Ma [AIT] Pages: 1624–1634

I. INTRODUCTION

The North Pacific Acoustic Laboratory (NPAL) experiment involved a network of acoustic sources and receivers with propagation ranges of a few megameters. Among these were a 75-Hz source north of Kauai, Hawaii, an array of receivers near the coast of central California at a range of 3889.8 km, and a vertical line array in the Gulf of Alaska at a range of 3336.2 km. One of the goals of the experiment was to investigate the state of the ocean—notably, its temperature structure—by means of acoustic tomography.

Ocean acoustic tomography was proposed by Munk and Wunsch as a means for examining the ocean interior.[1] Oceanographic features influence acoustic propagation through the dependence of sound speed on temperature and salinity. A series of experiments in the Pacific Ocean with ranges on the order of 1000 km demonstrated that long-range transmissions were detectable and could be used to characterize the water column. These include the Slice89 experiment,[2] the Heard Island Feasibility Test,[3] the Acoustic Thermometry of Ocean Climate (ATOC) project and its associated Acoustic Engineering Test (AET),[4,5] and the North Pacific Acoustic Laboratory (NPAL) experiment.[6] Often in long-range experiments with low acoustic frequencies, the propagation can be accurately modeled without a detailed treatment of the acoustic interaction with the seafloor. For megameter-scale experiments, the bottom can typically be modeled very simply if it is nearly flat and much deeper than the sound-channel axis. An approximately flat bottom does not greatly alter the inclination of the propagating sound energy so that bottom interactions (and the associated loss) are repeated. For a deep bottom, the sound intensity is low at the bottom interface due to the refractive effect of the ocean sound channel. The Slice89 experiment, for example, has been successfully simulated using an analytic, absorbing bottom boundary condition.[7] A more sophisticated treatment of the seafloor has been necessary to successfully simulate some long-range acoustic experiments. Propagation from a bottom-mounted source near Oahu, Hawaii to receivers near California, for example, was modeled using a fluid bottom with depth-dependent acoustic properties.[8] In this case, all of the acoustic energy that contributed to the long-range field interacted with the bottom. Because the depth at the source was more shallow than the sound-speed minimum, the sound traveled into deep water along the bottom interface.

Transmissions from the NPAL Kauai source are unusual in that purely refracted and bottom-interacting energy both contribute to the far-field arrival structure as a result of the seafloor geometry. The process of tomographic inversion requires that arrivals present in numerical simulations of the transmission be identified with those recorded in the data.

[a)]Electronic mail: michaelvera@earthlink.net; Currently at The University of Southern Mississippi, Long Beach, MS 39560.
[b)]J. A. Colosi is at Woods Hole Oceanographic Institution, Woods Hole, MA 02543. B. D. Cornuelle, M. A. Dzieciuch, W. H. Munk, and P. F. Worcester are at Scripps Institution of Oceanography, University of California at San Diego, La Jolla, CA 92093. B. D. Dushaw, B. M. Howe, J. A. Mercer, and R. C. Spindel are at the Applied Physics Laboratory, University of Washington, Seattle, WA 98105.

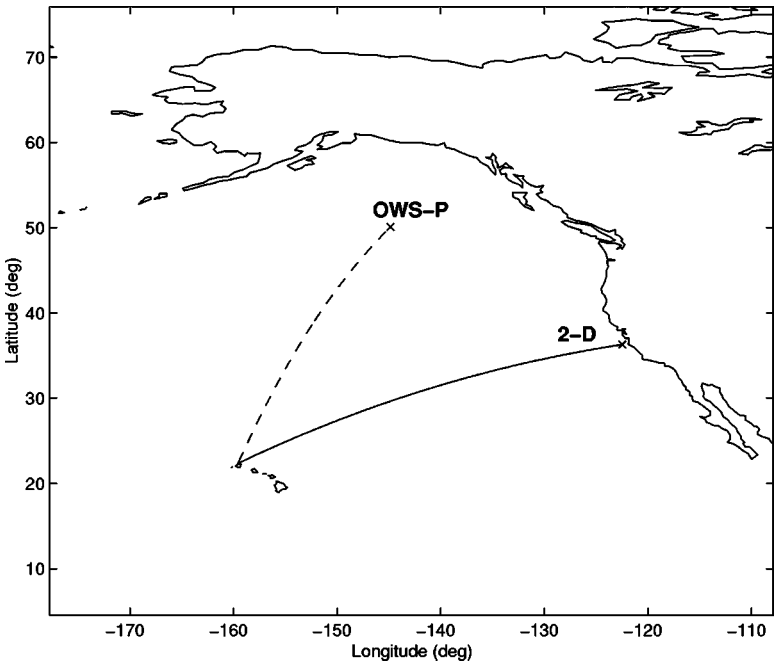


FIG. 1. The propagation path from the Kauai source to the 2D array near central California is shown as a solid line. The path from the source to OWS-P is shown as a dashed line.

The identification between simulated and measured arrivals is made more complicated by the simultaneous presence of both kinds of sound energy. Initial attempts at identification using only fully refracted simulated ray paths could not account for the entire arrival pattern. These refracted rays were also offset in travel time relative to the data.

This long-range, low-frequency experiment was not intended as an investigation into the geoacoustic properties of the seafloor along the propagation paths. Nevertheless, interaction with the bathymetry has clearly affected the acoustic field from the NPAL Kauai source. Perhaps the most obvious effect is the presence of very steep (greater than 15 deg from the horizontal) acoustic arrivals at the receivers near California. They can be explained by reflection from the seafloor in the vicinity of the receiver array. Recorded arrivals such as these cannot be replicated in simulations without a treatment of acoustic interaction with the bathymetry. A treatment of the acoustic interaction with the seafloor, therefore, can aid in tomography by protecting against misidentification of receptions generated by bathymetric interaction with fully refracted sound in simulations. Models that incorporate bottom interaction still give rise to a net travel-time difference from the data, though the arrival pattern can be more successfully replicated and the identification is much less ambiguous. Other factors must be considered to explain this offset; one candidate is an inaccuracy of the Del Grosso sound-speed equation at high pressures.

II. NORTH PACIFIC ACOUSTIC LABORATORY EXPERIMENT

Two broadband ($Q=2$) acoustic sources with a center frequency of 75 Hz were employed in the Acoustic Thermometry of Ocean Climate (ATOC) and NPAL experiments; the one considered in this work was deployed off the shore of Kauai, Hawaii. Receptions were recorded on several vertical and horizontal receiving arrays. This paper involves a vertical line array (VLA) of receivers at Ocean Weather Station-P (OWS-P) in the Gulf of Alaska and a 2D receiver array near the California coast. This 2D array consisted of five vertical line arrays oriented transverse to the propagation path; it was designed to examine the coherence of receptions with horizontal displacements of a few hundred meters to a few kilometers. Both of the paths are shown in Fig. 1.[6]

One of the propagation paths considered in this work has a range of 3889.8 km from the Kauai source to the NPAL 2D array. In the vicinity of this path, SEABEAM bathymetry measurements were performed with a horizontal resolution of approximately 50 m. The measurements taken near the source are displayed in Fig. 2. Portions of the measured bathymetry along the transect are displayed in Fig. 3. Significant acoustic interaction with the bottom could be expected at these locations. The first displayed region is near the source. At a range of roughly 2000 km, a seamount was detected that reached as high as 2500 m below the surface. Finally, the bathymetry near the receiver array slopes upward to depths of about 1800 m. The path from the Kauai source to OWS-P has similar bathymetry near the source, but lacks any significant seamounts and the upward slope near the receiver. The near-source bathymetry for this path is also displayed in Fig. 3. The same bathymetry measurements as for the first path are available for the first 4 km of this transect. At longer ranges, the Smith/Sandwell bathymetry database was used.[9]

Two extensive series of water column measurements were also performed during the summers of 1998 and 1999 along the transect from the source to the 2D array. High-resolution CTD/XBT surveys were used in the construction of objective maps of salinity and temperature. The profiles at adjacent ranges from an objective map with a horizontal seperation of 39 km were averaged in order to suppress small-scale features such as internal-wave effects. Sound-speed fields were derived using these data and the Del Grosso sound-speed equation.[10,11] The sound speeds derived from the 1999 measurements are displayed in Fig. 4. The surface

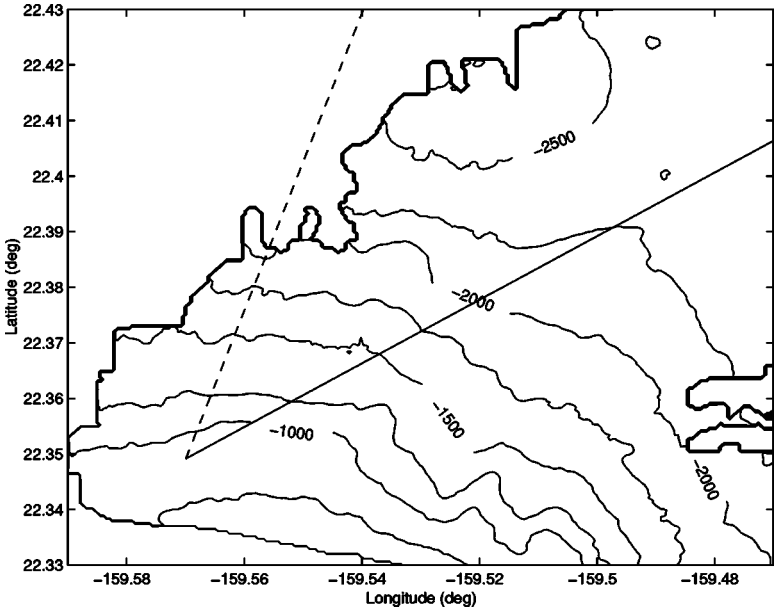


FIG. 2. The bathymetry measurements are shown in the vicinity of the Kauai source. The path to the 2D array is shown as a solid line. The path to OWS-P is shown as a dashed line. For the latter, bathymetry measurements are available for the first 4 km of the path; thereafter, the Smith/Sandwell database was used.

waters are cooler near the California coast than near Hawaii; the sound speeds near the surface decrease with range. Annual averages from the World Ocean Atlas were used to generate sound speeds along the path to the VLA at OWS-P.[12]

III. SIMULATION TECHNIQUES

A. Rays

The ocean depth at the location of the bottom-mounted Kauai source is approximately 800 m. The acoustic center of the source is less than 2 m above the seafloor.[13] Therefore, a significant amount of the energy that would contribute to the long-range refracted field for a source located in a region of abyssal depth instead interacts with the bottom. Ray methods can be used to represent the trajectories of sound propagating within the range–depth plane connecting the source and a receiver. The ray simulations discussed below allow for specular reflections from the bottom, up to a maximum of ten, as well as from the surface. Sound speeds were interpolated according to the requirements of the adaptive step-size, fourth-order Runge–Kutta ray integration. The 12 000 initial

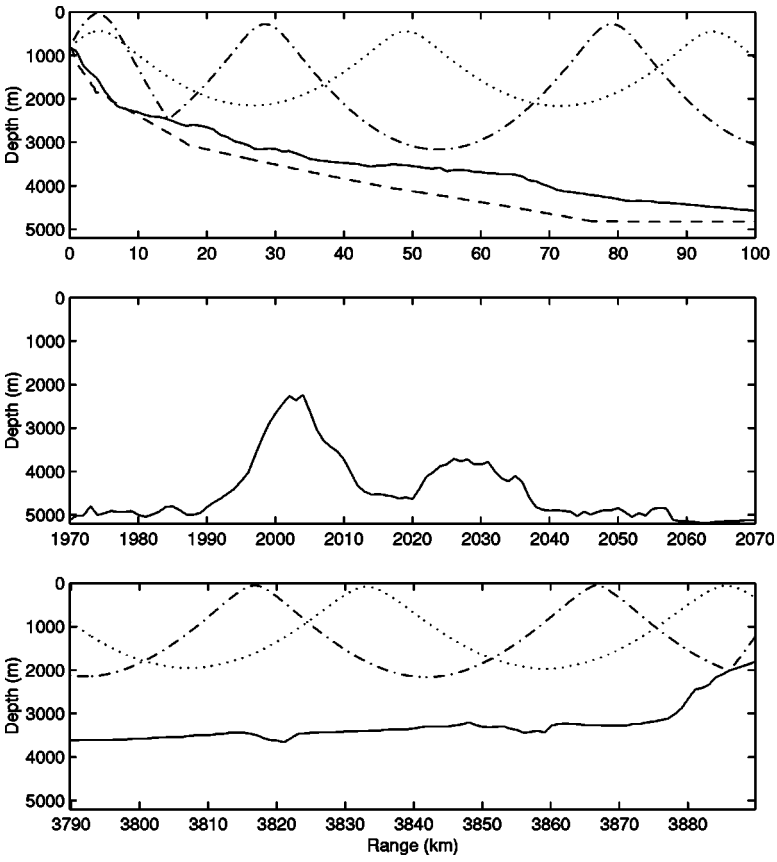


FIG. 3. The bathymetry along the propagation path from Kauai to the 2D array is shown as a solid line. The top panel is the source region; the bathymetry along the OWS-P transect is shown in this vicinity as a dashed line. Two sample rays are shown: the dotted ray is fully refracted, the dash-dotted ray reflects from the surface, and the bottom in the vicinity of the source. The center panel displays the seamount at approximately half the range along the path to the 2D array. The bottom panel is the bathymetry near the 2D receiver array. The same refracted ray as in the top panel is included; this dash-dotted ray only reflects from the bottom near the receiver.

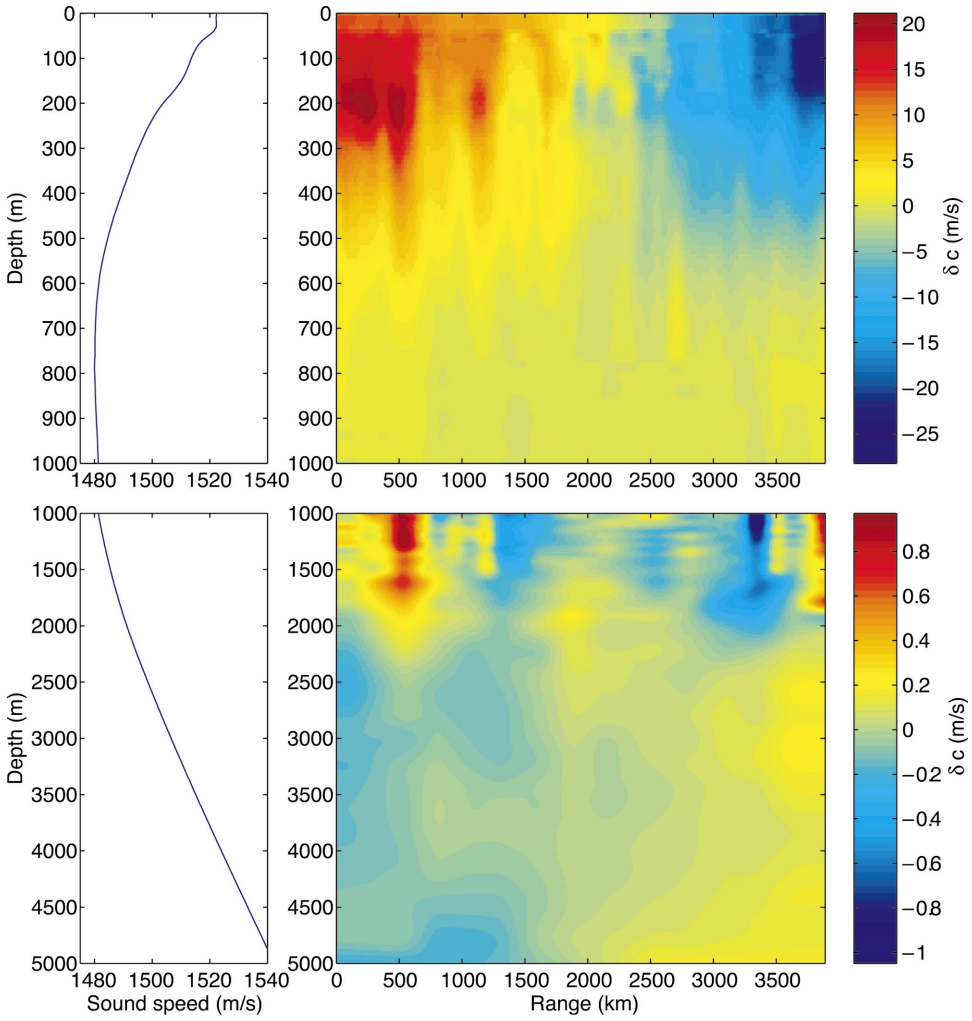


FIG. 4. Sound speeds derived from environmental measurements taken during the summer of 1999 along the 2D array transect. The left panels contain the range-averaged sound speed as a function of depth. The panels on the right display the difference from the range average as a function of both depth and range. The sound speed is less variable at great depth, as can be seen in the color scales of the right panels.

launch angles spanned vertical inclinations from −30 to +30 deg relative to horizontal. Rays were discarded if they reflected from the bottom at a range shorter than the nominal wavelength of 20 m.

Rays in these simulations can be divided into groups according to their interactions with the bottom and surface. Near the source, the bottom slopes downward from 800 m to abyssal depths for ranges beyond $O(100\,\mathrm{km})$. The bottom slope is approximately 5 deg in the immediate vicinity of the source. Since specular reflection from a downslope changes the propagation angle by twice the grazing angle, relatively steep downward-propagating rays can be scattered into shallower paths and thereafter propagate as refracted rays. The ray with a launch angle of −10 deg, for example, undergoes a single bottom interaction and then propagates nearly horizontally. Other rays, typically steep positive launches, reflect between the surface and bottom until the repeated bottom interactions decrease the ray inclinations sufficiently for them to propagate as refracted paths. These rays can leave the source region with steeper angles than are accessible to the other ray types. Finally, there exist some rays that interact with neither the surface or the bottom; it is these rays that are typically used in tomographic inversions.

Some of the paths that reflect from the bottom near the source but not the surface differ only slightly from paths that are entirely water borne. At a given range, the arrival times and depths of these paths closely mirror the purely refracted rays (Fig. 5). Those rays that interact with both the surface and the bottom, however, yield timefront segments that evolve differently in range. At relatively short ranges, such as the 100-km example in Fig. 5, these rays are later than the others due to the extra path length resulting from traversing the entire vertical extent of the water column multiple times in a relatively short range. At longer ranges, however, the effect of the extra distance on the travel time is overcome by the sampling of higher-speed water that is caused by the greater inclination. Then, these rays give rise to the earliest arrivals.

As the acoustic field approaches the 2D receiver, the depth again decreases from abyssal values. The water depth at the NPAL 2D array is approximately 1800 m. A subset of the incoming rays interacts with this upslope, is reflected into steeper angles—in some cases, angles too steep for significant long-range propagation—and then received. These near-receiver bottom-interacting rays, unlike those that only interact with the bottom near the source, are easily distinguished from the refracted paths by this increase in inclination. The bathymetry along the path to the VLA at OWS-P remains relatively deep (greater than 4000 m) as the receiver is approached. Therefore, only the near-source bathymetry plays a prominent role.

Turning-point filter representations of the ray arrivals at

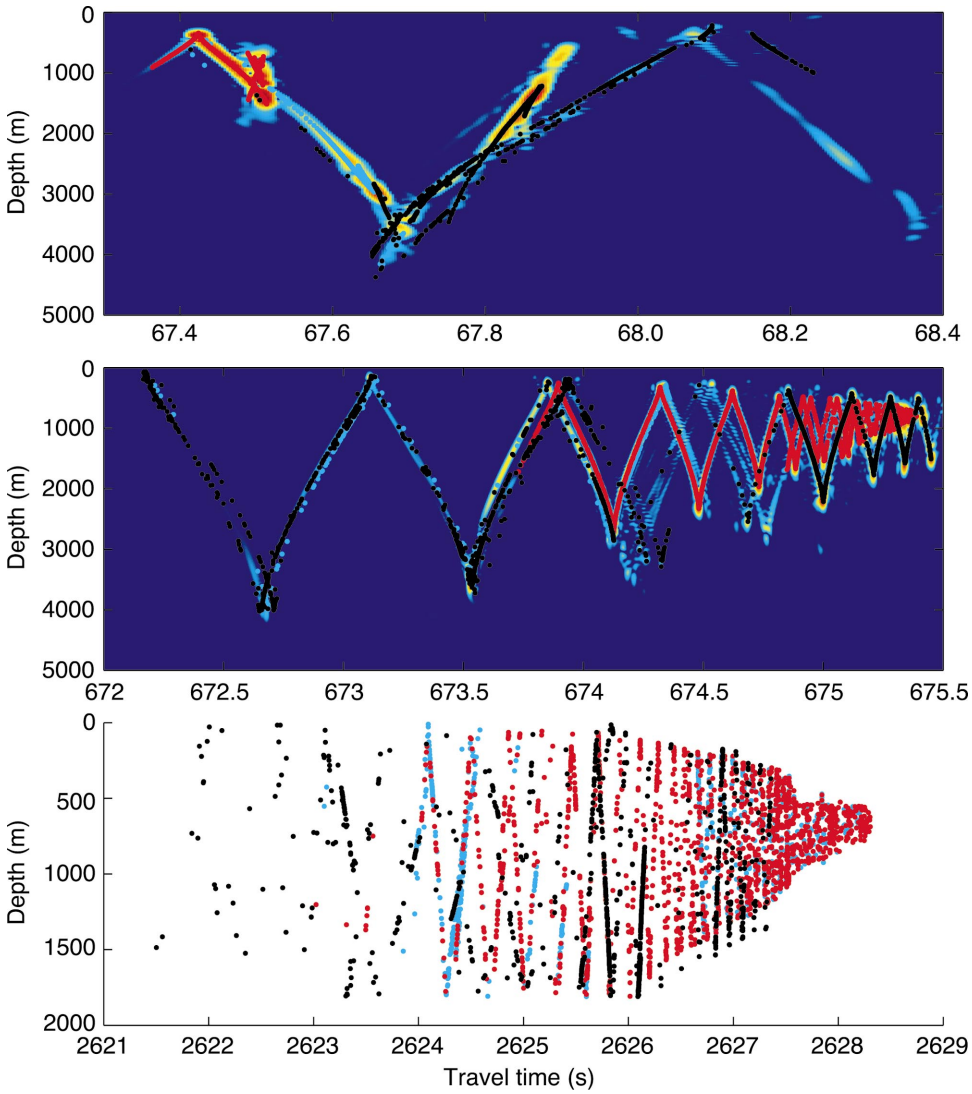


FIG. 5. Timefronts at ranges of 100, 1000, and 3889.8 km along the 2D array transect. Results from PE simulations using bottom type 1 are shown at ranges of 100 and 1000 km with a 30-dB dynamic range. Ray arrivals are indicated at all three ranges. Fully refracted trajectories are shown in red. Paths that reflect only from the bottom are marked in blue. Paths that bounce from both the bottom and the ocean surface are shown in black. At short ranges, the paths that reflect from both the bottom and the surface are later than the refracted rays. At longer ranges, however, this lag decreases since the reflected rays can attain higher angles from the horizontal and, thus, sample water with a higher sound speed.

the OWS-P VLA are displayed in Fig. 6.[14] The arrivals at the 2D array are displayed in Fig. 7. The refracted paths yield the usual arrival pattern. The rays that have reflected only from the bottom can be organized into two distinct groups. The first group is virtually indistinguishable from the refracted rays; these have interacted with the bottom in the immediate vicinity [within $O(1\ \mathrm{km})$] of the source. The second set, found only in the simulations of propagation to the

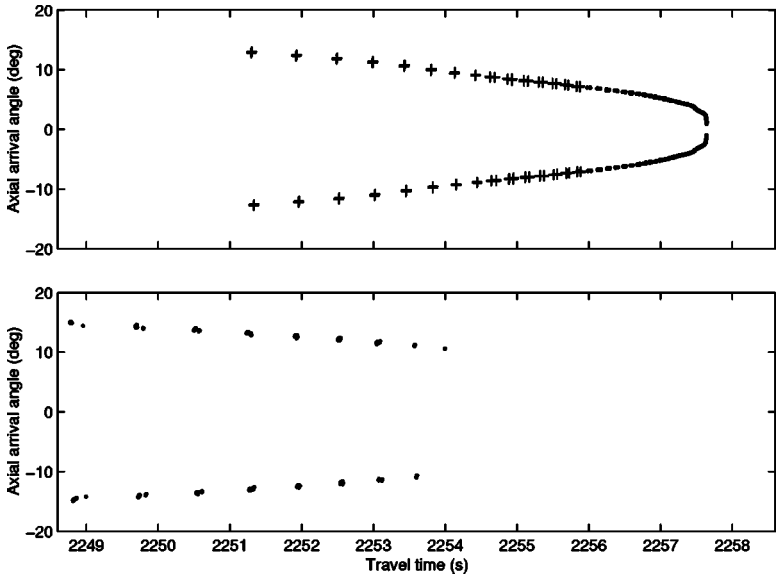


FIG. 6. The application of a turning-point filter yields axial travel times and arrival angles for the rays of the OWS-P propagation path. The top panel shows the fully refracted ray arrivals as dots and those that reflect only from the surface as crosses. The bottom panel displays the rays that reflect from both the surface and the bottom (with 2 or fewer bottom bounces); these are the earliest arrivals.

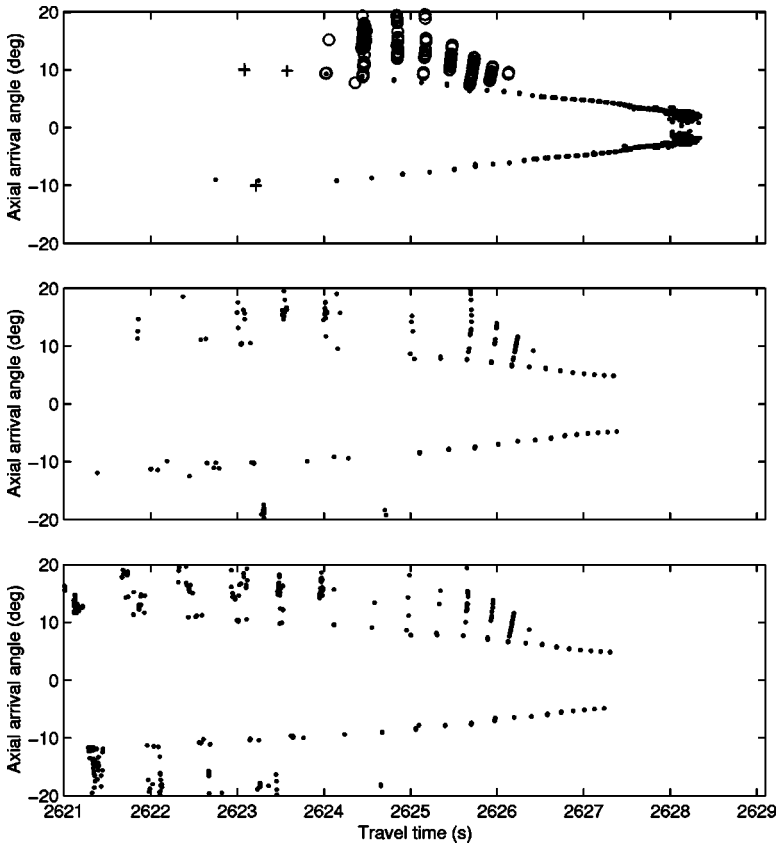


FIG. 7. The top two panels are the same as in Fig. 6 except for the propagation path to the 2D array. An additional category of ray paths is included in the top panel; these reflect only from the bottom near the receiver and are marked as circles. The bottom panel displays the rays that have reflected from both the surface and the bottom in simulations with the seamount removed; in this case, rays with up to 5 bottom bounces are included.

2D array, arrive at much steeper angles and have reflected from the bottom near the receiver.

The distinct range evolution undergone by the rays that interact with both the surface and the bottom affects the pattern at the full range. Some of these rays yield distinct points in the time–angle plane similar to those in the refracted pattern. Relative to refracted rays near the same travel times, these arrivals are slightly steeper. Rays of this type also yield the earliest arrivals. The ray trajectories along the transect to OWS-P are affected only by the near-source bathymetry.

Ray interaction with the bathymetry along the 2D array transect can also occur at approximately half the full range where a seamount reaching up to depths of approximately 2500 m was detected. The small horizontal extent of this feature suggests that a two-dimensional ray simulation could exaggerate its importance as an acoustic blockage. That the earliest arrivals have interacted with both the surface and the bottom is more clearly illustrated in ray simulations performed with this seamount feature removed (see Fig. 3). The measured bathymetry along the transect to the 2D array is replaced in these simulations with a straight line in the range interval from 1900 to 2100 km.

B. Normal modes

The propagation from the Kauai source to the NPAL 2D array can also be examined within the mathematical framework of acoustic normal modes. Acoustic normal modes are eigenfunctions of the single-frequency, depth-separated wave equation. Since modes can be characterized by turning points, an examination of the propagation using modal methods allows for some of the geometrical insight provided by ray techniques in a way that accounts for the finite frequency of the field. Although they are an exact solution only for range-independent environments, modes provide useful tools for evaluating the severity of the range dependence and its effect on the long-range signal. For mildly range-dependent environments, the energy distribution of modes is constant and the propagation is referred to as adiabatic mode propagation. The bathymetric changes along the path from the Kauai source to the NPAL 2D array induce mode coupling, a redistribution of energy amongst the modes. By expressing the effect of the seafloor in terms of mode coupling, it can be compared to the effect of the range dependence in the sound-speed field used in the model (see Sec. II), and to the effect in other studies of oceanic fronts[15] or of smaller-scale range dependence induced by internal waves.[16]

To examine these issues, a simplified, single-frequency (75 Hz) parabolic-equation (PE) computation was performed. At regular range intervals, the local acoustic normal modes were computed. The mode spectrum was derived from the PE field solution using the orthonormality of the normal-mode wave functions

$$\int_z \frac{\psi_m(z,r)\psi_n(z,r)}{\rho(z,r)}\,dz = \delta_{mn}, \tag{1}$$

$$b_n(r) = \int_z \frac{\psi_m(z,r)P(z,r)}{\rho(z,r)}\,dz, \tag{2}$$

where ψ_m is the mth mode function, δ_{mn} is the Kronecker delta function (1 only when $m=n$ and 0 otherwise), $P(z,r)$ is the single-frequency parabolic-equation solution, and b_n is the nth mode amplitude at a particular range. For mildly

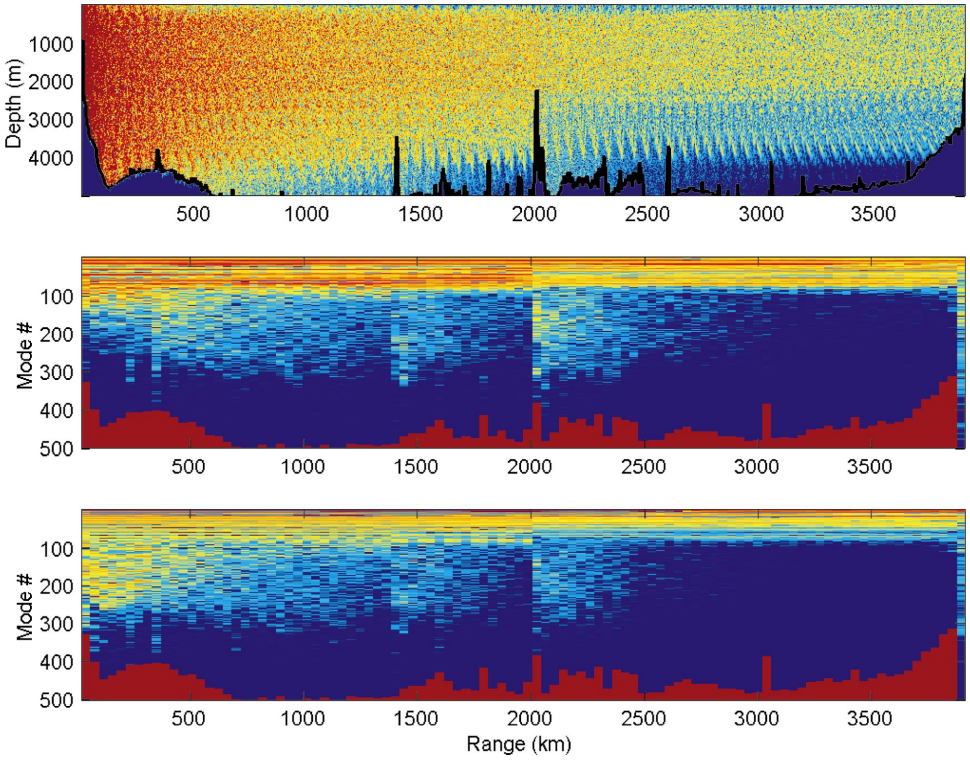


FIG. 8. The modal decomposition of the single-frequency acoustic field illustrates the ranges for which range dependence induces significant coupling. The top panel (a) displays the 75-Hz acoustic field between the Kauai source and the 2D array. The middle panel (b) shows the modal decomposition of this field. The bottom panel (c) is similar, but with the near-source bathymetry removed.

range-dependent environments, b_n is expected to be independent of range (except for attenuation and cylindrical spreading).

The single-frequency PE field is shown in the top panel of Fig. 8. The bathymetry along the path has been superimposed on the field. The dominant bathymetric interactions occur at the source, at the seamount near 2000 km and at the receiver. The cyclical interference pattern in range and depth is due to convergence zones (roughly 50 km apart) where acoustic paths (rays) of different angles converge. The field plot indicates that the seamount strips energy from the water column.

The mode spectrum (in dB) from Eq. (2) is shown in Fig. 8(b) after correcting for cylindrical spreading ($+10\log R$). For comparison with the ray view of propagation, the WKBJ approximation is used and mode phase speeds determine equivalent ray turning points. The low modes (1–60) are trapped (refracted–refracted paths). Modes 60–100 are surface reflecting and bottom refracting (depending on the depth). Modes greater than 100 are surface reflecting and bottom reflecting (SRBR) regardless of depth. For the lowest order modes ($N<20$), the propagation path appears to be adiabatic; the mode amplitude is constant in range. This verifies that the range dependence in the measured sound-speed profile (after smoothing in the horizontal) is indeed mild enough to be considered adiabatic. All of the ranges at which significant mode coupling occurs correspond to regions of bathymetric interaction. The seamount at 2000 km induces substantial mode coupling. Very low-order modes appear unaffected (these modes, or shallow rays, pass over the seamount). The intermediate modes (50–150) are stripped of energy. The high modes (>150) are excited as the sound radiates from the backside of the seamount. There is additional mode coupling at the receiver, as the seafloor rises up the continental shelf approaching the receiver array.

Mode scattering is also evident at a range of approximately 1400 km. There is a seamount at this location which has a minimum depth of 3500 m. This illustrates an interesting point about propagation in deep water. It appears that the acoustic field can be sensitive to variations in the seafloor topography that are well (2500 m) below the sound-channel axis. High mode numbers, or equivalently high-angle rays, have very deep turning points and can be scattered by changes in the ocean depth. This can be significant because it is precisely these high-angle rays/high modes that are stable, identifiable, and therefore useful in ocean acoustic tomography.

As an additional examination into the significance of near-source bathymetric interaction on the long-range acoustic transmission, the same calculation was performed with the bottom depth set to 4000 m within 300 km of the source. The resulting mode decomposition for the ''deep-water source'' is shown in the lower panel of Fig. 8. The propagation over the path is similar to that of the middle panel, with the exception of the excitation spectra near the source. The presence of the seafloor leads to significantly more energy in the mode numbers from 40–100. This is precisely the energy that has scattered off the slope near Kauai and has entered into deep water in modes (angles) that are no longer bottom interacting.

C. Parabolic-equation simulations

The role of bathymetry in influencing acoustic propagation in the NPAL experiment was also examined by performing broadband simulations using a parabolic equation. Since the bottom is solid, acoustic energy can be lost to conversion into elastic shear waves, as well as compressional sound

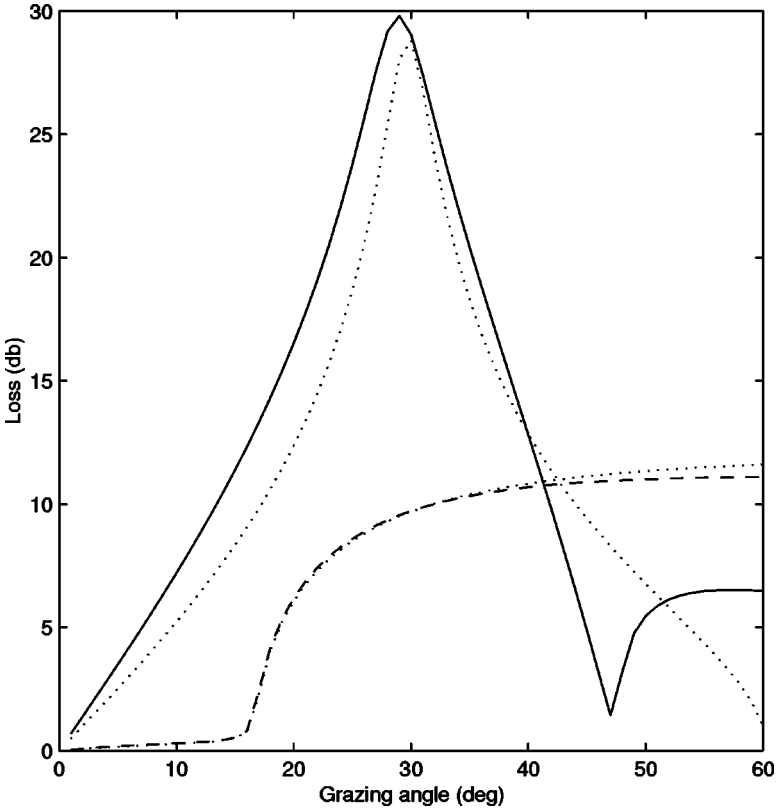


FIG. 9. Bottom loss $[-10\log|V^2|$; see Eq. (3)] versus grazing angle for the bottom types considered in this work. The solid curve follows from the elastic reflection coefficient for a compressional speed of 2200 m/s, a shear speed of 1100 m/s, and a density 2.1 times that of the water (bottom type 1). The corresponding parameters for the dashed curve are: 1550 m/s, 200 m/s, and 1.7 (bottom type 2). The dotted curves are given by the complex-density, equivalent-fluid parameters in each case.

waves, in the bottom. Because the acoustic field in the water was the focus of the simulation effort, an equivalent-fluid technique was used to model the combined effects of the shear and compressional parameters of seafloor material. A technique presented by Zhang and Tindle[17] involves the calculation of an effective density, a complex number, that models the important effects of shear in a simulation that involves only fluid media. The complex density is chosen in order to approximate the reflection coefficient of an elastic bottom. The plane-wave reflection coefficient, $V(k)$, for an acoustic wave in the water incident on an elastic seafloor can be written as

$$V(k)=\frac{\gamma_1[\rho_2 P(k)/\rho_1]-i\eta_2}{\gamma_1[\rho_2 P(k)/\rho_1]+i\eta_2}, \tag{3}$$

where k is the horizontal wave number and γ_1 and η_2 are the vertical wave numbers for compressional waves in the water (medium 1) and the bottom (medium 2). The effects of shear are contained in the factor, $P(k)$, given by

$$P(k)=\left[1-\frac{2k^2}{(\omega/c_s+i\alpha_s)^2}\right]^2+\frac{i4\eta_2\gamma_s k^2}{(\omega/c_s+i\alpha_s)^4}, \tag{4}$$

where ω is the angular frequency, c_s is the speed of shear waves, γ_s is the shear vertical wave number, and α_s is the shear attenuation. The complex density presented by Zhang and Tindle corresponds to using a complex number to represent the product $\rho_2 P(k)$ at low grazing angles. The equivalent fluids used in the simulations discussed below were generated by a search in which compressional speed, and the real and imaginary parts of the density are treated as free parameters. The best match to the reflection coefficient of the elastic bottom was chosen for the grazing angle interval from 0 to 45 deg. This method is intended to extend the validity of the technique to higher shear speeds and larger grazing angles. The bottom losses calculated using elastic media are compared to this complex-density equivalent fluid in Fig. 9. For sufficiently large grazing angles, this technique breaks down for some media. This method must also fail in circumstances where interface waves or the propagation in the bottom media are significant. None of these limitations appears directly relevant to simulations of the long-range NPAL experiment.

The RAM code, developed by Collins,[18] was used in simulations with two sets of seafloor geoacoustic parameters. A set of 901 discrete frequencies over the 75-Hz total bandwidth was propagated forward using a computational grid with 0.25-m vertical spacing and 10-m range steps. Four terms were included in the Padé expansion. This grid is based upon the convergence of single-frequency transmission loss at the center frequency of 75 Hz; a travel-time comparison with the ray results was successful. Sound speed in the water was interpolated at 300 equally spaced ranges (for a separation of approximately 13 km along the path to the 2D array). The bottom was treated as homogeneous in all cases, except for an absorbing layer at depths greater than 5750 m. The first medium (bottom type 1) was intended to approximate the porous basalt of the Hawaiian islands. Seismological investigations indicate that 2200 m/s is a reasonable estimate of the compressional speed in the solid near the water/bottom interface.[19,20] Empirical relations stated by Hamilton then suggest a shear velocity of approximately 1100 m/s (Ref. 21) and a density 2.1 times that of water (ρ_w).[22] The second parameter set is a characterization of a loose sediment (bottom type 2); the compressional speed is

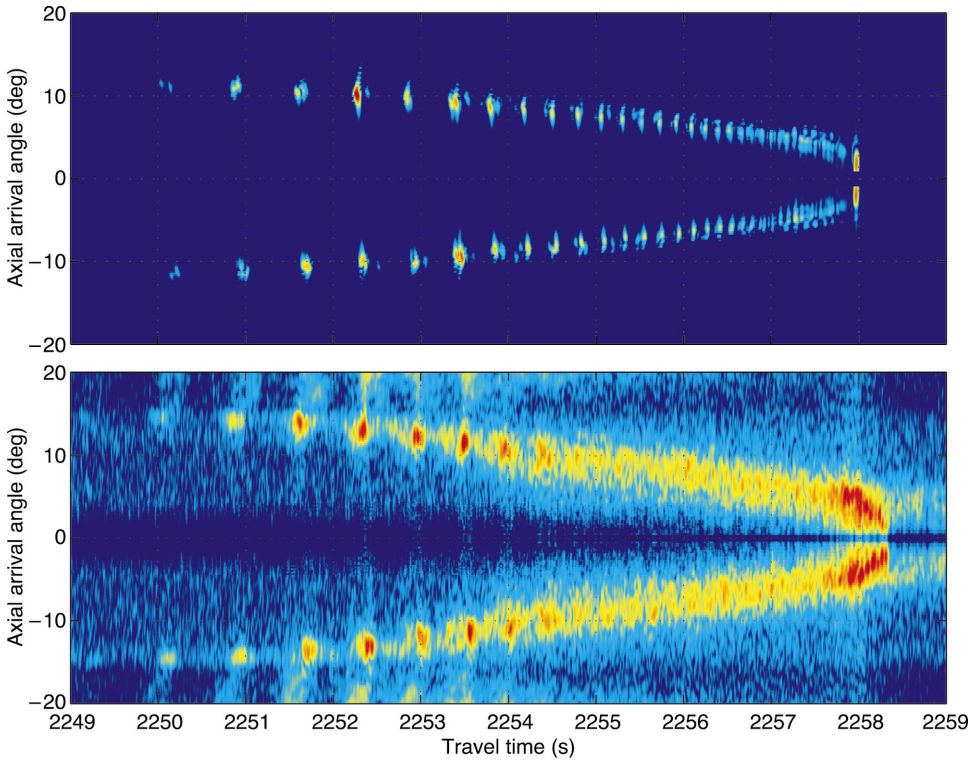


FIG. 10. Turning-point filter results from the parabolic-equation simulations (top) and the data (bottom) are shown for the path to OWS-P with a 30-dB dynamic range. The simulation depicted in the top panel used bottom type 1 and has been offset by 0.4 s.

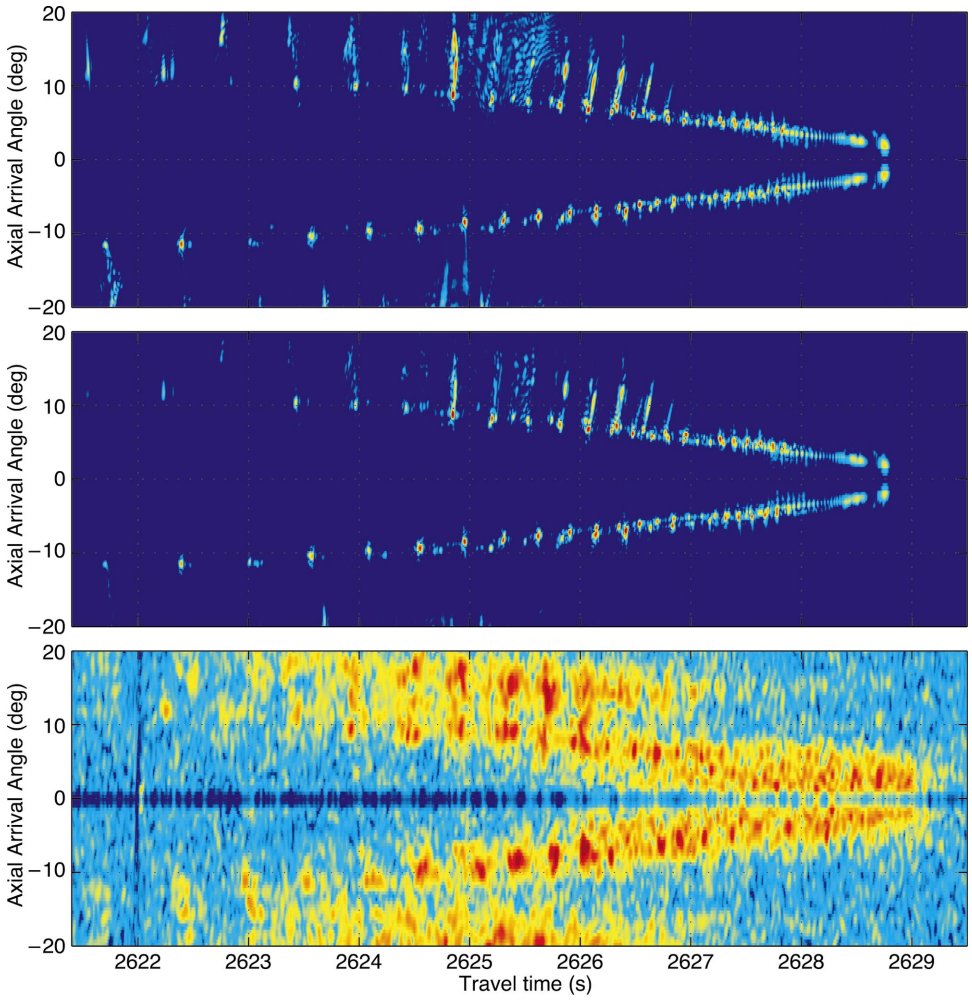


FIG. 11. Turning-point filter results from the parabolic-equation simulations (top and middle) and the data (bottom) are shown for the path to the 2D array with a 30-dB dynamic range. The simulation depicted in the middle (top) panel used bottom type 1 (2) and has been offset by 0.4 s.

1550 m/s, the shear speed is 200 m/s, and the density is $1.7\rho_w$.[21] For both bottom types, the shear attenuation was set to 0.2 dB/λ and the compressional attenuation was 0.1 dB/λ. These bottom types were chosen in an attempt to include an estimate of the actual geoacoustics near the Kauai source (type 1) and a limiting case that was softer and looser (type 2) than any seafloor parameters that were likely to be encountered over the entire propagation range of the experiment.

IV. COMPARISON

Turning-point filter representations of the acoustic field at the full range are displayed in Fig. 10 for the OWS-P VLA and in Fig. 11 for the 2D array. The data in each case are shown in the bottom panel and consist of a 1-day average. For the field at OWS-P, the daily average refers to 1998 yearday 273 and includes six transmissions. The 2D array data shown were collected on 1998 yearday 544 at the central VLA receiver (VLA 3) and include four transmissions. Coherent averaging is performed during the transmission period of approximately 20 min, and different transmissions are averaged incoherently. The turning-point filter transformation is performed using the entire 20-element array.[6] Corresponding simulations are included in Figs. 10 (using bottom type 1) and 11 (using both bottom media). A travel-time offset of 0.4 s yields an unambiguous identification between the early-time arrivals in the PE simulations and in the data.

These simulations generate the same basic arrival structure seen in the ray analysis. The ray arrivals in Figs. 6 and 7 can also be identified with the data with an overall travel-time offset of 0.4 s. This identification illustrates that the early-time arrivals represent energy that has interacted with both the ocean surface and seafloor. These early arrivals are especially useful in performing tomography because they are relatively stable and distinct. The recognition that this energy has reflected from both the sea surface and the bottom in the NPAL geometry has been a useful step in the inversion of the received acoustic field for ocean properties.

Initial efforts to estimate the overall time offset between the data from the 2D array and simulations using only fully refracted ray paths resulted in substantial nonzero values. Incorporation of bottom interaction into the models helps to protect against misidentification of reflected arrivals in the data with refracted arrivals in simulations. It also illustrates an additional complication; refracted and bottom-interacting acoustic energy can both arrive at similar travel times. Such problems in the identification process, however, do not appear to have been the sole cause of the offset; travel times from the expanded simulations still differ from those in the data. The simulated arrivals are earlier than the data by about 0.4 s. The simulations presented here used sound speeds derived from environmental measurements taken during 1999. Travel times from the acoustic data change rapidly during this period. If, instead, the 1998 environmental measurements are used, simulated travel times are offset by approximately 0.2 s. This offset is not sensitive to bottom parameters within the limits of a complex-density equivalent-fluid representation of the solid elastic media. An attempt to accurately characterize other features of the received acoustic field, such as the intensity of the reflected arrivals, may depend on the use of realistic geoacoustic parameters in numerical simulations.

Comparisons between simulations and data in other long-range experiments have also yielded an offset. Some authors have suggested that these offsets can be attributed to inaccuracy of the Del Grosso sound-speed equation at high pressures. The analysis presented by Worcester *et al.* of the Acoustic Engineering Test (AET), part of the Acoustic Thermometry of Ocean Climate (ATOC) project, indicated that the simulated arrivals differed from the measurements by approximately 0.2 s for a range of 3250 km.[23] Spiesberger and Metzger presented simulated travel times based on the Del Grosso equation that differed from observations of a 3-Mm experiment in the northeast Pacific by about 0.6 s. They derived an additive correction, a function of pressure, to the Del Grosso equation.[24] Despite extensive environmental information, simulations of the NPAL experiment are similarly offset from the data.

Incorporation of a model for bathymetric interaction into the simulations has improved the correspondence between the recorded and simulated arrivals sufficiently to permit an identification. There are, however, important limitations of this analysis. The homogeneous representation of the seafloor is sufficient for the purposes of an identification in travel time, but it is obviously unrealistic and probably unsuitable for analyzing other acoustic characteristics such as intensity. The complex-density method discussed in Sec. III C provides a computationally efficient approximation to an elastic bottom of sufficient accuracy for most long-range applications, but data regarding the actual parameters of the elastic solid to be modeled are often unavailable. The changes in the travel time of an acoustic arrival over the course of the experiment have been neglected for the purposes of the identification, even though these changes may provide some insight into the different travel-time offsets from the data for simulations based on 1998 and 1999 environmental measurements. Finally, small-scale perturbations of the sound speed, such as those caused by internal waves, have been neglected. It is not expected that these perturbations will change the travel-time structure so dramatically as to undermine an identification of the earliest arrivals; but rather will reshape the signal finale and induce small-scale fluctuations.[7]

V. CONCLUSION

Interaction with the bathymetry alters the long-range acoustic field generated by the Kauai source. For ranges greater than $O(1000\,\mathrm{km})$, sound that has reflected from both the sea surface and the seafloor in the vicinity of the source arrives at travel times similar to, or earlier than, purely refracted energy. Relative to refracted energy at the same travel time, this reflected energy arrives at a steeper angle. Propagation to the 2D array is also affected by a seamount at a range of approximately 2000 km and by an upslope to the receiving array depth of about 1800 m. Analysis of the propagation to OWS-P illustrates the effect of the near-source interactions without the additional complications that occur downrange for the 2D array transect. Energy that interacts with both the surface and the bottom near the source

constitutes the earliest arrivals in both cases. Sound that has reflected only from the bottom in the vicinity of the source is virtually indistinguishable from purely refracted energy; because it does not yield a distinct arrival, this energy does not greatly influence the process of tomography. Sound that reflects only from the bottom near the 2D array can be easily distinguished from arrivals generated by other types of trajectories and does not lead to difficult identification ambiguities.

The investigation into the effect of bottom interactions on the acoustic propagation from the Kauai source to the NPAL 2D array was intended to address the need for a reliable identification between the recorded acoustic arrivals and the arrivals in numerical simulations. Models that account for the acoustic interaction with the bathymetry yield an arrival pattern that corresponds to the data much more closely than is possible in models that treat only the water column. Ray simulations, because they generate a specific trajectory, make it possible to group arrivals according to the number of reflections from the sea surface and the bottom. A modal analysis indicates the ranges at which interaction with the bathymetry plays an important role. Parabolic-equation simulations allow the geoacoustic parameters of the bottom media to be varied in order to examine the sensitivity of the received field. A reliable determination of the overall best match between simulated arrival patterns and the data can be used as the first step in a detailed tomographic inversion.

ACKNOWLEDGMENTS

This work was supported by the Strategic Environmental Research and Development Program through Defense Advanced Research Projects Agency (DARPA) Grant No. MDA972-93-1-0003, by the Office of Naval Research (ONR) through Grant Nos. N0014-1-97-1-0258 and N00014-03-1-0182, and by the National Oceanographic Partnership Program (NOPP) through University of Washington Subcontract No. 551302.

[1] W. Munk, P. Worcester, and C. Wunsch, *Ocean Acoustic Tomography* (Cambridge University Press, Campridge, 1995).
[2] T. Duda, S. M. Flatté, J. Colosi, B. Cornuelle, J. Hildebrand, W. Hodgkiss, Jr., P. Worcester, B. Howe, J. Mercer, and R. Spindel, ''Measured wavefront fluctuations in 1000-km pulse propagation in the Pacific Ocean,'' J. Acoust. Soc. Am. **92**, 939–955 (1992).
[3] W. Munk, R. Spindel, A. Baggeroer, and T. Birdsall, ''The Heard Island feasibility test,'' J. Acoust. Soc. Am. **96**, 2330–2342 (1994).
[4] J. A. Colosi, E. K. Scheer, S. M. Flatté, B. D. Cornuelle, M. A. Dzieciuch, W. H. Munk, P. F. Worcester, B. M. Howe, J. A. Mercer, R. C. Spindel, K. Metzger, and T. G. Birdsall, ''Comparisons of measured and predicted acoustic fluctuations for a 3250-km propagation experiment in the eastern North Pacific Ocean,'' J. Acoust. Soc. Am. **105**, 3202–3218 (1999).
[5] J. Colosi, F. Tappert, and M. Dzieciuch, ''Further analysis of intensity fluctuations from a 3252-km acoustic propagation experiment in the eastern North Pacific,'' J. Acoust. Soc. Am. **110**, 163–169 (2001).
[6] P. Worcester and R. Spindel, ''North Pacific Acoustic Laboratory,'' J. Acoust. Soc. Am. **117**, 1499–1510 (2005).
[7] J. Colosi, S. M. Flatté, and C. Bracher, ''Internal-wave effects on 1000-km oceanic acoustic pulse propagation: Simulation and comparison with experiment,'' J. Acoust. Soc. Am. **96**, 452–468 (1994).
[8] F. Tappert, J. Spiesberger, and M. Wolfson, ''Study of a novel range-dependent propagation effect with application to the axial injection of signals from the Kaneohe source,'' J. Acoust. Soc. Am. **111**, 757–762 (2002).
[9] W. Smith and D. Sandwell, ''Global sea floor topography from satellite altimetry and ship depth soundings,'' Science **277**, 1956–1962 (1997).
[10] B. Dushaw, P. Worcester, B. Cornuelle, and B. Howe, ''On equations for the speed of sound in seawater,'' J. Acoust. Soc. Am. **93**, 255–275 (1993).
[11] V. D. Grosso, ''New equation for the speed of sound in natural waters (with comparisons to other equations),'' J. Acoust. Soc. Am. **56**, 1084–1091 (1974).
[12] M. Conkright, R. Locarnini, H. Garcia, T. O'Brien, T. Boyer, C. Stephens, and J. Antonov, World Ocean Atlas 2001: Objective analyses, data statistics, and figures, CD-ROM Documentation (National Oceanographic Data Center, 2002).
[13] B. Howe, ''Acoustic Thermometry of Ocean Climate (ATOC): Pioneer Seamount Source Installation,'' Technical Report No. APL TM3-96, Applied Physics Laboratory, University of Washington (1996).
[14] M. Dzieciuch, P. Worcester, and W. Munk, ''Turning point filters: Analysis of sound propagation on a gyre-scale,'' J. Acoust. Soc. Am. **110**, 135–149 (2001).
[15] B. McDonald, M. Collins, W. Kuperman, and K. Heaney, ''Comparison of data and model predictions for Heard Island acoustic transmissions,'' J. Acoust. Soc. Am. **96**, 2357–2370 (1994).
[16] J. A. Colosi and S. M. Flatté, ''Mode coupling by internal waves for multimegameter acoustic propagation in the ocean,'' J. Acoust. Soc. Am. **100**, 3607–3620 (1996).
[17] Z. Zhang and C. Tindle, ''Improved equivalent fluid approximations for a low shear speed ocean bottom,'' J. Acoust. Soc. Am. **98**, 3391–3396 (1995).
[18] M. Collins, ''A split-step Padé solution for the parabolic equation method,'' J. Acoust. Soc. Am. **93**, 1736–1742 (1993).
[19] D. Hill, ''Crustal structure of the island of Hawaii from seismic-refraction measurements,'' Bull. Seismol. Soc. Am. **59**, 101–130 (1969).
[20] F. Klein, ''A linear gradient crustal model for south Hawaii,'' Bull. Seismol. Soc. Am. **56**, 1084–1091 (1981).
[21] E. Hamilton, ''Geoacoustic modeling of the seafloor,'' J. Acoust. Soc. Am. **68**, 1313–1340 (1980).
[22] E. Hamilton, ''Sound velocity–density relations in sea-floor sediments and rocks,'' J. Acoust. Soc. Am. **63**, 366–377 (1978).
[23] P. F. Worcester, B. D. Cornuelle, M. A. Dzieciuch, W. H. Munk, B. M. Howe, J. A. Mercer, R. C. Spindel, J. A. Colosi, K. Metzger, T. G. Birdsall, and A. B. Baggeroer, ''A test of basin-scale acoustic thermometry using a large-aperture vertical array at 3250-km range in the eastern North Pacific Ocean,'' J. Acoust. Soc. Am. **105**, 3185–3201 (1999).
[24] J. Spiesberger and K. Metzger, ''A new algorithm for sound speed in seawater,'' J. Acoust. Soc. Am. **89**, 2677–2688 (1991).

The Kauai Near-Source Test (KNST): Modeling and measurements of downslope propagation near the North Pacific Acoustic Laboratory (NPAL) Kauai source

Kevin D. Heaney[a)]
OASIS Inc., 6304 Anneliese Dr., Falls Church, VA 22044

(Received 27 April 2004; revised 26 August 2004; accepted 27 August 2004)

In April 2002, a series of recordings were taken near the North Pacific Acoustic Laboratory (NPAL) Kauai source. This source transmits broadband low-frequency signals to receivers in various locations across the Pacific. Travel-time calculations based upon the nonbottom interacting PE or ray-trace modeling for the Kauai source lagged the actual first arrival by 1 second or so, a disturbing discrepancy that needed answering. The Kauai Near-Source Test (KNST) was designed to take measurements near the Kauai source to examine the effect of the local bathymetry on the signal transmitted into the ocean. Six receptions, using a 100 m VLA, at ranges from 0.2 to 52 km were taken. Parabolic Equation (PE) simulations of the channel impulse response for the measurement ranges are in good agreement with the observations. PE simulations using an exposed basalt seafloor match the data much better than those with a silty-sand cover. A primary result of this test was the indication that there may be significant energy in the first bottom bounce, comparable to the direct-path (purely refracted) ray. This implies that the seafloor near the source influences the arrival times and patterns of the field at long distances, and should be taken into account. [DOI: 10.1121/1.1854990]

PACS numbers: 43.30−k; 43.30Qd; 43.30Zk [AIT]

Pages: 1635–1642

I. INTRODUCTION

The field of long-range ocean acoustics has advanced to the point where forward modeling is understood well enough to permit the inversion of acoustic travel times for ocean sound speed information.[1] The presence of a sound channel in the deep ocean makes possible the transmission of low-frequency acoustic energy to very long distances. In 1991 the Heard Island Feasibility Test[2] demonstrated that sound could be effectively transmitted across the globe. Using bottom mounted sources (to limit source motion and provide longevity) the Acoustic Thermometry of the Ocean Climate (ATOC)[3] and North Pacific Acoustic Laboratory (NPAL)[4] experiments have been transmitting broadband acoustic signals intermittently since 1996 to receivers distributed throughout the North Pacific Ocean.

For the Pioneer Seamount source of the ATOC experiment (located on Pioneer Seamount off the coast of central California at a depth of 938 m) the acoustic predictions from ray trace modeling matched the arrival times of the receptions on a variety of receivers (Vertical Line Arrays, Horizontal Line Arrays) relatively well. For the Kauai source this was not the case. Using ray trace models including only purely refracted paths an off-set of as much as 1 second was required to align the early arriving wave-fronts. The 1-second offset was required for deep bottom mounted SOSUS arrays as well as for the NPAL 2-D array off Point Sur, California. Was there a source timing error? Was the physical interaction of the sound with the bottom different for the Kauai source than it was for the Pioneer Seamount source?

A critical question is what is the probe pulse being sent into the ocean waveguide? The signal sent from the power amplifier to the source is well known. The impulse response of the source is stable and well measured prior to the experiment. The ocean waveguide is deep for most of the propagation path and deep-water propagation is well understood. Changes between predictions and receptions for this portion of the path are rightly attributed to fluctuations in the sound speed profile through the inversion process.

In order to begin to understand the nature of the differences between the Pioneer source and the Kauai source, recordings near the source were required. In April 2002 the Kauai Near-Source Test (KNST) was conducted. Measurements were taken directly above the source with a short (100 m) vertical line array deployed over a period containing two days of transmissions. A total of six transmissions were recorded at ranges between 200 m and 52 km. Careful attention was paid to timing and array calibration. The conclusions are that there is no error in the source timing and that the first bottom bounce has as much energy as the purely refracted path.

The outline of this paper is as follows. Numerical modeling results for the downslope environment near Kauai are given in Sec. II. These simulations help explain the physics of downslope propagation, illustrate the affect of sediment type on propagation modeling, and help in planning the geometry of the near-source recordings. In Sec. III the experiment is described and the results from it are presented. The measurements are compared with the PE modeling results in Sec. IV. Section V is the conclusion.

[a)]Electronic mail: Oceansound04@yahoo.com

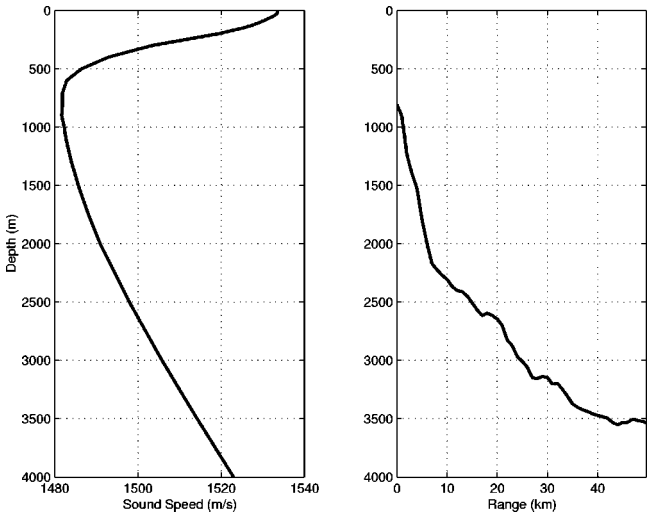


FIG. 1. Kauai Near-Source Test (KNST) sound speed profile and local downslope bathymetry from archival data.

II. MODELING DOWNSLOPE PROPAGATION FROM THE KAUAI SOURCE

A. The NPAL Kauai source

The Kauai source for the NPAL experiment is mounted on the seafloor, at a depth of 811 m, 14.8 km north of Kauai (159.570 °W, 22.349 °N). In addition to receiving the transmitted signals on the U.S. Navy SOSUS network, five vertical line arrays were deployed in 1800 m of water on the continental slope off the coast of Point Sur California. The arrays were deployed broadside to the expected arrival angle for the purpose of measuring horizontal coherence lengths for very long-range propagation experiments. The details and results of this experiment are summarized by Worcester *et al.*[5] in this special issue.

The source transmits 44 periods of a phase-coded pseudo-random *m*-sequence with a center frequency of 75 Hz ($Q=2$). Each period is 27.28 s long, yielding a 20-minute transmission. The broadband source level is 195 dB *re* 1 μPa at 1 m. Each transmission is preceded by a five-minute ramp-up period during which the source level is gradually increased to full power. The source transmits every fourth day, with six 20-minute transmissions at four-hour intervals on that day. The transmission schedule was designed to satisfy marine mammal mitigation concerns and permit the accurate estimation (and removal prior to the performing the tomographic inversion) of the tidal component.

The source location was chosen to meet the constraints of being close to shore (for power) and at a location where the bathymetry intersected the sound channel axis. The sound channel axis in the central Pacific is at a depth of roughly 800 m. The source is placed on the axis so that it excites acoustic energy at all propagation angles, providing the best sampling of the water column for ocean acoustic tomography. The bathymetry from the source into deep water along the path from the Kauai source to the NPAL receiver array is shown in Fig. 1. The slope of the island is much steeper than typical continental shelf slopes and therefore abyssal depths are achieved within 50 km. The sound speed profile at the source (in deep water) is also shown in Fig. 1.

B. Single frequency propagation modeling

In order to examine the effects of the bathymetry near the bottom mounted source on acoustic transmissions into deep water, broadband Parabolic Equation (PE) runs were performed using RAM.[5] A difficulty with precise modeling of the downslope propagation is the lack of knowledge of the geo-acoustic parameters. Given the volcanic creation of the island of Kauai, it can be assumed that the basement is basalt, with a very high compressional sound speed on the order of 3500 m/s. Uncertainty in the exact compressional speed is not an important issue due to the severe sound speed contrast between the speed of sound in water and in basalt. The significant geo-acoustic questions are the extent of sediment covering the basalt (the north-slope of Kauai is the wettest place on earth so water run off and sedimentation can be significant), the presence of shear waves (and the associated shear speeds), and the accurate modeling of surface roughness, which has a strong effect on acoustic scattering. In this paper we have addressed the issue of sediment coverage and have not addressed the shear speed or surface roughness issues.

Figure 2 shows the received level at 75 Hz (for a source level of 195 dB *re* 1 μPa at 1 m) for two different sediment assumptions. The top panel is for a sediment modeled as exposed basalt (compressional speed 3500 m/s). It is clear that there is significant acoustic energy at high-angles and that waterborne acoustic energy does propagate after interacting with the seafloor. For this case the assumption that the purely refracted acoustic paths contain all of the energy that arrives at the receiver several thousands of km away is not valid. The second geo-acoustic example (shown in the lower panel of Fig. 2) has a 100 m silty-sand sediment (compressional sound speed 1520 m/s) overlying a basalt basement. From the resulting field plot (Fig. 2) it is clear that the bottom interacting energy is completely stripped from the water column. The geo-acoustic profile in the sediment is depth dependent, assuming a homogenous sediment and using relations based upon Hamilton[6] and Bachman.[7] For this sediment, the approximation that all bottom interacting energy is gone and only rays which do not bottom interact are present and contain significant energy appears valid.

Two paths of significance to this experiment are evident in Fig. 2. The direct path, which goes up at the source and is fully refracted, has an upper turning point at a range of 50 km. There is also a path that interacts with the seafloor at 5–10 km and has an upper turning point at 25 km. This is a critical path to the description of the effect of the seafloor scattering on the long-range tomographic signal. This path has the features that it interacts with the seafloor once and then is fully refracted for the rest of the propagation path. If this is the case, and there is substantial energy in this path, then it will have a longer travel time than the direct path when it leaves the near-source region, due to the surface and bottom interaction near the source. A full-field, broadband,range-dependent propagation run is required to

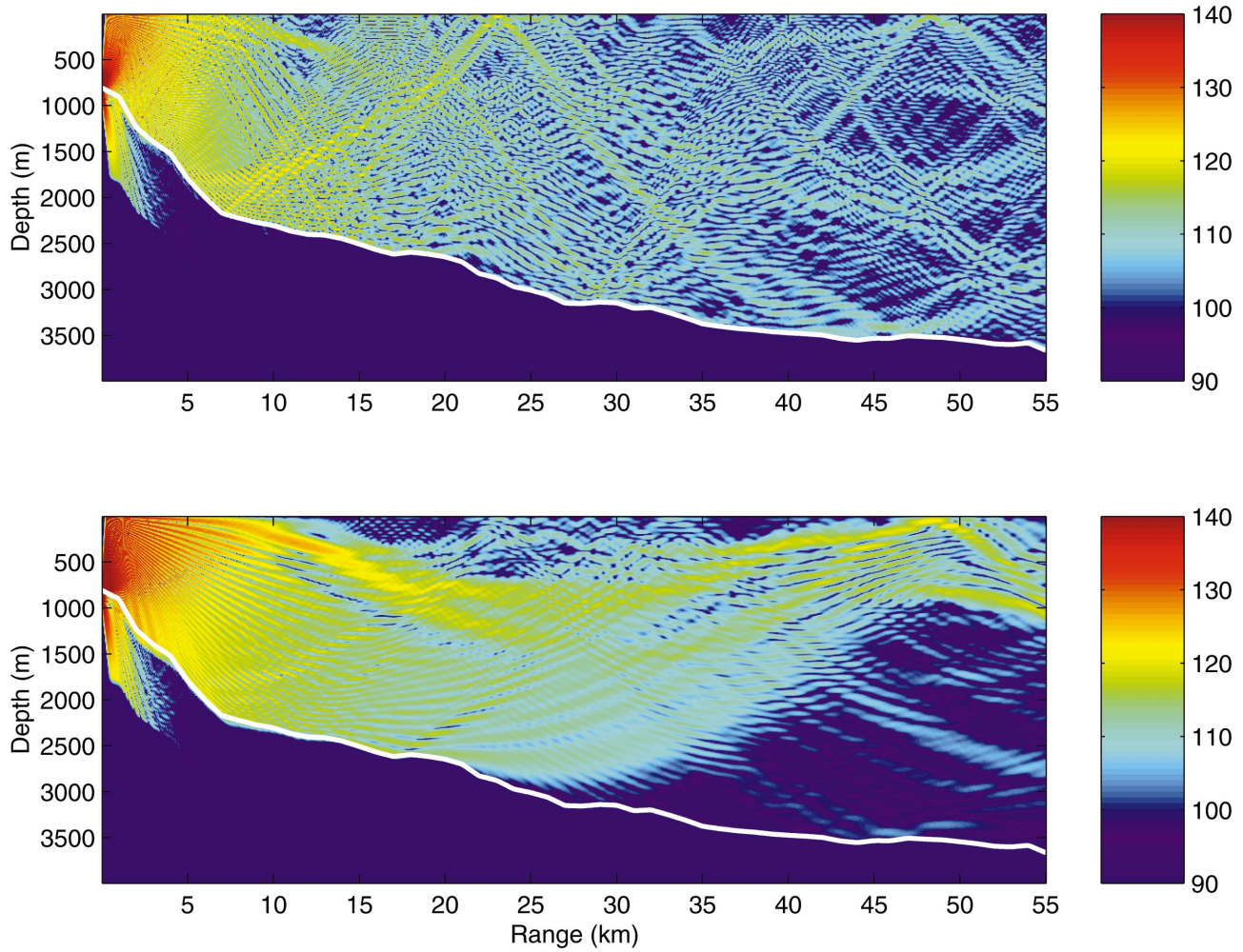


FIG. 2. KNST parabolic equation received level (dB *re* 1 μPA) simulation for two sediments (top) exposed basalt ($cp = 3500$ m/s) and (bottom) 100 m of silty-sand ($cp = 1520$ m/s).

fully understand how much energy is contained in each set of paths.

C. Broadband modeling

The impulse responses at 50 km for the two sediments are shown in Fig. 3. With the soft sediment layer, the bottom bounce is present, but it is weak and does not compete with the direct path. For the exposed basalt the first order bottom bounce has as much energy as the direct path. The path which interacts with the bottom between 5 and 10 km has a lower turning point at 50 km. It is visible at a depth of 2000 m and a time of 41 seconds, which is half a second behind the purely refracted path. This is significant to the tomography problem because this energy could propagate to long ranges.

D. Discussion

There are two mechanisms at work here to permit sound to come off the shelf delayed by 0.5 s. The first is a shallow water propagation phenomenon. In deep water the acoustic dispersion is such that the group velocity is driven by refraction of the sound speed profile and therefore higher angle rays (or higher modes) have higher group velocities because they spend more time (and turn) in regions with higher compressional sound speeds. In shallow water, the opposite is true. In shallow water propagation, the steep angles refract little, and their group velocity is dominated by interaction with the bottom and effectively becomes $v_o \cos\theta$. While the sound is propagating down the shelf, interacting with the bottom in the first 10 km, the propagation is effectively shallow water, giving the high angle energy very low group velocities. Horizontal group velocities can be as low as 1300 m/s for high angle energy. Once in deep water, the modal dispersion switches to the more common SOFAR propagation and the high angle energy is now the fastest. We postulate that there will be a range at which these two paths cross and most of the energy collapses into a single pulse. In order to have the sound effectively propagate to the receiver at great distances, the sound cannot continue to interact with the bottom. The second mechanism, which permits sound to propagate to long ranges, is the angle conversion (or mode coupling) that occurs with each bottom bounce on the continental shelf. As the high angles hit the seafloor, they are converted to lower angles. There is a set of propagation angles that are delayed by 0.5 s coming off the shelf that are scattered to angles that are no longer bottom interacting.

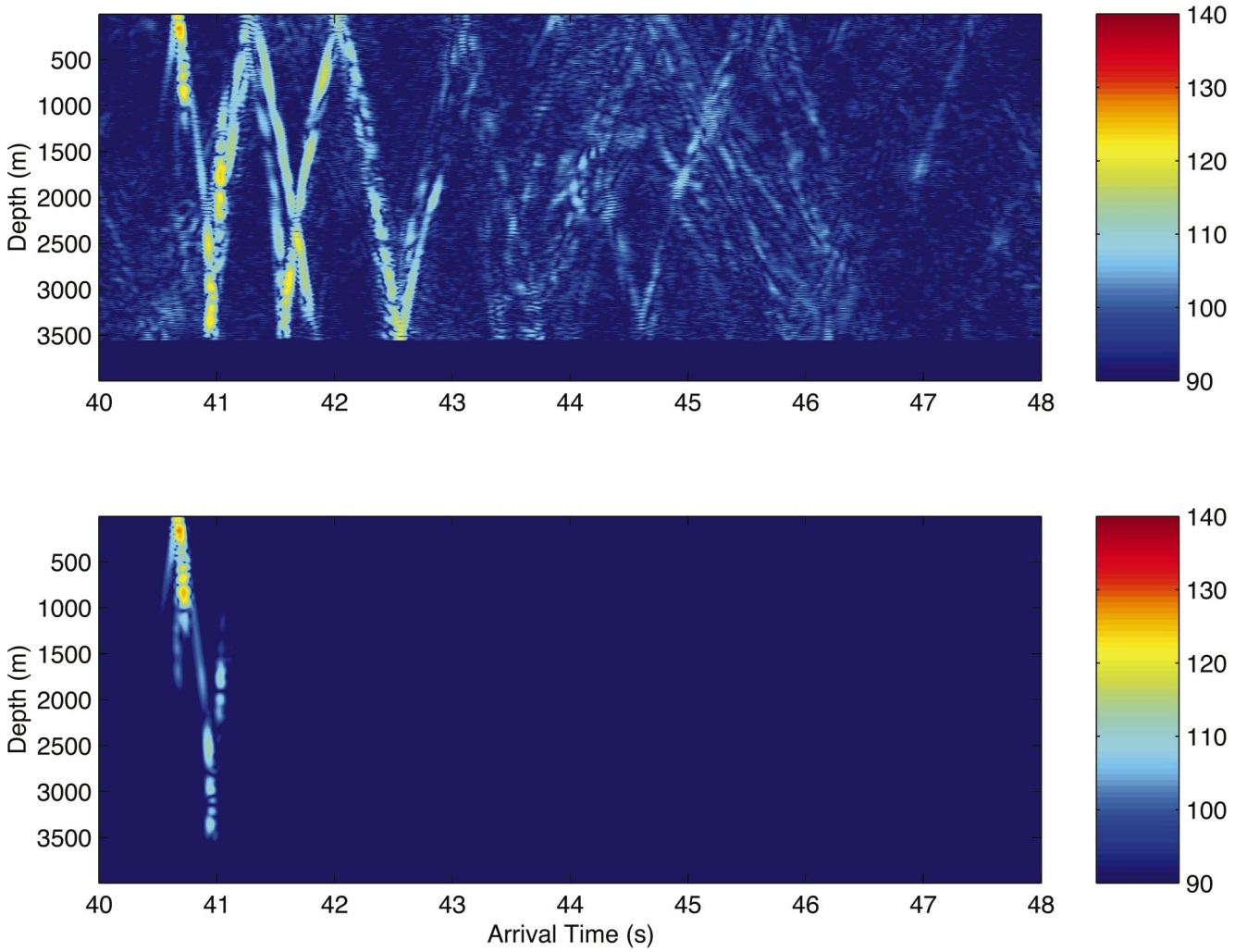


FIG. 3. Broadband PE predictions at 52 km; the range of Transmission 6. (Top) Exposed basalt; (bottom) silty-sand sediment over basalt.

III. THE KAUAI NEAR SOURCE TEST

A. Experiment geometry

In April 2002, the Kauai Near-Source Test was conducted to test the accuracy of the source timing and to determine if the bottom interacting energy was observable at ranges of 25 and 50 km. A 100-m vertical line array was deployed to a depth of 300 m from a small boat off the coast of Kauai. Over two separate transmission days (GMT 97 and 101) six transmissions of the NPAL source were successfully recorded on the vertical line array at ranges of 0.2, 3.2, 5, 25.2, 38 and 50.4 km from the source.

The array contained eight equally spaced elements (12 m spacing) with a total aperture of 100 m. The array had 300 m of cable permitting deployment to a maximum array center depth of 350 m. The inability to deploy the array at the depth of the sound channel was a significant constraint. Wavefronts could only be observed at or near their upper turning points. From the numerical modeling (Fig. 2) it was determined that the direct path has an upper turning point at a range of 50 km, corresponding to the first convergence zone. The first order bottom bounce, however, is off by one half of a cycle and has an upper turning point at 25 km. Thus the two paths can be measured but only for different transmissions at different ranges. The array had no depth sensors for array element localization.

Extra care was taken to synchronize the clocks of the recording system with a GPS satellite IRIG-B clock to make sure travel time could be accurately measured. The IRIG-B signal was added directly to the acoustic signal from the first hydrophone to ensure accurate timing. The hydrophones were well calibrated (when new) and a calibration tone was put on each acoustic data file to permit the accurate calibration of the output signals. The range to the source was computed using the D-GPS location of the boat. Moderate currents led to a drift speed of 0.5 m/s (approximately 1 knot).

B. Measurement results

For all of the recordings the signal-to-noise ratio (SNR) was high enough to see the transmission on a single element spectrogram without matched filtering the impulse response. The single element spectrogram for Transmission 2 (3.2 km) is shown in Fig. 4. Due to the difficulty of array element location during this test, all results will be presented for a single hydrophone (hydrophone #4). The 60–90 Hz broadband signal is clearly visible from $t=400$ s to 1600 s. The source amplitude was stepped up slowly over five minutes, and this is also visible in the spectrogram. The striping mo-

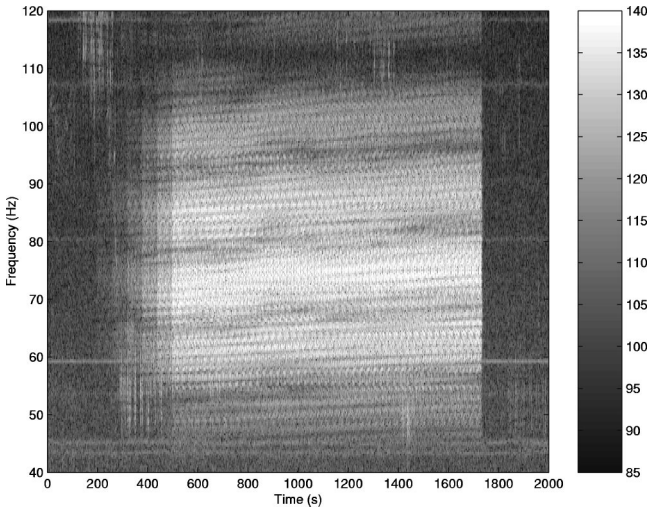


FIG. 4. Spectrogram of received signal for Transmission 2, at a range of 3.2 km. The variations in the signal structure are due to coherent interference and receiver drift.

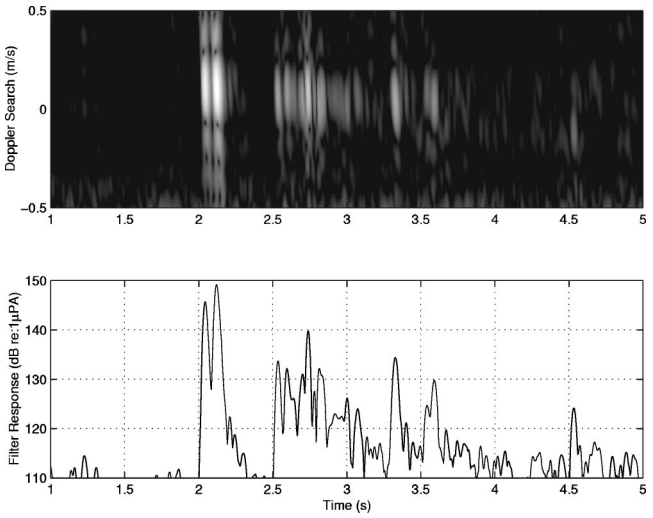


FIG. 5. The 4 period average of the matched filter response of Transmission 2 at a range of 3.2km. The top panel is the result of the search in Doppler space. The peak level in Doppler at each time is plotted in the lower panel. The direct path and surface reflection is visible at 2 s.

tion in the nulls between 50 and 100 Hz is an acoustic path interference pattern caused by the drift of the boat. In downward refracting shallow water environments, a relation exists between the phase and group velocities of the normal modes (and equivalent rays) that yields stable interference patterns in range and frequency.[8] The relation between the differential group and phase speeds amongst different interacting paths (modes) is the waveguide invariant.[9] This phenomenon demonstrates the importance of shallow water propagation, even in this long-range ocean acoustic experiment. Shallow water striations have been used to perform geo-acoustic inversions in other shallow water regions.[10]

To increase the SNR and to determine the impulse response of the ocean channel, the *m*-sequence matched filter is applied. Four periods of the transmission were processed coherently and a search over (source–receiver) Doppler conducted. The Doppler search is to accommodate receiver motion in this case because the source is stationary. The Doppler search and resulting received signal for Transmission 2 at a range of 3.2 km is shown in Fig. 5. The received signal is computed as the peak (in Doppler) arriving at each time.

The pre-matched-filtered SNR for Transmission 2 (3.2 km) was 30 dB. This was computed using Fig. 4, looking at the level difference at the end of the transmission ($t \sim 1800$ s). After matched-filtering (and losing energy due to dispersion) the SNR was 40 dB. This was computed using Fig. 5, by taking the highest early reception, and estimating the noise as the level before the signal arrives. The direct path at this range has a peak received level of 145 dB *re* 1 μPa, the cluster of arrivals with a single bottom bounce have levels near 133 *re* 1 μPa, and the second bottom bounce paths have levels near 130 dB *re* 1 μPa. This indicates approximately a 3 dB per bounce attenuation in the sediment, which is clearly a strongly reflecting sediment at high-angles.

The Doppler is related to the horizontal component of the acoustic velocity vector and as such is different for each arriving ray (due to the vertical angle of the ray). Thus, a simple approach using the Doppler of an arriving ray on a single phone can be used to determine the propagation angle of a wavefront, according to Eq. (1):

$$v_{\text{Doppler}} = v_{\text{boat}} \cos \theta_{\text{ray}},$$

$$\theta_{\text{ray}} = \cos^{-1}\left(\frac{v_{\text{Doppler}}}{v_{\text{boat}}}\right), \tag{1}$$

where v_{Doppler} is the observed Doppler shift, v_{boat} is the drift motion (horizontal component) and θ_{ray} is the angle of the wavefront with the horizontal. The Doppler of the three bundles of arrivals are 0.101, 0.083 and 0.035 m/s. This corresponds, using Eq. (1), to angles of 18°, 35° and 70°, matching those expected for the direct path, the first bottom bounce, and the second bottom bounce.

Transmission 4 was recorded at a range of 25.2 km. The SNR before matched filtering was 10–15 dB with a 75 Hz received level on the order of 125 dB *re* 1 μPa. The processed Transmission 4 (25.2 km) is shown in Fig. 6. The SNR after matched filtering is 25 dB. The single bottom bounce path is clearly visible at 17.7 s and a level of 135 dB *re* 1 μPa. Multiple bottom bounce paths are also visible.

Transmission 6 was recorded at a range of 50.2 km. The 75 Hz received level was on the order of 115 dB *re* 1 μPa. The SNR before matched-filtering was less than 5 dB. The Doppler search and matched-filter response is shown in Fig. 7. Clearly *m*-sequence processing (with a 4-period average) is required to extract the signal at this range. The direct path arrival is visible at 40.8 s with a peak level of 127 dB *re* 1 μPa. This is weaker (even when considering the range doubling with cylindrical spreading ~ 3 dB) than the single bottom bounce path observed in Transmission 4. This will be examined more fully in the next section on model–data comparisons. The bottom and multiple bottom bounce paths are visible with each down approximately 10 dB.

IV. DATA–MODEL COMPARISONS

The received signal for Transmission 2, at a range of 3.2 km, is shown in Fig. 8, along with PE predictions for two

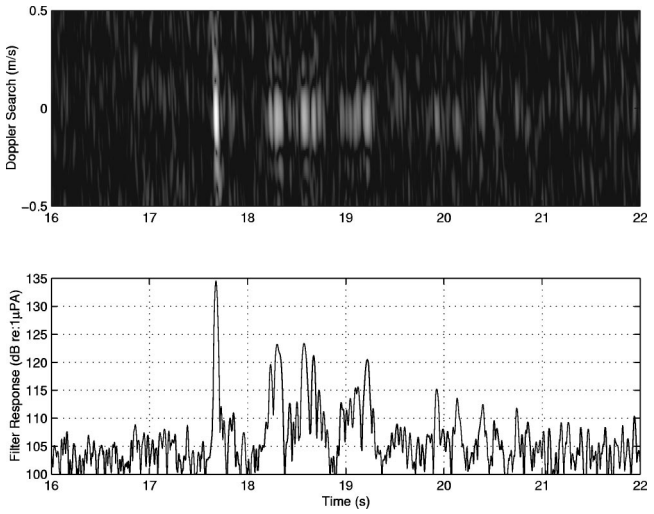


FIG. 6. The 4 period average of the matched filter response of Transmission 4 at a range of 25.2 km. The top panel is the result of the search in Doppler space. The first arrival is identified as the single bottom bounce path with a received level of 135 dB *re* 1 μPa.

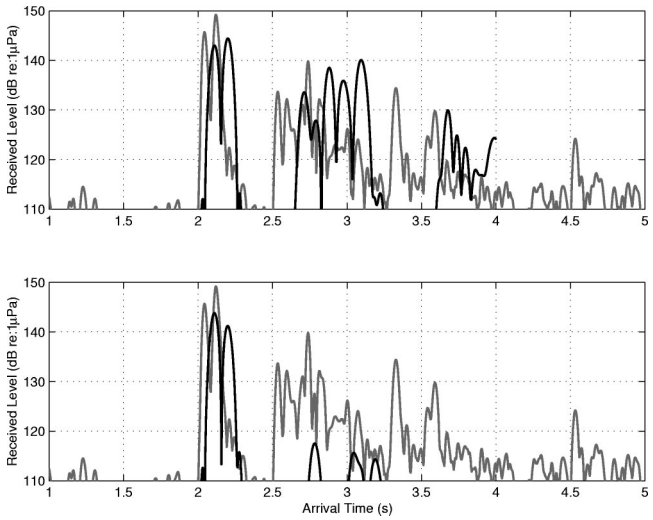


FIG. 8. Model–data comparison for Transmission 2 at a range of 3.2 km. The top panel is the PE simulation with exposed basalt (dark) against the measured data (gray). The lower panel is the PE with a silty sand cover (dark) against the measured data (gray).

sediment types. The first arrival is split into a direct and a surface-reflected path. The first order bottom bounce paths arrive 0.6 to 0.8 s later. The PE simulation captures both of these paths quite well. The silty-sand sediment (lower panel) has single bottom bounce paths that are 20 dB lower than observed or modeled with the exposed basalt sediment. The structure of the bottom bounces is not well resolved by the PE modeling indicating a mismatch in the bathymetry, the geo-acoustics, or the approximations that shear and roughness can be neglected.

The comparison of the measured and predicted impulse response at 25.2 km is shown in Fig. 9. Both the soft and hard sediment geo-acoustic PE computations have a travel time for the fastest path of 17 s, whereas the data has a minimum travel time of 17.7 s. This is a significant difference. This would correspond to an error in the receiver position of at least 1 km. Without a full water column array and the ability to find other acoustic paths, it is difficult to diagnose whether the cause of this error is location error or mismatch in the propagation modeling. The silty-sediment has only one clear arrival, and this is clearly not well matched to the data. The time-spread and the levels of the PE simulation using the basalt sediment matches the measured data well.

Finally, the measured and modeled arrival structure for Transmission 6 (50.2 km) is shown in Fig. 10. At this range the direct path is most clearly visible. The presence of a single purely refracted path arrival, at a level of 128 dB *re* 1 μPa and a time of 40.8 s, is well matched by the PE model using both geo-acoustic assumptions. This is to be expected because of the lack of dependence on the way the sediment is handled in the modeling. Again, the silty-sand sediment predicts only one arrival where in the data and basalt model 3–4 are clearly seen. For this reception, the PE model with a

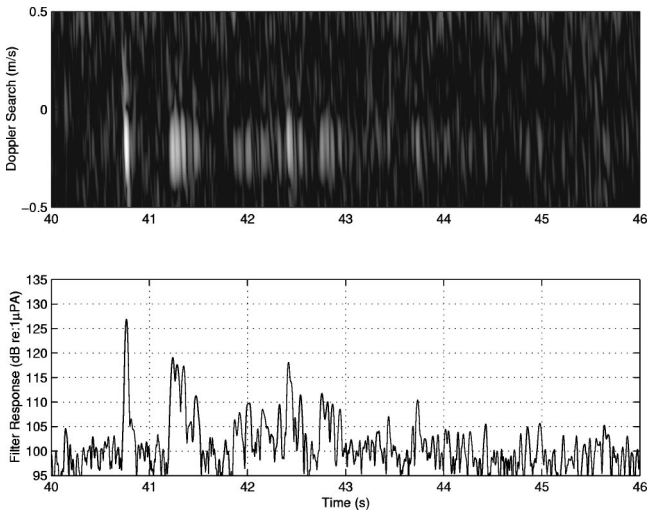


FIG. 7. The 4 period average of the matched filter response of Transmission 6 at a range of 50.2km. The top panel is the result of the search in Doppler space. The first path is identified as the purely refracted path with a level of 127 dB *re* 1 μPa.

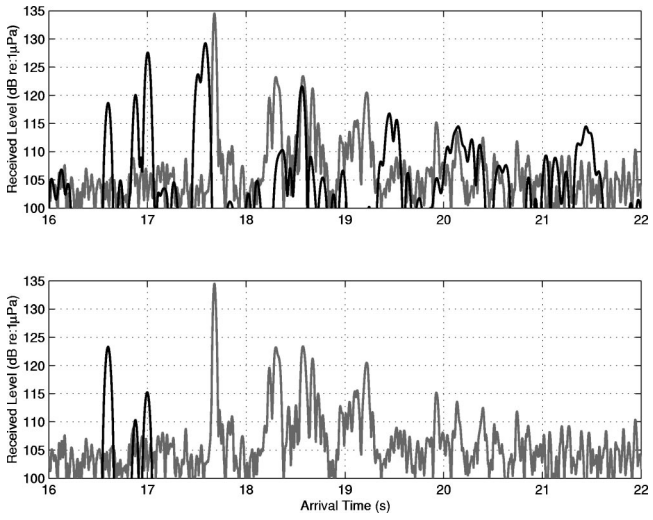


FIG. 9. A comparison of data with simulations for Transmission 4 at a range of 25.2 km. This is the location of the upper turning point of the first bottom bounce. The top panel is the PE simulation with exposed basalt (dark) against the measured data (gray). The lower panel is the PE with a silty-sand cover (dark) against the measured data (gray).

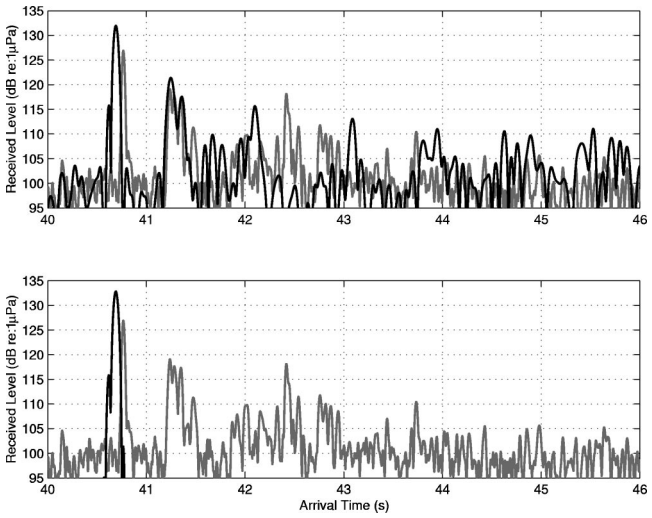


FIG. 10. A comparison of data with simulations for Transmission 6 at a range of 50.2 km. This is the location of the upper turning point of the direct path. The top panel is the PE simulation with exposed basalt (dark) against the measured data (gray). The lower panel is the PE with a silty-sand cover (dark) against the measured data (gray).

basalt basement reproduces the structure of the first bottom bounce very well. This further validates the timing, the calibration and the interpretation of each path based upon PE modeling.

V. CONCLUSION

The effects of interaction with the seafloor on long-range acoustic transmissions from a bottom-mounted source have been investigated. The approach has been a combination of long-range acoustic propagation measurements (the NPAL experiment), near-source computational acoustic modeling and measurements of transmissions near the source. It was found that the bottom bounce path has comparable (even as much as 8 dB more) energy to the direct path. This energy arrives roughly 0.5 s later than the direct path when the sound enters deep water. Comparisons of broadband PE computations with the measurements show qualitative agreement, although the full geo-acoustic inversion has not been completed. The geo-acoustic parameters used in the PE modeling indicate that the seafloor off the island of Kauai is exposed basalt and not thick soft sediment.

The apparent discrepancy between predicted (using PE and ray models) arrival times and those measured at ranges of over 3000 km motivated this work. Any form of source timing error was ruled out by direct acoustic measurements over the source. The results presented here are extended to the effects of seafloor interaction along the entire path from the source to the 2-D NPAL receiver in Vera *et al.*[11]

This analysis was hindered in two respects. Experimentally the lack of accurate array element location (AEL), as well as the limited aperture and deployment depths, were issues. To address this issue, the Basin Acoustic Seamount Scattering Experiment (BASSEX) is planned for the fall of 2004. This test will be a substantially more thorough investigation of the near-source bottom interaction than the KNST. The Office of Naval Research Five Octave Research Array (FORA) will be deployed, permitting both horizontal and vertical measurements to be taken with dense sampling on a 200 m-aperture array. The second hindrance was the lack of roughness and shear in the acoustic modeling. For this limited dataset it is unlikely that these effects could be experimentally isolated and well understood. To accurately model these effects a modeling approach using both the Range-Dependent OASES[12] model and the RAM-S[5] (with shear) model will be used. The combination of higher precision propagation modeling and direct measurement at a variety of ranges will provide us with the opportunity to fully investigate the complex interaction of sound with the seafloor as it propagates from shallow water to deep water. This issue is important in the understanding of source excitation for long-range ocean acoustic tomography, as well as the structure of ambient noise in the deep ocean basins.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research, Code 32-OA, Ocean Acoustics. This work was motivated by the NPAL (North Pacific Acoustic Laboratory experiment, and the author would like to thank the NPAL group for their support, significant scientific discussions, for the use of their source (and its associated *m*-sequences) and for their help in conducting the experiment. The NPAL group is J. A. Colosi, B. D. Cornuelle, B. D. Dushaw, M. A. Dzieciuch, B. M. Howe, J. A. Mercer, W. H. Munk, R. C. Spindel, and P. F. Worcester. J. A. Colosi is at Woods Hole Oceanographic Institution, Woods Hole, MA 02543. B. D. Cornuelle, M. A. Dzieciuch, W. H. Munk, and P. F. Worcester are at the Scripps Institution of Oceanography, University of California at San Diego, La Jolla, CA 92093. B. D. Dushaw, B. M. Howe, J. A. Mercer, and R. C. Spindel are at the Applied Physics Laboratory, University of Washington, Seattle, WA 98105. The author would like to acknowledge Art Teranishi, Brett Castille (Lockheed-Martin ORINCON) and Daniel Sternlicht (Dynatec) for their help in the design of the electronics and in their help during the sea-test. The comments and reviews by Peter Worcester (Scripps Institution of Oceanography) and Bob Spindel (Applied Physics Laboratory, University of Washington) were very helpful. Matthew Dzieciuch (Scripps Institution of Oceanography) provided the *m*-sequence processing code and was extremely helpful in the signal processing. Thank you to the Marine Physical Laboratory of the Scripps Institution of Oceanography for lending us the vertical line array receiver (which was designed by Rob Sohn). Thank you to Sea Engineering for their expertise in small ship handling and deploying hardware at sea. The detailed comments of one reviewer improved this paper.

[1] W. Munk, P. F. Worcester, and C. Wunsch, *Ocean Acoustic Tomography* (Cambridge University Press, Cambridge, England, 1995).
[2] W. Munk, R. C. Spindel, A. B. Baggeroer, and T. G. Birdsall, "The Heard Island feasibility test," J. Acoust. Soc. Am. **96**, 2330–2342 (1994).
[3] ATOC-Consortium, A. B. Baggeroer, T. G. Birdsall, C. W. Clark, J. A. Colosi, B. D. Cornuelle, D. Costa, B. D. Dushaw, M. A. Dzieciuch, A. Forbes, C. Hill, B. M. Howe, R. C. Spindel, D. Stammer, P. F. Worcester, and C. Wunsch, "Ocean climate change: Comparison of acoustic tomography, satellite altimetry and modeling," in Science **281**, 1327–1332 (1998).
[4] P. F. Worcester, R. C. Spindel, and the NPAL Group (B. D. Cornuelle, J.

A. Colosi, B. D. Dushaw, M. A. Dzieciuch, B. M. Howe, J. A. Mercer, W. H. Munk, R. C. Spindel, and P. F. Worcester), ''North Pacific Acoustic Laboratory (NPAL) experiment,'' J. Acoust. Soc. Am. **117**, 1499–1510 (2005).
[5] M. Collins, ''A split-step Pade solution for the parabolic equation method,'' J. Acoust. Soc. Am. **93**, 1736–1746 (1993).
[6] E. L. Hamilton, ''Geoacoustic Modeling of the Seafloor,'' J. Acoust. Soc. Am. **68**(5), 1313–1340 (1980).
[7] R. T. Bachman, ''Parametrization of geoacoustic properties,'' Scripps Institution of Oceanography Report, 1989.
[8] G. L. D'Spain, J. J. Murray, W. S. Hodgkiss, N. O. Booth, and P. W. Schey, ''Mirages in shallow water matched field processing,'' J. Acoust. Soc. Am. **106**, 3245–3265 (1999).
[9] S. D. Chuprov, ''Interference structure of sound field in a layered ocean,'' *Akustika Okeana: Sovremonnoe Sostoyanie*, edited by L. M. Brehkovskikh (Nauka, Moscow, 1982), pp. 71–91.
[10] K. D. Heaney, ''Rapid geoacoustic characterization: applied to range-dependent environments,'' IEEE J. Ocean. Eng. **29**(1), 43–50 (2004).
[11] M. D. Vera, K. D. Heaney, and the NPAL Group (B. D. Cornuelle, J. A. Colosi, B. D. Dushaw, M. A. Dzieciuch, B. M. Howe, J. A. Mercer, W. H. Munk, R. C. Spindel, and P. F. Worcester), ''The effect of bottom interaction on transmissions from the Kauai source,'' J. Acoust. Soc. Am. submitted , (2004).
[12] J. T. Goh, H. Schmidt, and P. Gerstoft, ''Benchmarks for validating range-dependent seismo-acoustic propagation codes,'' IEEE J. Ocean. Eng. **22**, 226–240 (1997).

Statistics and vertical directionality of low-frequency ambient noise at the North Pacific Acoustic Laboratory site

Arthur B. Baggeroer
Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Edward K. Scheer
Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

The NPAL Group (J. A. Colosi, B. D. Cornuelle, B. D. Dushaw, M. A. Dzieciuch, B. M. Howe, J. A. Mercer, W. H. Munk, R. C. Spindel, and P. F. Worcester)[a)]

(Received 16 August 2004; revised 9 December 2004; accepted 9 December 2004)

We examine statistical and directional properties of the ambient noise in the 10–100 Hz frequency band from the NPAL array. Marginal probability densities are estimated as well as mean square levels, skewness and kurtoses in third octave bands. The kurotoses are markedly different from Gaussian except when only distant shipping is present. Extremal levels reached ~150 dB re 1 μ Pa, suggesting levels 60dB greater than the mean ambient were common in the NPAL data sets. Generally, these were passing ships. We select four examples: i) quiescent noise, ii) nearby shipping, iii) whale vocalizations and iv) a micro earthquake for the vertical directional properties of the NPAL noise since they are representative of the phenomena encountered. We find there is modest broadband coherence for most of these cases in their occupancy band across the NPAL aperture. Narrowband coherence analysis from VLA to VLA was not successful due to ambiguities. Examples of localizing sources based upon this coherence are included. kw diagrams allow us to use data above the vertical aliasing frequency. Ducted propagation for both the quiescent and micro earthquake (T phase) are identified and the arrival angles of nearby shipping and whale vocalizations. MFP localizations were modestly successful for nearby sources, but long range ones could not be identified, most likely because of signal mismatch in the MFP replica. [DOI: 10.1121/1.1855035]

PACS numbers: 43.30.Bp, 43.30.Pc, 43.60.Pt, 43.60.Rw [AIT] Pages: 1643–1665

I. INTRODUCTION

Studies of the statistics and vertical directionality properties of low-frequency ambient noise, which for the purposes of this paper are defined to be in the recording band of the NPAL arrays of 10–100 Hz, are unusual, especially in the vertical dimension. In fact, measurements for the vertical apertures of 700–1400 m and horizontal apertures of 3600 m are not available in either the unclassified (and classified) literature. In just the horizontal dimension we have array data from the Navy SOSUS (Sound Surveillance Underwater System), towed arrays such as the SURTASS (Surface Towed Array Sound System), and the TB-29 (a long array towed by submarines), but their classified nature has a limited publication; moreover, their lengths are short compared to the NPAL horizontal extent. Notable exceptions to this are the papers by Curtis *et al.*, and Andrew *et al.*, which summarize single sensor data from many of the Pacific SOSUS arrays within the limits of Navy classification guidelines.[1,2]

Historically, there has been a lot of interest in the vertical structure of ambient noise, especially in the deep ocean. Urick has an excellent summary of most of the work done prior to 1966.[3] Chapter 7 of Urick here is especially relevant. The earliest measurements of vertical directionality cited by Urick were done in 1965 by Axelrod *et al.*,[4] wherein the lowest frequency reported was 112 Hz, which is higher than the band of the NPAL arrays. There have also been a very large number of publications on the role of bubbles. These analyses, however, also concern frequency bands much higher than those being considered here.

Low-frequency, vertical directionality data are scarce, primarily because of the large apertures required for spatial resolution. There are Navy vertical arrays used for both environmental and operational purposes; however, the sparse amount of data from these has been published with most as spectral densities on a sensor by sensor basis.[5] In 1982, data from a 1000 m Vertical Line Array (VLA) in the Arctic[6] and a 1000 m VLA in the east Pacific, which was a prelude to the High Gain Initiative (HGI) were acquired. See the discussion by Mikhalevsky in Baggeroer *et al.*[7] Subsequently, two large vertical apertures (VLAs) for the HGI itself were deployed, but results are not available in the open unclassified literature. The emphasis of these HGI experiments was on Matched Field Processing, so an extensive noise analysis was not done on the vertical structure. Nevertheless there has been a lot done for the horizontal noise structure, e.g., by Makris and Dyer and Langley.[8,9]

At low frequencies, analyzing the vertical structure re-

[a)]J. A. Colosi is at Woods Hole Oceanographic Institution, Woods Hole, MA 02543. B. D. Cornuelle, M. A. Dzieciuch, W. H. Munk, and P. F. Worcester are at Scripps Institution of Oceanography, University of California at San Diego, La Jolla, CA 92093. B. D. Dushaw, B. M. Howe, J. A. Mercer, and R. C. Spindel are at the Applied Physics Laboratory, University of Washington, Seattle, WA 98105.

quires large vertical apertures. When combined with the waveguide effects, this requires a resolution of less than 5°, corresponding to at least an 1800 m array at 10 Hz or a 180 m array at 100 Hz, to account for ducting and noise notch phenomena. Even these dimensions are not adequate for identifying a closely spaced multipath according to vertical angles. At low frequencies the finite aperture, turning point issues, and wave front curvature for modal separation of ducted propagation must be addressed.[10] A very important consideration is the equipment. The very long VLAs and the data recording plus the strum isolation and element positioning for hydrophones are major oceanographic deployments and tricky at best.

Uniformly applicable statements about ambient noise are risky at best. Probably, the most common model used in this low-frequency range is a superposition of (i) the surface noise derived from the Kuperman–Ingenito model because it incorporates the propagation waveguide, (ii) monopole and occasionally dipole point sources generated by Green's function propagation codes, and (iii) sensor white noise, i.e., uncorrelated among sensors.[11] These models are limited, nevertheless, by the ubiquitous background level and the use of simple sources, which does not usually include the real multipole nature, nearby, discrete shipping contributions, clusters of marine mammals and other biologics, or earthquakes, While these model components can be modified to incorporate them, there are a lot of parameters with which to play for inverting data to a model. Certainly one reason against postulating uniformly applicable statements about ambient ocean noise is because so much depends upon the environment: shallow *versus* deep, low *versus* high frequency, wind/rain driven *versus* shipping dominated, static *versus* dynamic environments, etc. Moreover, much is dominated by the array, or lens, which is used to resolve the several noise phenomena. Here, we are concerned with deep water, low-frequency, mostly shipping or biologically dominated, dynamic environments. Even with these caveats, we do not capture all the phenomena and some remain unexplained. Nevertheless, spectral entropy methods indicate structured spectra in 83% of the NPAL records for a single channel and 92% with array gain. There are times of quiescent low level scenarios (10 dB) below the average, but they are not the norm. While ambient noise data is easily obtained simply by an *in situ* measurement, it can be difficult to predict at a specific time and place since one or several phenomena may dominate a particular measurement.

The emphasis here is on the observed levels and the implications for sonar signal processing. We did not have the observational controls such as weather, sea state, shipping density, etc. to argue for causal relations to these environmental variables. The volume edited by Carey and Monahan,[12] and the monograph by Kerman[13] are good sources for a general background. In the Carey and Monahan volume there are several articles of note—Kennedy and Goodnow,[14] who used an array measuring vertical directional spectra down to 75 Hz, Carey *et al.*,[15] who used a parabolic equation approach to model wind excited noise, and Hodgkiss and Fisher,[16] who obtained vertical array data at sites south of NPAL looking at slightly higher frequencies. All these articles were motivated by wind speed correlates. A theoretical discussion of the wind-noise dependence was recently published by Oguz,[17] while Yang examined the environmental impacts on the vertical structure.[18] Recently, Booth *et al*. have examined vertical noise directionality using polarization processing, but at frequencies above the NPAL data.[19]

The NPAL site was close to shipping, the known transit lanes of marine mammals, and micro-earthquakes associated with the North American and Pacific plate boundaries, so these discrete events were probably more common at the NPAL site than at others. As previously mentioned we found the ''quiescent'' wind noise driven environment unusual in contrast to specific discrete sources. As a result we have categorized our data analysis in terms of four types of noise environments:

(1) ''quiescent'' environments where there is no clear spectral structure on a single spectrogram;
(2) ''shipping'' environments where the spectrograms have a clear indication of nearby shipping and the spectral lines associated with the machinery;
(3) ''biologic'' environments where the spectrograms indicate the vocalization ''chirps'' associated with marine mammals typically near the NPAL site;
(4) ''micro-earthquake'' environments where high level transient events with coda and T phases are found with micro-earthquakes.

It is useful to provide the reader with a guide to the several processing algorithms applied to form our conclusions. These are as follows.

(1) Spectral classifications: We first describe a method for classifying spectral structure based upon signal spectra entropy. This measurement has the property that flat spectra in the band maximize the entropy while tonal and high structure minimize it. This has subjective aspects since the data can always be ''whitened'' to modify the entropy. We have not done this. See Gallager or Pratt[20,21] for a complete discussion of the entropy of a time series or image. The result of this is four single channel spectra indicative of the four categories of data dominant at the NPAL site.

(2) Marginal densities and histograms: Virtually all of the performance predictions for sonar are made, assuming a background of Gaussian noise. We make the case that in these discrete environments these are questionable assumptions since the single-most common measure of non-Gaussianity, the kurtosis, is far from Gaussian. The extremals, i.e., the tails of distribution are much too high. This does not, however, suggest that this measure may be used as a classification tool, but simply that the tails on a receiver operating characteristic (ROC), where we want to operate, may understate false alarms. We also explore an unexpected result of the dependence of kurtoses *versus* depth.

(3) Broadband cross-spectral coherence: One of the outstanding questions framing the NPAL experiment was the coherence of the signal from a source near Kauai across the arrays. We examine both the vertical and horizontal coherence of the four categories of ambient noise. We find surprisingly high levels for the quiescent environment and excep-

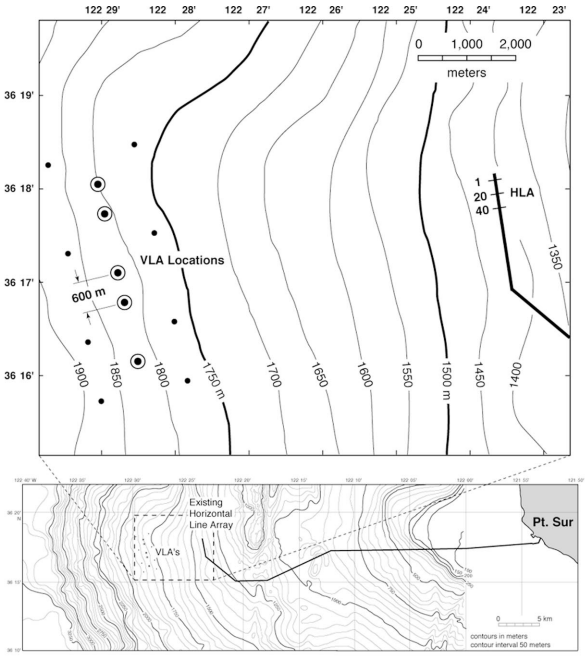
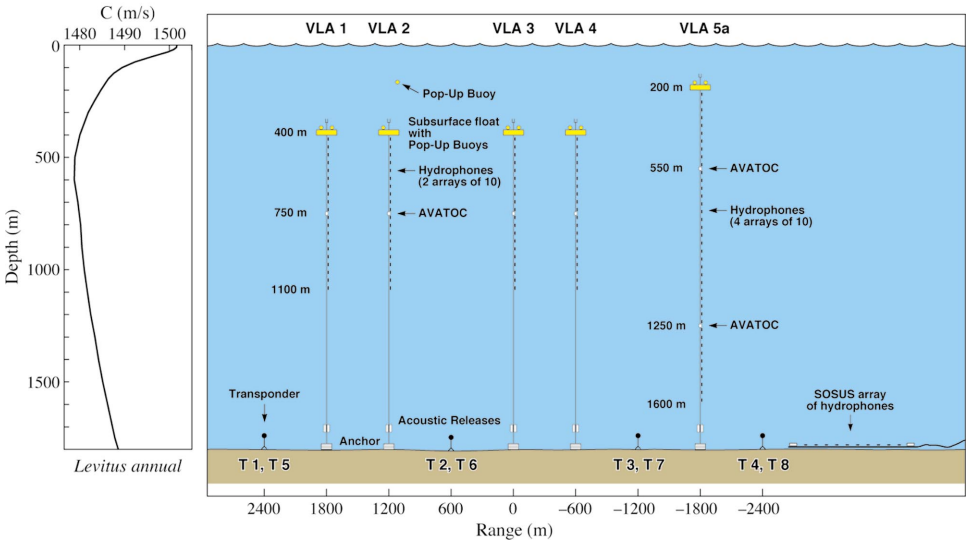


FIG. 1. Plan view and profile of the NPAL vertical line arrays (courtesy of Worcester and Spindel[22]).



tionally high ones for the discrete ones across the full NPAL aperture. For sonars this is especially important since the noise gain (or reduction) is usually where the performance ''dBs'' are found since the very best signal gain is 0 dB gain degradation. The multipath structure is clearly identified in these figures.

(4) $k-\omega$ spectra: The directionality observed on an array is usually summarized in a frequency-azimuth (FRAZ) display or a frequency wave number, or $k-\omega$ spectral distribution. We find well-defined directionality including ducted power in the quiescent spectra and well-defined directions in the others according to the apparent vertical phase speed of the signals. We are able to extend the $k-\omega$ spectra into the aliased region by noting the continuity from the unaliased regions and the aliased images at the higher wave numbers.

(5) Source localization: Vertical arrays can estimate the range and depth of the source using a number of methods. The most robust is multipath ranging, while the most sensitive is matched field processing (MFP). When five VLAs are used, azimuth can also be measured. We use two approaches to source localization, one of the ultimate objectives of a sonar. The wide separation of the VLAs make them work more as an interferometer than a filled aperture, so we can use broadband processing for bearing and multipath matching for depth. We also use MFP as a comparison. It is well known that MFP is quite sensitive to sound velocity profile (SVP) mismatch, so our expectations for long range localizations were not high; nevertheless, we do get good comparisons between the two methods.

Finally, since this article is one of a series in this volume on NPAL, all the common experimental details of the NPAL arrays and system may be found in the accompanying article by Worcester and Spindel.[22] The location is on the continental shelf off Point Sur, California near a SOSUS array, as illustrated in the top of Fig. 1. There are four arrays, 20

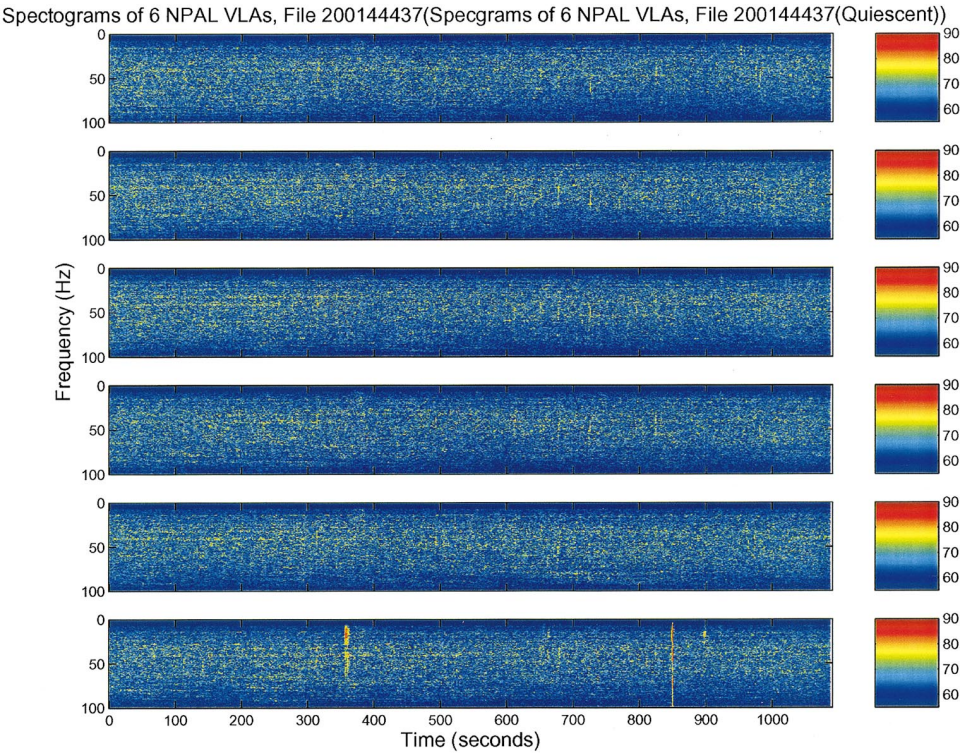


FIG. 2. Quiescent period on the NPAL arrays. Each panel illustrates the spectrogram of the channels noted earlier. The average level is approximately 75 *re* 1 μPa/$\sqrt{\text{Hz}}$.

channel arrays spanning 400–1000 m, and one 40 channel array spanning 200–1600 m. The sensor separation in the vertical is 37.5 m, giving a 20 Hz aliasing vertical frequency. The horizontal spacing was a minimally redundant array[23] with 600 m leading to a 1.2 Hz horizontal aliasing frequency. The sound speed profile used is also illustrated in Fig. 1. Clearly, a single profile is not adequate for long range modeling, so much of the NPAL data analysis has used profiles taken before and at the end of the experiment with archival information used for interpolations. See Worcester and Spindel[22] for more details.

II. SPECTRAL CLASSIFICATIONS OF THE NPAL NOISE

Since the noise at the NPAL arrays is selectively dominated by several physical mechanisms, it is appropriate to illustrate spectra for which one mechanism dominates. One issue was to classify the structure of the spectra among the four. We used a spectral entropy approach that was quite successful in isolating quiescent events into one category and into the three others. Surely additional work using more sophisticated methods of pattern recognition could be done here and have been done with other datasets, but we do discuss them here. For simplicity, the spectral entropy is given by

$$H(S(f))=c_1+c_2\int_{f_{\min}}^{f_{\max}}\ln\left[\frac{S(f)}{P}\right]df,$$

where P is power in the spectral band given by

$$P=\int_{f_{\min}}^{f_{\max}}S(f)df.$$

The constants c_1 and c_2 depend upon the number of samples. They are constant for this analysis since all the segments of the NPAL data are all equal. $f_{\min}=10$ Hz and $f_{\max}=100$ Hz define the spectral analysis region. We found a threshold for $H(S(f)$ empirically by examining all the spectra visually, which led to the best classification choice. (We did not have *a priori* training sets as in most pattern recognition problems.) This led to the conclusion that 0.83% of the 600 data segments had a spectral structure in which tones or features were apparent. For a selected number of beamformed segments 0.92% had a structure that is representative of the well-known problem that array gain just keeps resolving weaker events. We think that we never hit the so-called noise floor or uniform background.

With the 600+NPAL segments, we have chosen four for discussion purposes to illustrate the four classifications. These are illustrated in Figs. 2–5. Six plots appear in each of these figures, containing, from top to bottom, spectra from VLAs 1 through 6. For each VLA spectrum the channel selected is the ''middle'' one.[24]

(1) Quiescent: Fig. 2 shows one of the quieter times in these data when the level was 10 dB below the average level for the NPAL experiment.[25] While we can still discern some shipping lines looking horizontally across the figures, these are most likely very distant and the spectra are quite unremarkable without any distinguishing features. We believe, however, that the ''spike'' in the bottom subplot (VLA6) is real, based upon examining the time series around it in de-

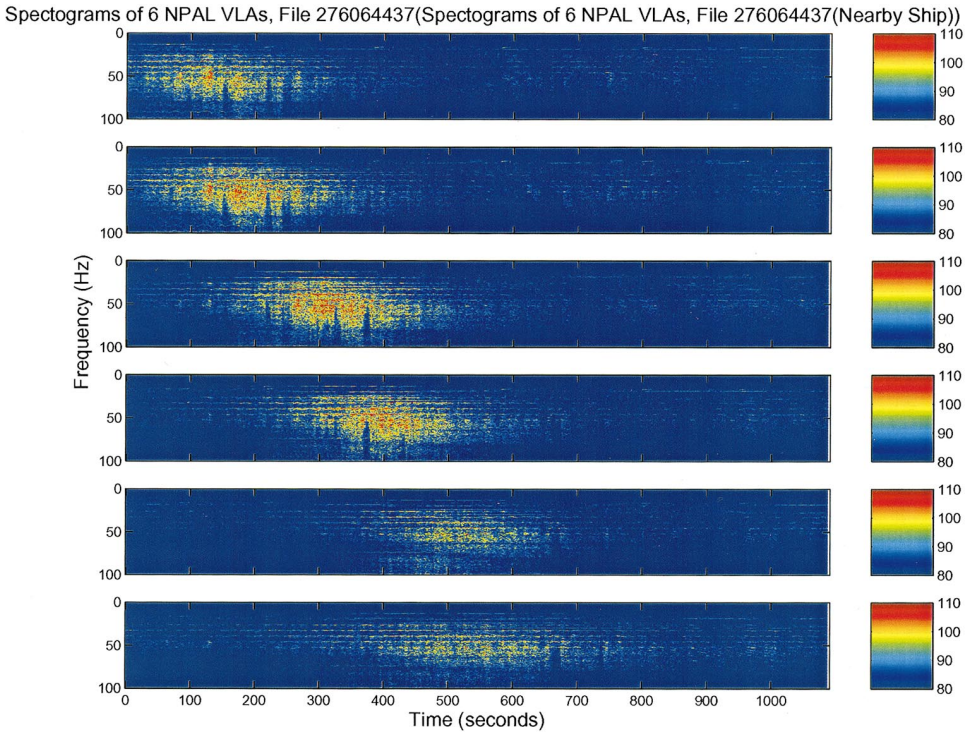


FIG. 3. Ship passing near the NPAL arrays in the band 0–100 Hz. One can observe the moveout as the ship passes VLA 1 to VLA 5–6.

tail. This phone, one of the deepest of all the deployed sensors, contains many more ''spikes'' than the other channels. When one examines adjacent channels they are also present, but with a much lower amplitude. We might associate this with ''fish bite'' a well-known problem for moorings. Overall, the NPAL arrays observed signals over a large dynamic range. Some pressure levels reached 150 dB re 1 μPa, associate with shipping, and almost saturate the digitizer.

(2) Shipping: Fig. 3 is indicative of the shipping observed at the NPAL arrays. (The arrays were located where ships steering southwesterly out of San Francisco would run quite close to the array.) We can observe a ship transiting

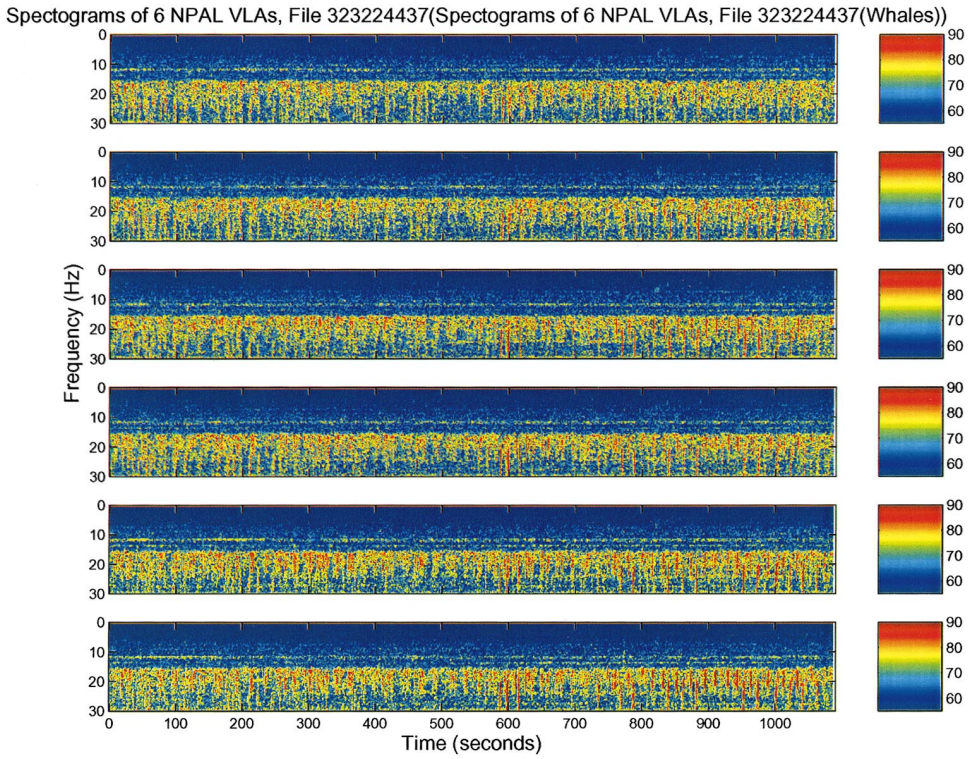


FIG. 4. Whales near the NPAL arrays in the band 0–30 Hz. One can observe the distinctive multitone chirps in each call; moreover, there appear to be two distinctive calls—one multitone while the other concentrated in a single sweep.

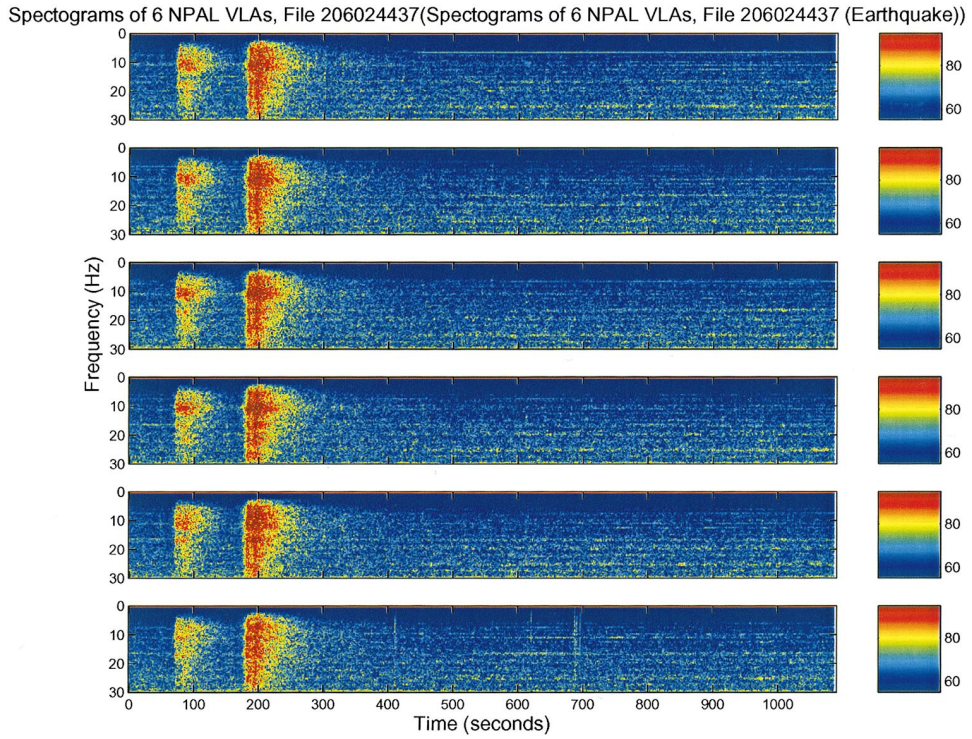


FIG. 5. One of many earthquakes observed in the band 0–30 Hz.

across the arrays and a continuous spectral lineup of the machinery lines. Later we use these data to demonstrate MFP on the NPAL arrays. A careful examination of some of the later times in the upper channels reveals interference effects. Note that the time intervals among the peak levels are approximately 80 and 160 s, which corresponds to a 7.5 m/s or 15 kt speed if the ship was a course going directly over the NPAL arrays. The speed would be higher if the track was offset or not parallel to the array lineup. We attempted to correlate the time series with the time delay of the acoustic propagation times across arrays (0.4–2.6 s), but were not able to do so consistently on a narrow band basis. This is this discussed in the section on coherence. The ship is close by and not large. The proximity is suggested since it fades in and out as it transits the VLAs, so it must be on the steep part of a transmission loss (TL) curve. Second, the peak spectral levels are approximately 110 dB and just 20 dB above the ambient noise level of 90 dB. Allowing 50–60 dB TL for the slant range to the sensors leads to levels of 160–170 dB, which is not an especially high source level.

(3) Whale vocalizations: Fig. 4 illustrates whale vocalizations. Whale calls, characterized by chirps (most with harmonics), are present in 50%–75% of the NPAL data. Most are centered near 16–20 Hz, suggesting that the source is finback whales; however, there are occasions when they are in the 25–45 Hz band. See Curtis *et al.*[1] for a more complete discussion on species identification. It turns out that these calls are exceptionally coherent, leading to excellent beamforming results. While straightforward to do, we have not attempted any pulse compression of the chirps, for example, using one pulse as a replica and then pulse compressing (matched filtering) earlier and subsequent chirps.

(4) Micro-earthquakes: The NPAL site is quite close to the plate boundary separating the North American and the Pacific plates. It should not be surprising that we observed many earthquakes during our observations. Figure 5 illustrates two of these microseisms. We did not attempt to beamform them or make an azimuth estimation since the conical angles and the high ambiguities introduced by such a sparse horizontal array lead to very ambiguous results.

III. MARGINAL DENSITIES AND HISTOGRAMS

In this section we summarize the estimated statistics for single point measurements on the NPAL arrays. Figure 6 is a composite of time series from year day 197 to 343 of the power in dB for middle sensors of each VLA.

Table I summarizes the spectral densities in roughly octave bands plus the composite over the full band. The normalized standard deviation of each measurement is

$$\sigma = 1/\sqrt{c_w * B * 1200},$$

where c_w is the window factor (3/8) for the Hanning one used, and B is the processing bandwidth for each of the bands 10, 20, 40, and 90 Hz, respectively, and the 1000 sec is the time duration of an NPAL segment of data. These are approximately consistent with the data from Hodgkiss and Fisher,[16] once their units from perdegree levels are converted to an integrated level. The tabulated data are also close to the 89 dB measured at 57 Hz during the Heard Island Feasibility Test.[26]

There is a lot of interest in peak pressure levels naturally occurring in the ocean at low frequencies whether from ships, earthquakes, or marine mammals because of the concern of the impact of anthropogenic noise from marine life. Table II summarizes the peak pressure levels observed. The

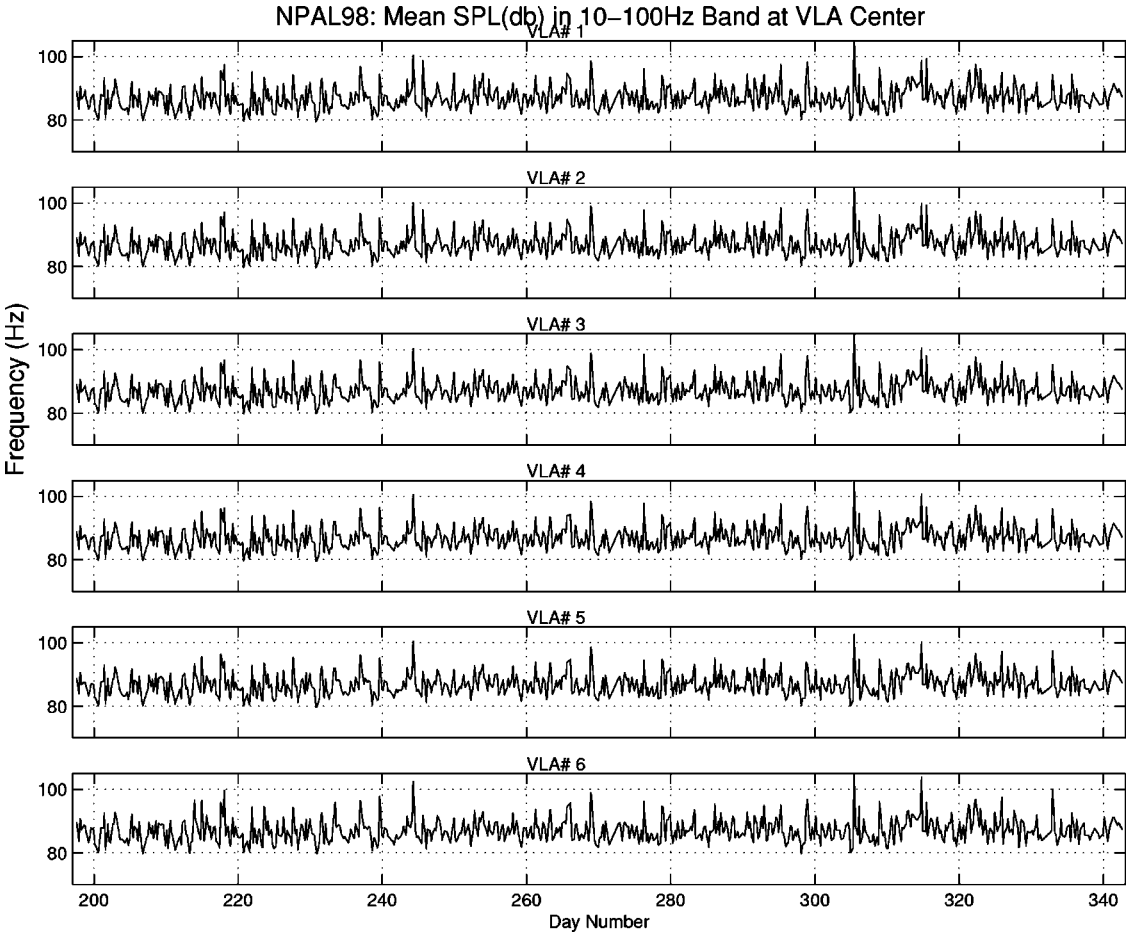


FIG. 6. Mean sound pressure level versus yearday, 6 VLAs on the mid-depth phones. Spectral densities are given in Table I.

full band has the highest level and indicates that 138–150 dB re 1 μPa were observed. The upcoming analysis of the kurtoses suggest that these were not rare events.

The assumption of Gaussian processes is embedded in most of sonar signal processing, e.g., see Urick or Van Trees.[3,27] Since the NPAL data are extensive and recorded with well-calibrated sensors that have proven to be quite reliable for several deployments since the Acoustic Thermometry of Ocean Climate (ATOC) experiment, we took the opportunity to check this assumption by measuring the kurtoses, one of the simplest indications of non-Gaussianity. If a real process is Gaussian, its kurtosis is 3.[28] (For complex data such as spectra the kurtosis is 2 if Gaussian.) If it is

TABLE I. Average spectral densities in approximately third octave bands for the three depths in the NPAL array (deepest to shallowest; see the text for depths of sensors).

Mean spectral densities (DB) re 1 μPa/$\sqrt{\text{Hz}}$						
			10–20 Hz band			
Channel	VLA1	VLA2	VLA3	VLA4	VLA5	VLA6
CH1	88.5	89.2	88.4	88.5	88.1	88.5
CH10	88.7	88.8	88.8	88.4	88.5	88.8
CH19	92.4	90.3	90.3	91.7	88.6	89.3
			20–40 Hz band			
CH1	90.5	90.7	90.5	90.5	90.3	90.6
CH10	90.7	90.8	90.8	90.4	90.5	90.7
CH19	91.2	90.9	90.7	91.0	90.6	90.7
			40–100 Hz band			
CH1	87.4	87.1	87.1	87.0	86.7	86.4
CH10	87.4	87.5	87.6	87.2	87.1	87.6
CH19	87.6	87.6	87.5	87.2	87.4	87.7
			Broadband: 10–100 Hz			
CH1	86.7	88.3	88.3	88.3	88.0	88.6
CH10	85.8	88.8	88.8	88.4	88.4	88.7
CH19	89.1	89.0	89.0	89.4	88.6	88.9

TABLE II. Peak pressure levels for all the NPAL three channels on all the NPAL time series in four bands.

Maximum Levels re 1 μPa						
			10–20 Hz band			
Channel	VLA1	VLA2	VLA3	VLA4	VLA5	VLA6
CH1	142.5	140.8	143.1	141.8	141.1	141.5
CH10	142.3	144.5	145.8	143.0	139.9	147.9
CH19	140.8	142.7	142.5	145.7	140.8	141.4
			20–40 Hz band			
CH1	145.6	143.2	145.4	140.4	144.3	141.6
CH10	147.0	142.5	139.7	138.5	146.7	150.0
CH19	149.1	145.3	146.5	148.7	135.3	139.4
			40–100 Hz band			
CH1	150.6	148.9	147.5	145.3	143.8	145.2
CH10	150.8	149.8	149.1	146.4	145.8	151.6
CH19	149.6	150.4	150.6	150.8	145.2	148.0
			Broadband: 10–100 Hz			
CH1	150.9	149.4	149.5	145.7	149.0	145.8
CH10	151.9	150.3	148.7	146.8	151.1	140.0
CH19	141.0	141.3	152.3	152.1	145.3	147.8

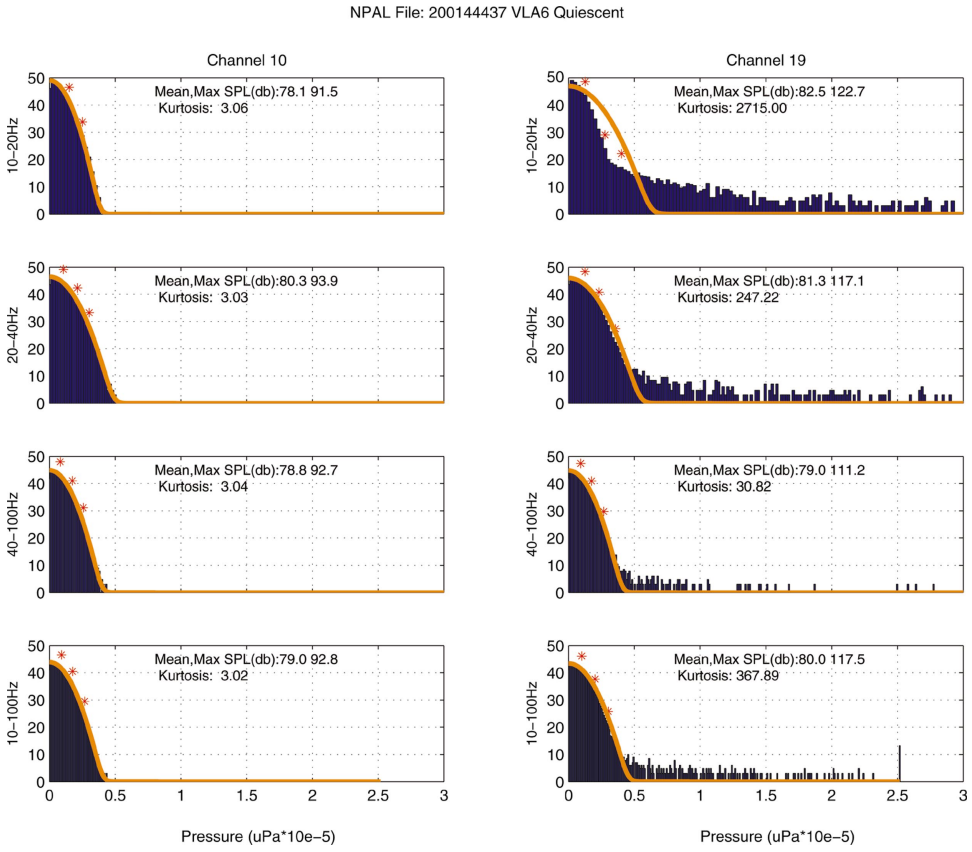


FIG. 7. Histogram of data on VLA-6, Channels 10 and 19 during the ''quiescent'' time of Fig. 2. Note the large number of extremals on Channel 19.

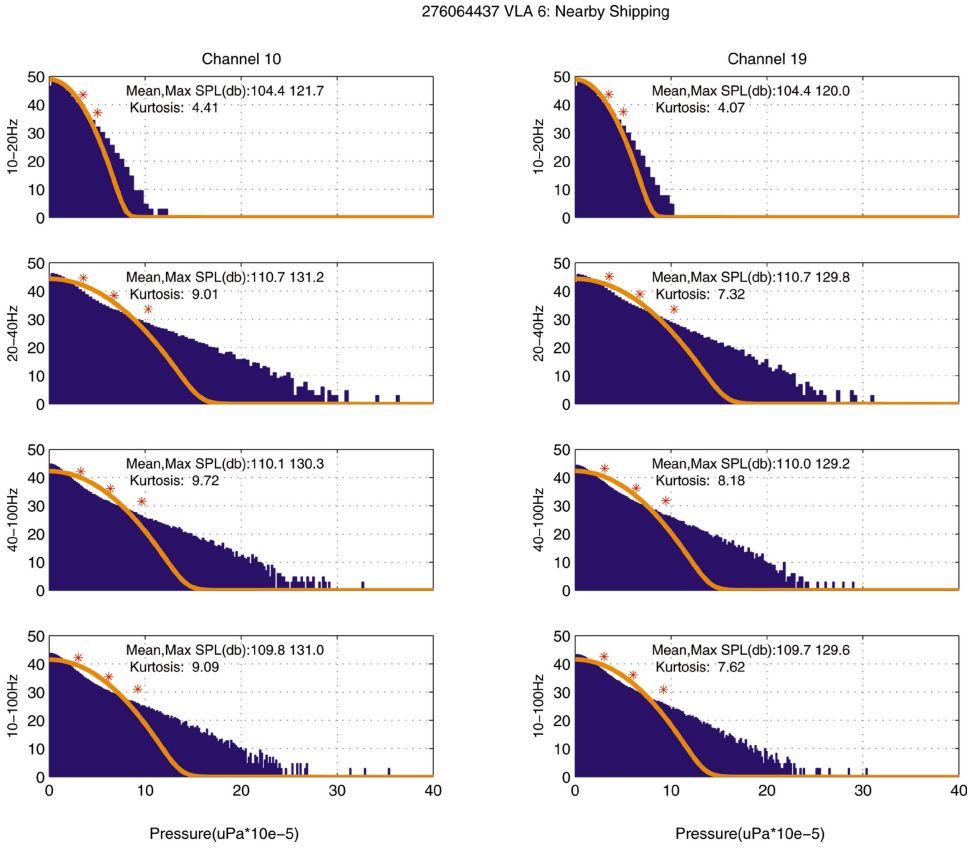


FIG. 8. Histogram of data on VLA-6, Channels 10 and 19 during the ''nearby shipping'' time of Fig. 3.

higher, the probability distribution is ''wider'' than a Gaussian, i.e., there are more frequent higher levels on the tail of the distribution and these are the extremals that are of concern and also impact false alarm estimates. Many non-Gaussian and bispectra algorithms exploit this, e.g., for machinery lines for shipping. The results are a bit perplexing and are discussed below.

Figure 7 illustrates the histograms for the ''quiescent'' spectra of Fig. 2 using the middle and lower channels (Channels 10 and 19) on VLA-6, corresponding to water depths of 1200 and 1530 m. We display the corresponding histograms in the same four frequency bands as Table I. For this there are 60 000 data points, partitioned into 301 bins, with each binwidth equivalent to 2000 μPa. Note: The vertical axis is in units of $10\log 10$ (the number of ''hits'' in a bin). (This can be confusing when applying 68% for 1 sigma and 95% for 2 sigma estimates of area.) The orange line is a zero mean log of a Gaussian distribution with the estimated variance, and the horizontal coordinate is the bin number divided by 10 000. The asterisks correspond to one, two, and three sigma points of a Gaussian distribution. The log scale on the vertical axis was used to accommodate the wide dynamic range at extremal and unusual levels.

We note that the shallower channel (Channel 10) is for all practical purposes well modeled with the Gaussian fit. The kurtoses are nearly equal to three and the fits almost overprint the data. On the lower phone (Channel 19) of the same data set and on the same array, the Gaussian model clearly has problems. The lowest-frequency band has an extremely high level with a significantly wider distribution and a corresponding high kurtosis. Skipping to the third band, the fit is better, but the extremal levels are still much too frequent. Recall that with the Gaussian model, bin occupancy decreases exponentially with a quadratic rate, so such high level events should be extremely rare.

The same histogram analysis and Gaussian fitting done on the ''nearby shipping'' spectra of Fig. 3 are shown in Fig. 8. Much work has been done on exploiting non-Gaussianity to identify ships and possibly for target classification in ASW. The lower-frequency band is closest to Gaussian with kurtoses near four. As we move up in the frequency toward the traditional shipping bands, both the levels increase since we use a constant bin width and the kurtoses do as well. Here, the differences between the two channels is less evident, suggesting that the incident shipping signal dominates.

The histogram for the ''whale'' example illustrated in Fig. 9 clearly shows the impact of the vocalizations. In the band of vocalization the histogram is definitively non-Gaussian with a very large number of points at modest extremal levels. As we move up toward the bands where there is a low vocalization content, the Gaussian fit is quite reasonable. Again, there are many more extremal levels for the deeper channel.

Finally, the histogram and probability density for the ''earthquake'' signal in Fig. 5 are indicated in Fig. 10. The data show non-Gaussian behavior within the bands where most of the earthquake energy is concentrated, but close to Gaussian at the higher bands, where it has little energy. Without more analysis it is hard to suggest a good probability model, but all statistical tests would certainly reject the Gaussian model for all bands. At very low frequencies it is often argued that these micro-earthquakes, whose number increases inversely by ten times the magnitude, lead to a superimposition of a large number of ''events'' in the data and lead to Gaussian data *via* the central limit theorem. (The central limit theorem asserts that the probability density of a sum approaches a Gaussian as the number of events increases if no one term in the sum remains dominant. The book by Doob is one of the well-known texts for the central limit theorem for random processes.[29]

A summary of kurtosis behavior for the three selected channels of the NPAL array for the broadband case (10–90 Hz) statistics is given in Fig. 11. Each of the six subplots gives the number VLA segments with a kurtosis less than the level, i.e., it is cumulative. The red is for the upper channel, the blue for the middle, and green for the deep channel. The black horizontal line is at a level of 3, the nominal value for a Gaussian process. The plots are for the entire NPAL ensemble. The horizontal axis is the number of segments. Examining each subplot we observe that for VLA-1 (upper left), there are very few data segments with low values, suggesting that almost all the data have extremals sometimes significantly larger than a Gaussian model. Both midwater channels almost overlap, but the deep channel has roughly one-sixth of its segments with kurtoses larger than 6, which is a very significant departure from Gaussianity. We can follow similar trends for the other VLAs (VLA-2 is the upper right); however, for the deepest of all the channels, VLA-6, the kurtoses now separate with depth with an increasing level as a function of depth. Finally, the mean and skewness of the data were close to zero for all the segments. The data acquisition equipment removes the mean because of a high pass filter at 10 Hz and the skewness near zero suggests no significant asymmetry in the sign of the data.

There is no evident explanation for this behavior *versus* depth. Again, the NPAL arrays and data recording have been used on many deployments and have been reliable. The data has a lot of scrutiny, so it is hard to assign the problem to the data acquisition equipment. *Conjectures* include the following: (i) Since the ''quiescent'' data were close to Gaussian, the sources for the other three categories may couple better to the deep sections of the arrays. (ii) ''Fish bites'' may be concentrated at the depth; however, the time series are not impulsive, which would suggest this. (iii) The line beneath the hydrophones may be exciting strumming, but the spectra on the deep sensors do not contain any more very low-frequency harmonic tonals characteristic of strumming.

IV. BROADBAND SPECTRAL COHERENCE

The covariance matrix across an array is the most important quantity in determining array gain and the ability of an array processing algorithm to reject noise interference. There are several models in the literature for the spectral covariance matrix, which describes the cross-spectra between two points separated by $\Delta\mathbf{x}$. Historically, the spectral covariance model for blackbody radiation, or three-dimensional (3D) isotropic noise, is often used, and this given by[30]

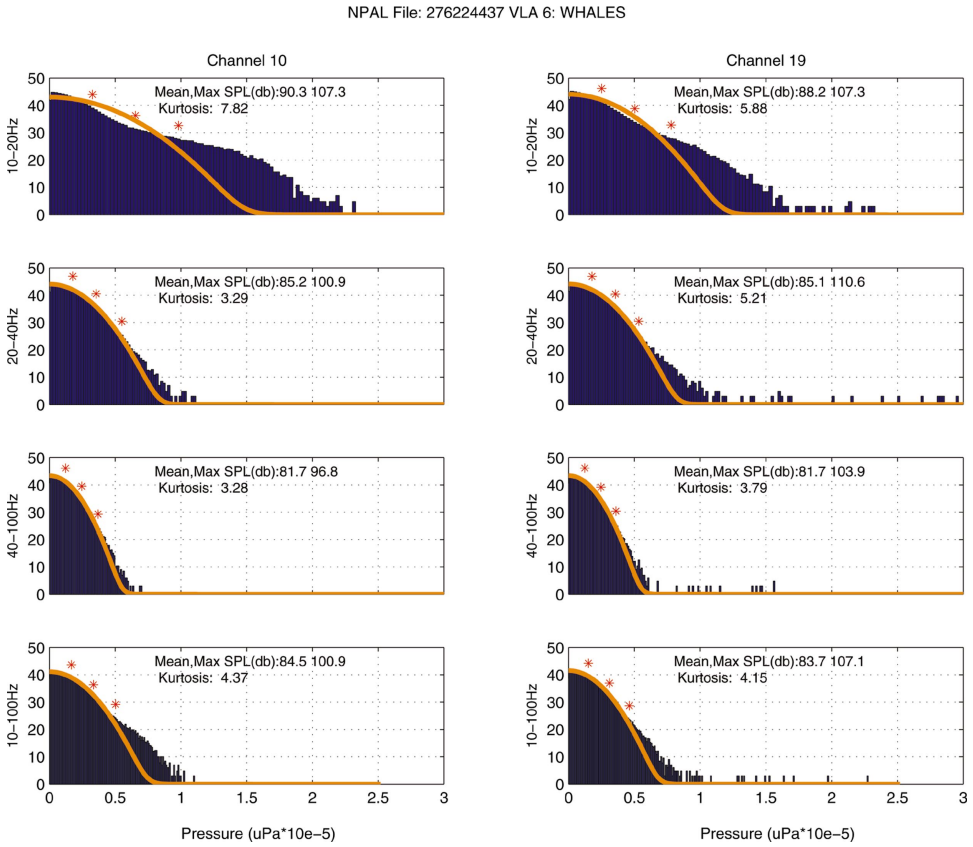


FIG. 9. Histogram of data on VLA-6, Channels 10 and 19 during the ''whale'' time of Fig. 4. Note the very non-Gaussian histogram in the vocalization band.

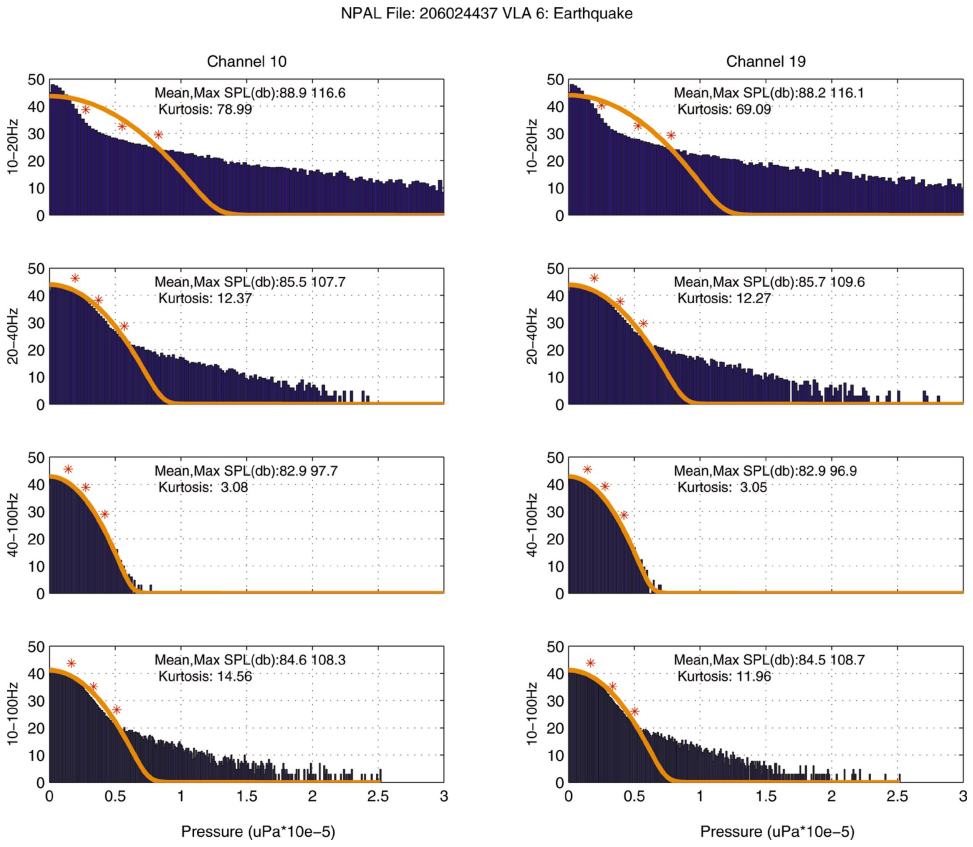


FIG. 10. Histogram of data on VLA-6, Channels 10 and 19 during the ''earthquake event'' of Fig. 5. Note the very non-Gaussian distribution in the lower-frequency bands occupied by the signal.

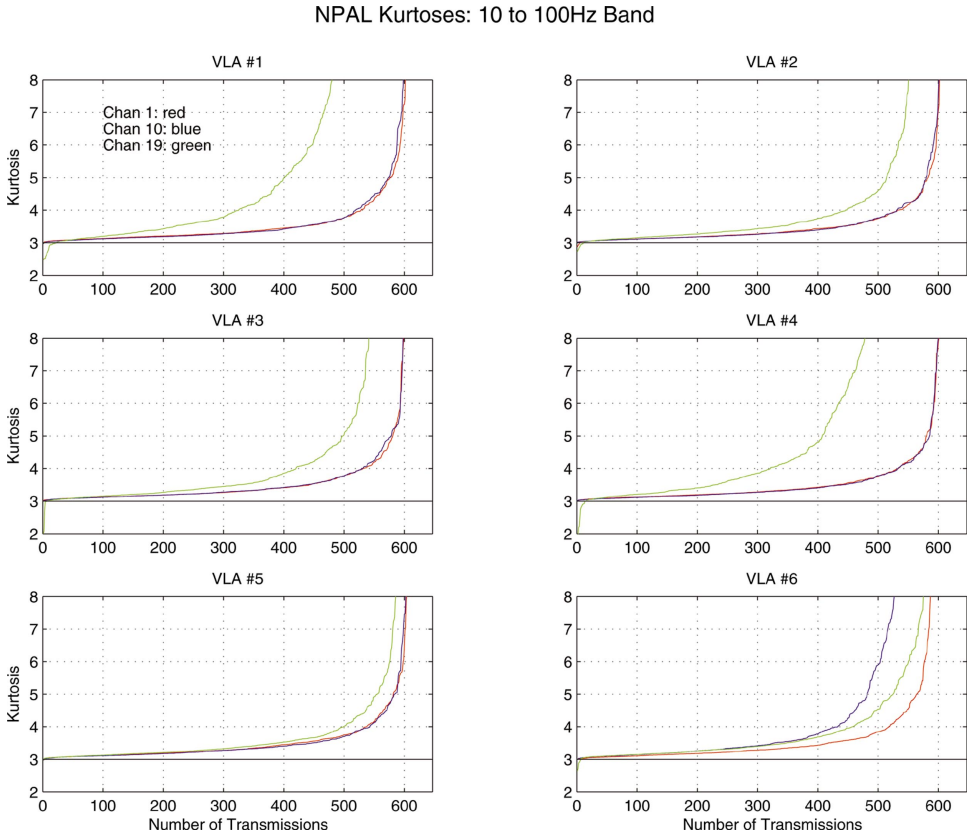


FIG. 11. The summary of kurtoses behavior on three channels of each of the VLAs. Note that the deepest channels on all arrays have higher levels. In addition, the two shallower channels on all arrays except VLA-6 overlap each other, whereas on VLA-6 they do not.

$$\mathbf{S}_n(f,\Delta\mathbf{x}) = \mathrm{sinc}\left(\frac{2\pi}{\lambda}|\Delta\mathbf{x}|\right), \tag{1}$$

where $\Delta\mathbf{x}$ is the separation, f the frequency, and λ the corresponding free space wavelength. Note the dependence on the absolute value of the separation; $\Delta\mathbf{x}$ implies the isotropic dependence. This model predicts the correlation to zero essentially among the VLAs, and a decaying oscillatory function over a scale of λ vertically on each array. This dependence can be seen in the quiescent segments as well as in spectral regions not populated by shipping or whales. Nevertheless, even in the quiescent segments array gain reveals a large horizontal extent of the covariance. Extensions to this model may be found in Cox[31], and Baggeroer[32]. (A more readily available version of this may be found in the recent text by Van Trees.[23] Both of these approaches specify an angular distribution on a sphere of radius $2\pi/\lambda$ and then uses spherical harmonic expansions for the correlation. Other expansions, e.g., Cron and Sherman, use a $\cos^n(\theta)$ model, where θ is the angle off grazing.[33,3] All the above are for deep water environments. Recently, the Kuperman–Ingenito model,[10] which can be used at all frequencies and water depths, has been used extensively. We should also note that spatial covariance functions for acoustic data are often similar to the sinc function since the propagating components of the spatial data are strictly bandlimited in wave numbers, i.e., $|\mathbf{k}|<2\pi/\lambda$.

V. SENSOR CROSS-CORRELATIONS

One of the objectives of the NPAL experiment is to determine the coherence properties for signals received from a NPAL source off Hawaii. In this paper we examine the coherence from the noise sources typically present near the NPAL array. There is a very practical concern for noise coherence matrices *versus* those for a signal. The maximum signal gain with perfect coherence is unity, so the signal issue in a signal-to-noise ratio (SNR) calculation is how closely we can achieve this. Generally, signal gain degradations, or coherence loss, cannot be less than -5 dB without significant focusing loss and spreading into other wave numbers, leaving a blur across all angles in a bearing-time recorder (BTR) display. The array gain indicates the reduction of the noise level and the noise coherence matrix determines how much gain can be achieved against a noise field. For noise uncorrelated among a sensor, this $10\log[N]$, where $[N]$ is the number of sensors; however, coherence among the sensors leads to higher array noise gains and signal-to-noise ratios that is of significant practical interest in predicting the sonar system performance. Since array gain calculations are complicated and more an issue for array processing, we refer to Refs. 23 and 34.

For the measurements below, Figures 13–16 indicate the analysis to measure the cross-coherences among the sensors for the four NPAL segments being discussed. We emphasize that the steps are the same as used in standard cross-correlation processing used for sonars:

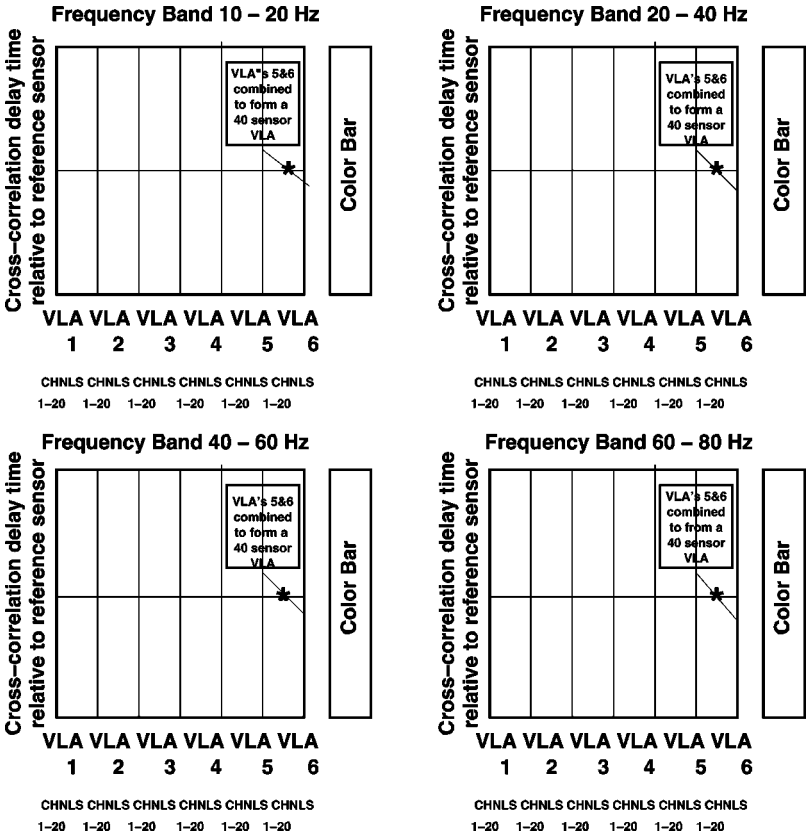


FIG. 12. Format for broadband cross-correlation among VLA sensor figures.

(1) Each channel was Fourier transformed with a Hanning tapered FFT with a duration of 200 s;
(2) The transforms at each frequency are cross-correlated for the elements of the $N \times N$ spectral cross-coherence matrix according to

$$\rho_{ij}(f:B) = \frac{X_i(f)X_j^*(f)}{\sqrt{\Sigma_{f\subset B}|X_i(f)|^2 \Sigma_{f\subset B}|X_j(f)|^2}},$$

where B is the frequency band processed.

Each NPAL data segment is 1200 s long, leading to 11 measurements when 50% overlapping is used, and these are averaged.

Each spectral cross-coherence element is inverse transformed back to the time domain, or $\rho_{(i,j)}(\tau:B)$, where τ is the lag, and the : B notation indicates that the result is specific to the band B. (There is a bit of duplicity in the use of ρ in both the frequency and delay domain; nevertheless, the symbol is the standard notation for a coherence value.) Array processing algorithms work in the frequency domain, but here the results are best viewed in the time domain where we can capture the correlation as a function of delay. This indicates the coherence among the various multipath components of the signal. Note that the denominator term simply guarantees unity coherence when $i=j$, i.e., the same sensor, with zero time lag.

In this section we want to capture the cross-coherence among all the sensors for a specific reference sensor, or one row of the matrix, or fixed i; moreover, we wish to capture this in a single figure so all the coherence on all the VLAs can be compared. Before considering each case it is useful to discuss the display format. The format is indicated in Fig. 12. The displays are the cross-coherence of sensor i with sensor $j=1, N$ organized by the VLA channel number on the horizontal axis and lag delay τ on the vertical. To interpret them we need to explain the NPAL data storage format. Each sensor is given a number, with the shallowest one on VLA-1 being number 1, and these are rasters lines 1–20 on the left with the coherence amplitude in dB encoded according to the color bar. The next 20 vertical rasters, 21–40, correspond to sensors 1 on VLA-2. The 41st raster starts with VLA-3 for the next 20 rasters. This continues until VLA-5, which starts with the 81st raster. The VLAs 5 and 6 represent the large 40 sensor array, so they have been aligned such that rasters 81–120 represent the 40 elements in this VLA array. A vertical grid of five red lines has been superimposed on the cross-coherence figures to aid in discerning the location of the coherence data among the six VLAs. All of the array data have been corrected for the accumulated clock offsets since they record time asynchronously. Since this is a coherence matrix, the peak level is unity (0 dB) on the reference sensor i for all the coherence plots.

We selected the reference channel to be the uppermost channel on VLA 6, at a mid array depth of 900 m and all delays are relative to this channel. This leads to the maximum horizontal separation for the coherence measurements. Since these are coherence plots, the color for the reference channel is 1 at zero delay is always red. There are roughly $c_w BT$ effective degrees of freedom, where again c_w is again the Hanning window factor, $T=100$ or 200, depending on the time duration and B is the bandwidth, 10 for the first band and 2 for the last three. The levels are carried to -25 dB to accommodate the wide dynamic range of the measurements.[35] [The degrees of freedom (dof's) can be

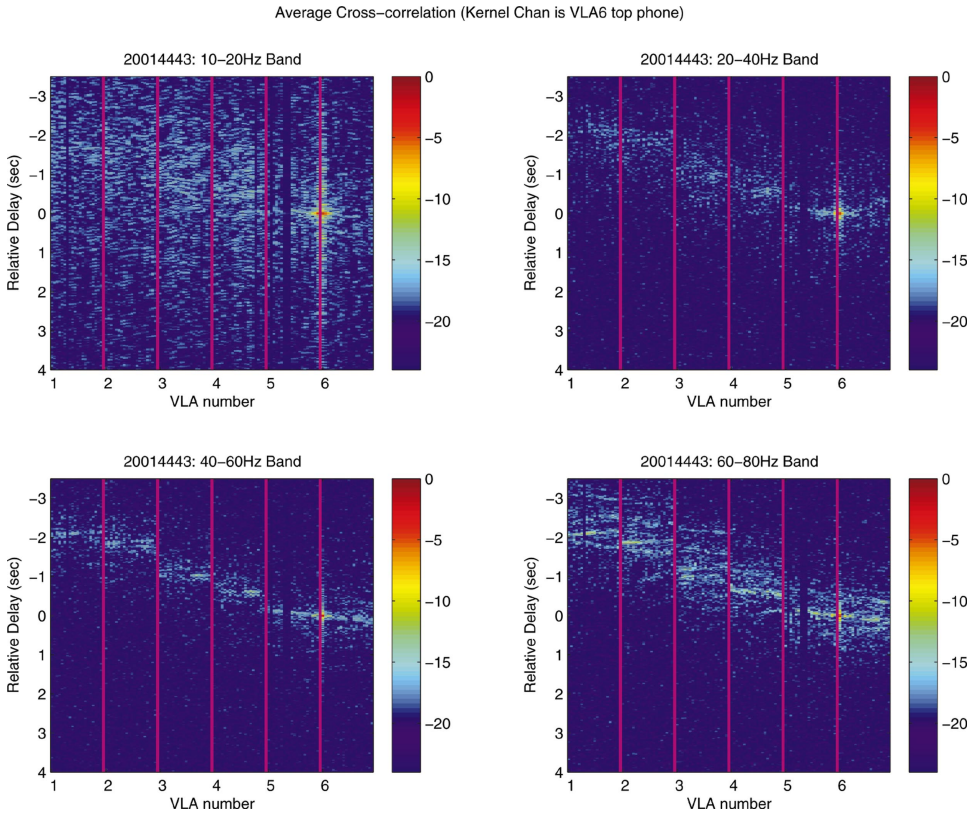


FIG. 13. Cross-correlations by the frequency band for the quiescent spectra of Fig. 2.

used to estimate the confidence intervals. Here the dof's range from 375 to 7500.]

A. Quiescent cross correlations

Figure 13 illustrates cross-correlation results for the ''quiet'' data of Fig. 2. The correlations are most evident as nearly horizontal ''stripes'' in each VLA band of the figure. The vertical red line indicating the division of the VLAs is adjacent to the raster of the reference channel, to the bottom sensor in VLA-6. If we ''expand'' this raster and the adjacent

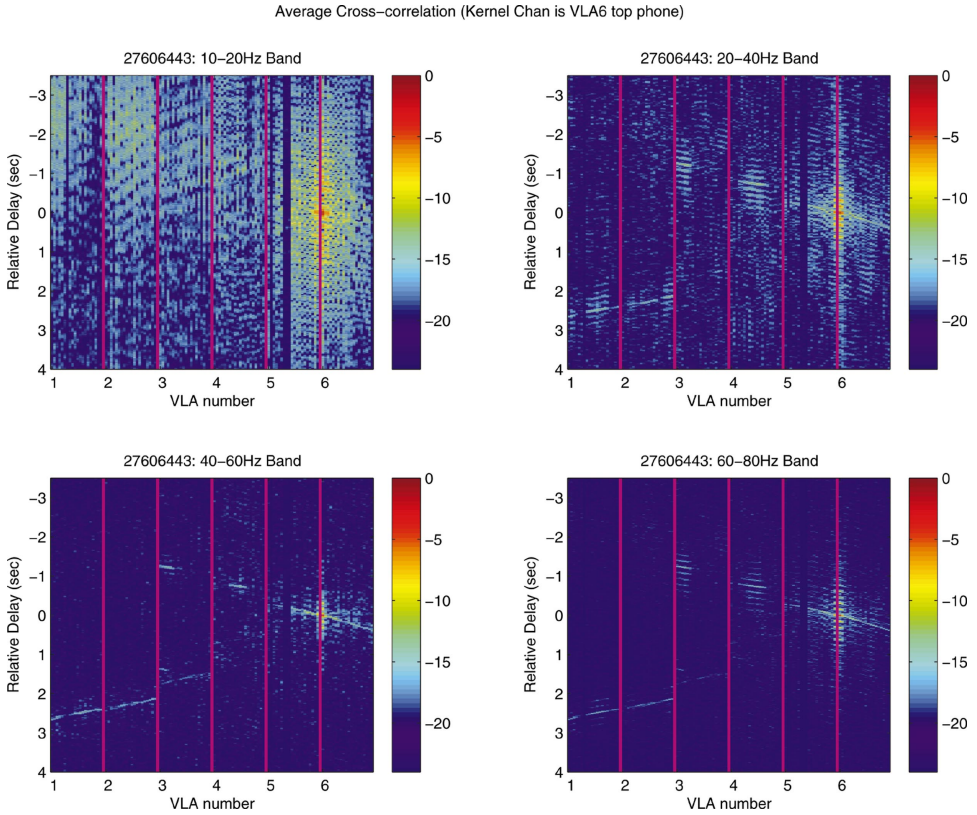


FIG. 14. Cross-correlations by frequency band for the nearby shipping of Fig. 3.

one to the left and right corresponding to VLA-5 with the sensors below and above the reference in the sensor in the 10–20 Hz frequency band (upper left), the result of convolving the "sinc$(2\pi f|\Delta \mathbf{x}|/c)$" spatial correlation with the effectively bandlimted white noise for the time–space coherence can be detected on these rasters. There should be significant coherence for approximately just three sensors each way horizontally, since the separation is 37.5 m and the wavelength at the center of the band (15 Hz) is 100 m. In the next band, 20–40 Hz (upper right) we can again observe this, but for fewer adjacent channels. In addition, we can observe a hint of coherence with the sensors on VLAs 1–4, suggesting some ducted signals can now be detected with the processing gain. This is an even stronger observation, low for the 40–60 Hz band (lower-left) and for the 60–80 Hz band (lower right). The coherence among some multipaths are discernible even though not visible on the original spectrum. The upward trend, which corresponds to negative delay or an advance on these panels, indicates that there is a discrete source north of the NPAL site with the signal advanced as it propagates from north to south. The cross-coherence levels across the VLAs are relatively low, less than -15 dB, so the background noise dominates the coherences observed.[36] (We discuss source localization subsequently with some stronger signals.)

B. Nearby ship noise cross-correlations

The cross-correlation diagram in Fig. 14 for the nearby ship spectra in Fig. 3 presents useful indications of the coherence of rays from such sources. First, the power spectral densities in Fig. 3 indicate that the ship passes VLA-1 first and VLA 5/6 last, so the cross-coherences with a reference sensor in the middle of VLA 5/6 should have negative delays (advances), as the signal has been "launched" prior to time for a window of the reference. The main horizontal stripe for VLA 5/6 indicates that the signal encounters the top part of the array first and then the bottom on a downward going ray. This is reasonable for a shallow source with deeper array. VLAs 1 and 2 also have signals with the opposite trend suggestive of a bottom bounce. This can be used for bottom bounce localization that we discuss later on localization. The coherences, nevertheless, are relatively low. This can be explained by Fig. 3, where the 200 s analysis window captures just one "loud" interval when passing VLAs 1–4 and 5/6, which indicates that it is passing very near the array. The level rapidly rises above and falls below the ambient noise so the ranges are close and on the steep level of the TL curve. Consequently, when we cross-correlate for the coherence, the data segments for VLAs 1–4 are below the ambient noise level for the 200 s analysis section; hence the low levels.

These shipping data example are more complicated to interpret. First, the oscillations at approximately 0.1 s are probably "DEMON"-like signals, where the propulsion of a ship is modulating a wideband signal at a frequency interval of 10 Hz seen on the spectra. The first panel (upper left) is complicated, reflecting the low level of power in the band. (There also seems to be a 15 Hz line from another source. The coherence on VLA 5/6 is strong. (There are standard "DEMON" programs used in the Navy for interpreting such shipping data with a lot of machinery tonals.) In interpreting the levels we observe that temporal correlation levels are relatively strong on the same and adjacent channels; however, they are modest, below -10 dB, on other channels. This reflects the fact that a 20 sensor VLA has a relatively low array gain for separating signals by vertical multipath. The NPAL array is also highly aliased horizontally at 1.2 Hz, again leading to low coherence estimates in the presence of other sources. Coherence measurements need to be done at high signal to noise ratios, especially if the level is high. While we did not have this in the upper band of 60–80 Hz, we still are able to see coherence across the full 3600 m aperture, or approximately for 140 wavelengths.

C. Whale vocalization cross correlations

Results illustrated in Fig. 15 are for the whale vocalization spectra in Fig. 4. It is one of the best data segments for demonstrating high coherence and well-defined multipath from the low-frequency band. Also, the sources are more distant than the ship and so lead to nearly constant TL to the VLAs. The coherence in the 10–20 Hz band (upper left) is very high, >-10 dB among the sensors for the VLA 5/6 array, and there is well-defined coherence across VLAs 1–4, the full extent of the NPAL array. The multipaths are well delineated, indicating both up and down paths at the arrays by the "x" pattern. The general trend toward positive delays indicates that the whale vocalizations are north of the array, arriving at the most southerly sensors early. In addition to the whale vocalizations, there is evidence of distant shipping from the opposite direction in the 40–60 Hz and 60–80 Hz bands, since the well-defined correlated, up/down multipath can be seen on both VLA 5/6 and VLAs 3 and 4.

D. Earthquake event cross-correlations

The cross-coherences for the earthquake event of Fig. 5 are given in Fig. 16. There is enough energy to dominate the signals and provide an estimate of the coherence limits across the array. We note that we used just the section containing the microearthquakes. We observe very high levels, >-7 dB across the full NPAL aperture. In addition, there are not up/down multipaths, indicated by low slopes for the array coherence, which suggest the near-axial propagation T wave propagation from the south. It is also a faint indication of a source north of the array in the middle two frequency bands also with near-axial propagation. It is a bit remarkable that there is mode coherence in the highest band with the same structure as the earthquake energy at the lowest; however, nearby seismic activity can have energy in this band.

VI. VERTICAL DIRECTIONALITY AND $k-\omega$ DIAGRAMS

We examine the vertical directionality by using $k-\omega$ diagrams where wave number k is the vertical wave number k_z and correspondingly are *vertical* phase speeds. (We actually use f, ν diagrams with ν the wave vector to eliminate annoying factors of 2π. This method for analyzing noise is well established for horizontal towed arrays, but less so for VLAs. This is an approximation because the vertical wave

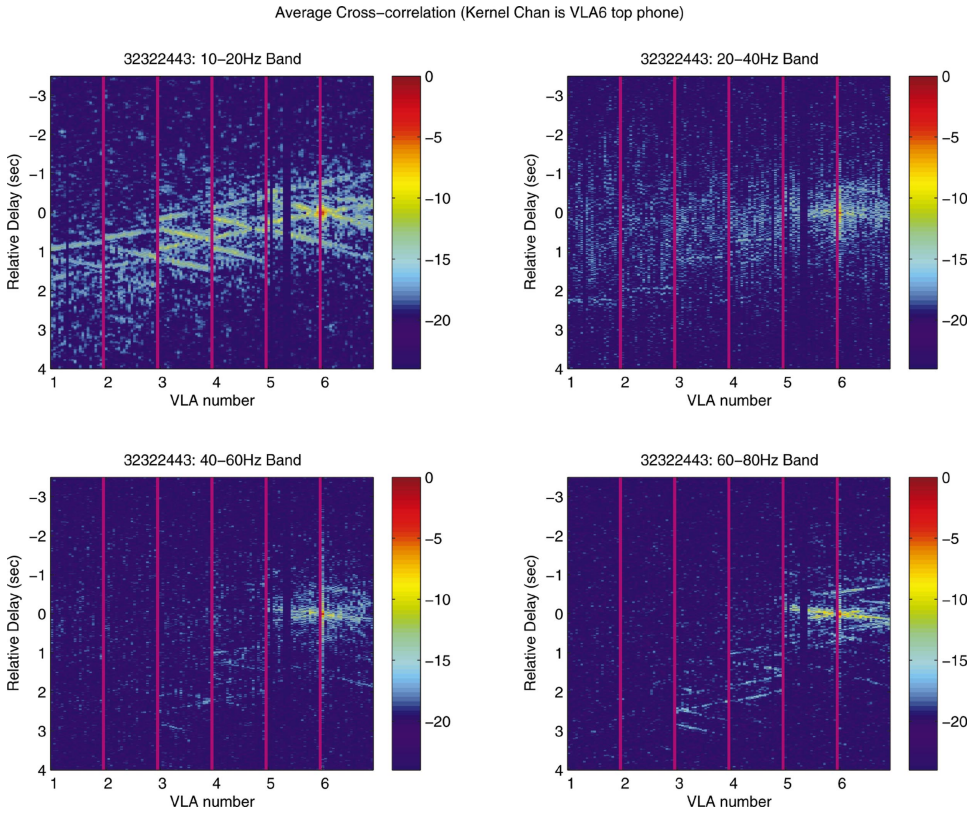


FIG. 15. Cross-correlations by the frequency band for the nearby whale vocalizations of Fig. 4.

number changes as a function of depth-specific frequency, horizontal ray parameter, and mode, but it does lead to easier interpretations instead of using an apparent phase speed across the array. Each VLA is aliased in the vertical at frequencies above 20 Hz. Unaliased broadband waveforms appear as ''stripes'' that converge at (0,0) for vertical slowness (cycles/meter) and frequency (Hz). Aliased components appear as replicas, offset from the unaliased stripe. This leads

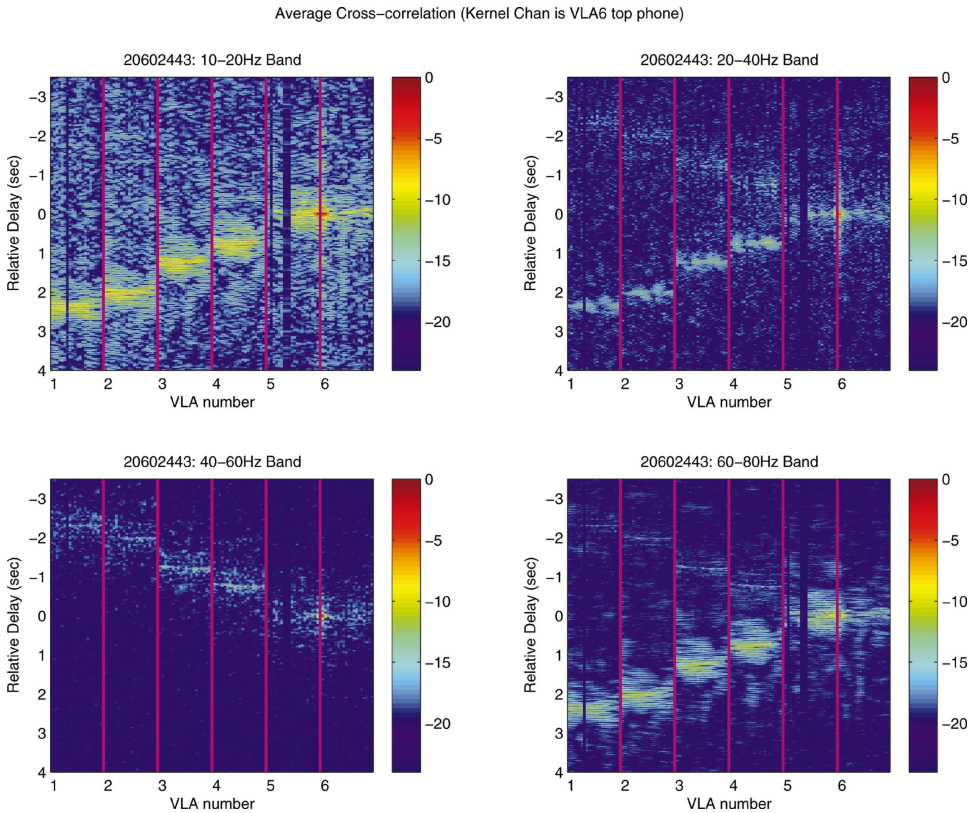


FIG. 16. Cross-correlations by the frequency band for the micro-earthquake of Fig. 5.

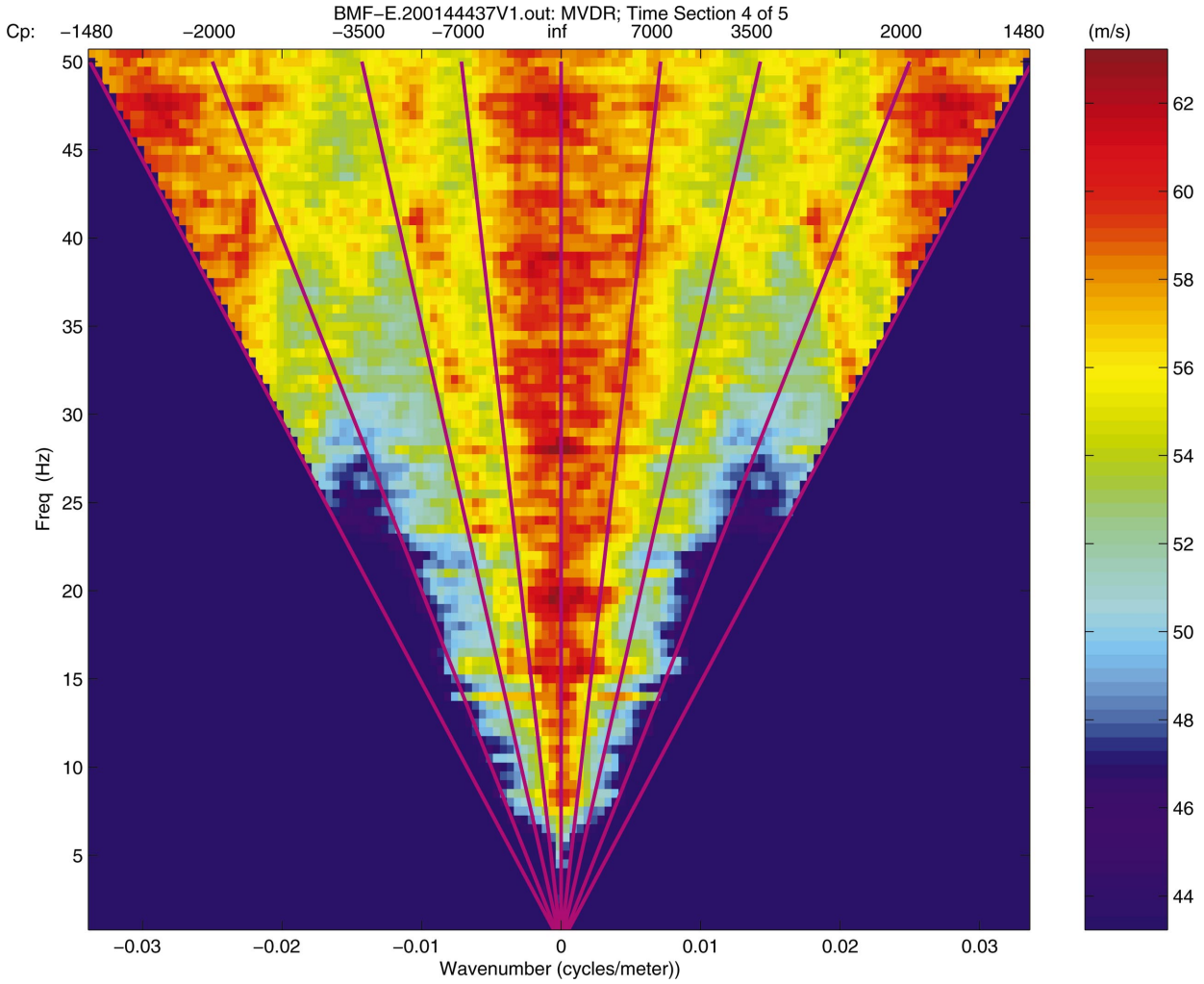


FIG. 17. Frequency wave number spectra for the quiescent spectra of Fig. 2. Note that in spite of the very low levels, there is enough array gain with the sensors (13 dB) to identify two weak shipping contributions at −5000 m/s and +3000 m/s, as well as the ducted noise.

to easier interpretations, especially when extrapolating signals into the aliased region, where we can observe the continuity along the slopes of lines corresponding to specific sources arriving at a vertical angle. The ''true'' contribution and the aliased ones are then very easy to sort. In viewing these diagrams, note that (i) the levels shift among the several plots in order to maximize the dynamic range of the displays, and (ii) the Minimum Variance Distortionless Filter (Capon) method was used for the spatial processing at each frequency.[23] We also note the discussion in Baggeroer and Cox regarding the level bias on these spectral estimates because of ''snapshot limits'' of the sample covariance matrices that are used[34] and that the MVDR is a local power estimate and not a true spectral density, i.e., power per Hz–vertical wave vector, ν_z. We defer the reader to the array processing literature for a complete discussion on MVDR.

There are also several diagonal lines superimposed to aid the conversion of ''stripes'' with vertical wave number and *vertical* phase speed diagrams to grazing angles. These are tabulated below:

± 1480 m/s $\pm 90°$ (vertical endfire directions),

± 2000 m/s $\pm 47°$,

± 3500 m/s $\pm 25°$,

± 7000 m/s $\pm 12°$,

∞ m/s $0°$ (broadside direction).

The general formula is given by

$$\Theta = 90 - \cos^{-1}\left(\frac{\text{slope } (d\text{Freq(Hz}/d\text{Wave number } (\nu_z))}{1480 \text{ m/s}}\right).$$

Lines with a positive slope correspond to power traveling down the array. The ducted power is concentrated in the acoustic cone to within 12° degrees of broadside. We note that these are equivalent to frequency-azimuth (FRAZ) displays, except that the azimuth is now the vertical.wave vector component.

A. $k-\omega$ directional spectra for quiescent data

Figure 17 illustrates the $k-\omega$ spectrum for the quiescent time section of Fig. 2. There is a slight asymmetry, suggesting a slightly wider contribution from ducted signals from below the array surface. There are also a couple of broadband low angle sources at $\approx -20°$ and $\approx 30°$, which is most

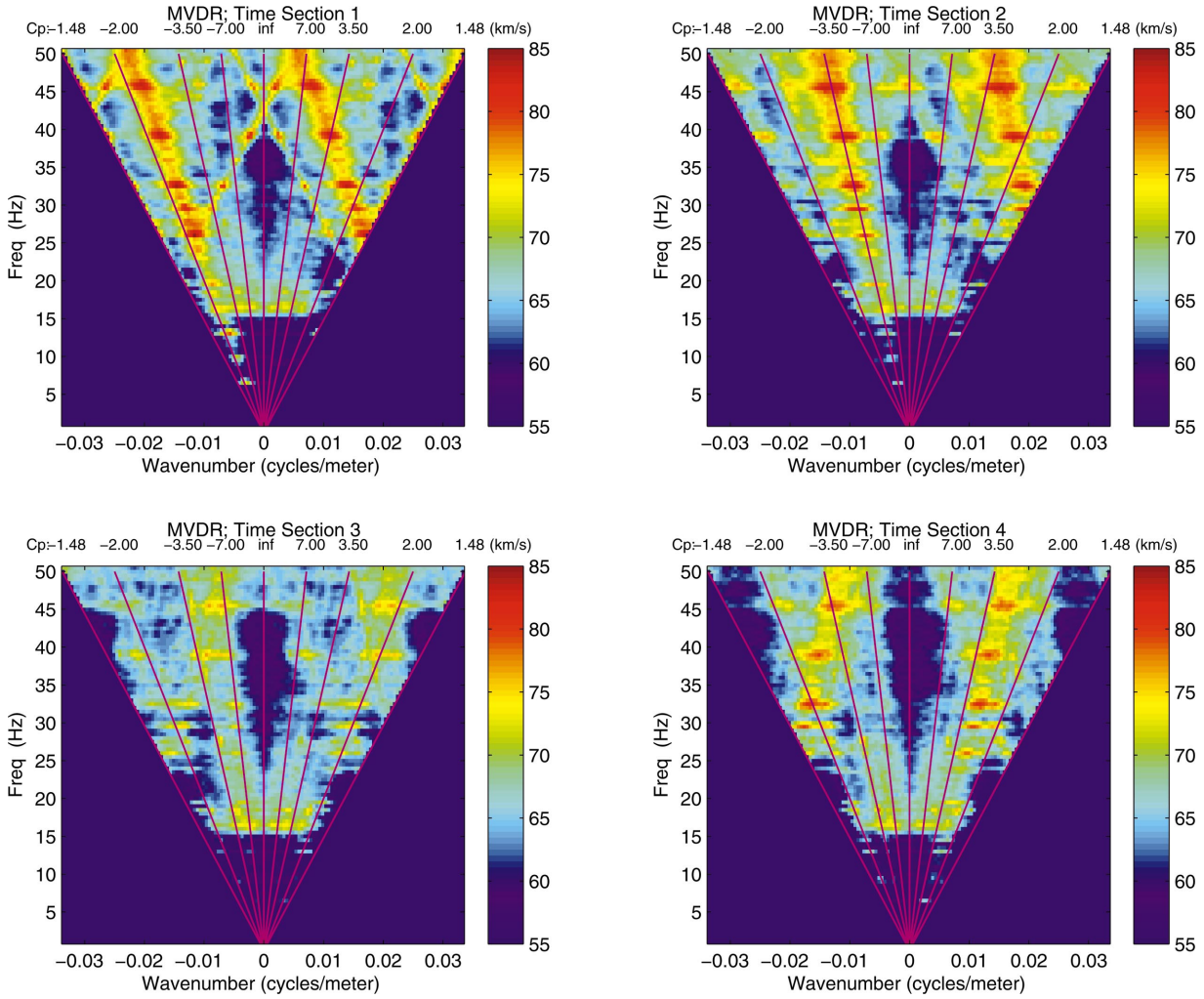


FIG. 18. Frequency wave number spectra for the shipping traffic interval. There are four intervals, each 200 s long. Note that there are at least two sources visible (see the text for discussion).

likely above the critical angle for the bottom, but still fairly distant. (This is complicated by the sloping bottom at the site.) There is also a suggestion of an all angle (completely within the acoustic cone) source of energy around 40 Hz that does not have an obvious interpretation. Finally, the high levels from 35–40 Hz at the large absolute wave number k values correspond to aliased power of the ducted components. Also, note the relatively low peak level of -63 dB on the color bar. This does not correspond to a spectral density level because of the known problems in the MVDR estimates.[30]

B. $k-\omega$ directional spectra for nearby shipping

Figure 18 is remarkable because it illustrates the nonstationary behavior of the nearby ship in Fig. 3. VLA-3 was used for the analysis. First, note that the levels are 20 dB higher than the quiescent levels, but still modest for shipping. There are four panels evolving in observation time, upper left to right on the top and then lower left to right on the bottom. In the first (upper left) there is a very large stripe at -2 km/s, or approximately downward at 45°. The dependence with respect to frequency peaks at the tonals observed in the spectral density. There is also little power below 25 Hz, as suggested by the spectral density. The aliased version can also be observed starting at 20 Hz, the aliasing frequency. There are also upgoing and downgoing signals and their aliases at ± 1480 Hz and above 25 Hz suggesting endfire noise from the surface then reflecting off the bottom with more than propagating downward (surface to depth). (The aliased components forming an "X" are easier to observe.) Note that the color maximum is 85 dB.

The next panel suggests the ship moving away leading to more horizontal angles consistent with Fig. 3; moreover, the multiples can be observed as they also move off vertical incidence; they are just inside the acoustic cone. (Recall these are vertical arrays, so endfire directions are at ± 1480 m/s. In the next panel (lower left) the direct path from the ship has disappeared most likely into a shadow zone. In the lower right, just the bottom bounce multiple (the apparent) phase speed is positive at 3 km/s. A ray path or parabolic equation solution using the sound speed profile in Fig. 1 supports these interpretations. There is a very faint suggestion of a second source closer to a ducted path (see the bottom two panels). Finally, Fig. 19 is a composite $k-\omega$ spectrum based upon the entire record in forming the sample covariance matrix for the MVDR. This indicates the blurring caused by the nonstationarity of the environment, which

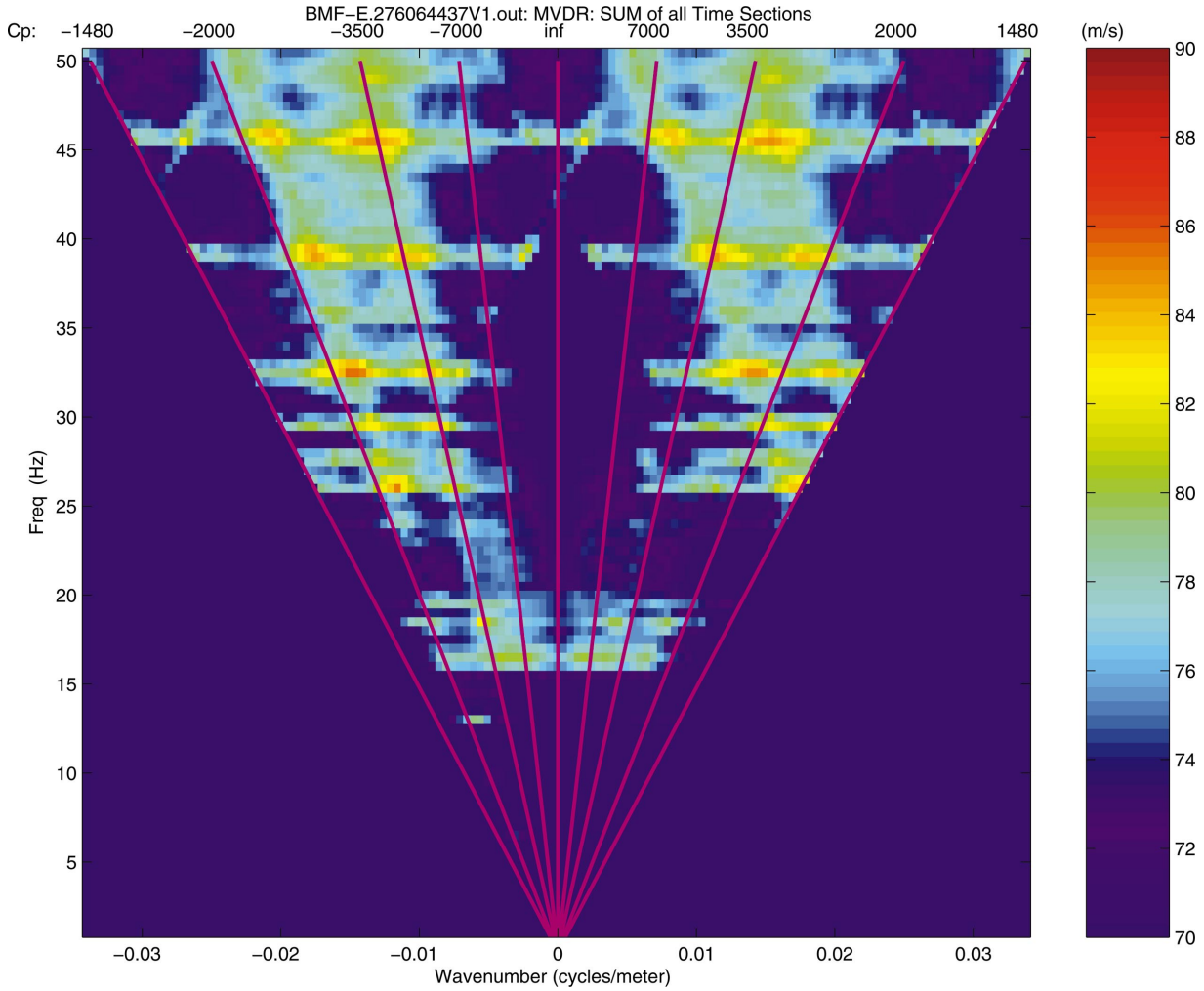


FIG. 19. The superposition of all the $k-\omega$ of five spectra of the passing ship. Note that the ray paths are more complicated to interpret because of the changing geometry.

leads to significant problems in ''snapshot-limited'' adaptive beamforming for passive sonars. Also note the presence of whale vocalizations around 16 Hz.

C. $k-\omega$ directional spectra for whale vocalization

Figure 20 illustrates the $k-\omega$ spectra for the whale vocalizations in Fig. 4 using VLA 5/6. The nominal peak levels are 78 dB—lower than a passing ship, but certainly not at the quiescent levels, and again complicated by the bias issues of the MVDR algorithm. The dominant contribution is concentrated around 16 Hz, as noted before. There are also harmonics at 32 and 48 Hz. There is also power at 39 Hz that is not a harmonic and possibly another mammal species. The notable aspect is that the ''stripes'' appear to be concentrated at phase speeds near 7 km/s, which are in the region ducted of $\pm 12°$ nominal propagation. There is virtually no power in either of the near vertical directions ± 1480 m/s, suggesting just long range vocalizations. (The array resolution and SVP knowledge did not permit identification or which ray loop for ranging.) We have not attempted to use ''invariant''-based approaches to date for ranging because of the uncertainty in the range rate.[37] For this data segment it is certainly appropriate to note that the vocalizations dominate the ambient noise, and this was very common in many of the NPAL segments. There is also ducted power at 12 Hz within the ducted region suggestive of some distant shipping.

D. $k-\omega$ directional spectra for a micro earthquake

Figure 21 indicates the evolution of the $k-\omega$ spectra for the two microseisms of Fig. 5. VLA 5/6 was used for the array $k-\omega$ analysis. Similar to Fig. 18, there are four panels proceeding from top left to right and then bottom left to right. The first upper left, corresponding to the first 200 s appears to be well-known ducted T phase propagation. The entire high-power level is concentrated in the SOFAR duct of ± 7 km/s vertical phase speed. There is very little other except at 28 Hz, suggesting shipping or mammal vocalizations. A few of these can be identified, especially examining the data after the earthquake, but they are difficult to interpret. The upper right panel still includes the micro-earthquake and has a similar format. After this, in the lower panels all the seismic activity has gone. Note that the earthquake bearing can be determined from Fig. 15 with propagation from south to north. Figure 22 is a composite $k-\omega$ spectrum of the entire 20 min record.

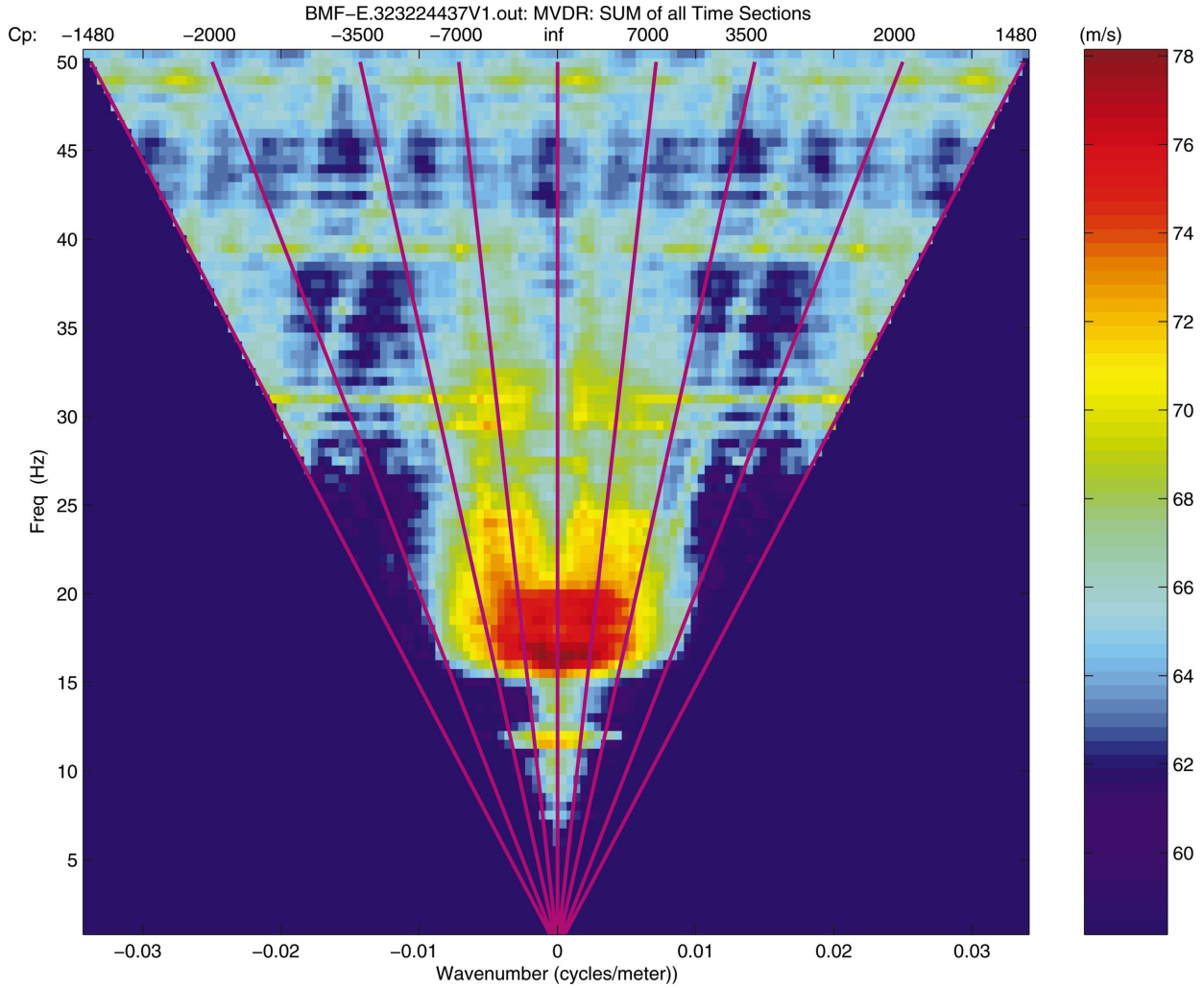


FIG. 20. Frequency wave number spectra for the whale spectra of Fig. 4. Note the strong level near 16 Hz and its harmonics plus other sources at 39 Hz, which is not a harmonic (another species?). There is also a strong ducted component around 12 Hz suggestive of distant shipping.

VII. SOURCE LOCALIZATION

One of the objectives of the noise analysis for the NPAL program is to determine how well sources can be located. Most of the effort concentrated on the Kauai source, whereas here we concentrate on sources of opportunity. Two methods were used to obtain estimates of the source locations of signals. The first method uses matched field processing (MFP) techniques on a narrow band of data from the long, 40 phone array, VLA 5/6. The second uses broadband correlations across the VLAs for the direction of arrival (DOA) and then models the crossing patterns of the multipath such as in Fig. 14, which might be considered a broadband MFP without precise phase matching. We note that there are many methods for source localization from the various submarine based methods to multipath arrival time separation to MFP methods ranked according to knowledge of the environment required.

A. Matched field processing approach

The MFP analysis uses just the one array, VLA 3, as before, because we could not obtain stable phase estimates of the cross-covariance matrix. This has been a common problem for MFP using multiple arrays reported in the Santa Barbara Channel Experiment,[38] so only the range of the source are considered unknowns. (The cause of this problem has been perplexing because the array localization is usually accurate to the one-tenth wavelength criteria; however, even small changes in the intersensor spacing over the long time needed for a stable estimate may be the cause since this smears the phase. Specifically, we use the ''nearby ship'' data file shown in Fig. 3 and we focused upon finding the position of the source of the 13 Hz ship line. We employed the OASN (OASES Ambient Noise) module from the OASES seismoacoustic modeling package[39] to generate replicas of the acoustic field in a 1 Hz band around 13 Hz for possible source positions on a grid ranging from the surface to 200 m in depth, and for horizontal ranges from 100 m to 10 km. The sound speed profile in Fig. 1 is used for the modeling. Sample covariance matrices are constructed from the data in the usual way by averaging the outer products of 10 s segments of data over 50 s intervals. Here 20 such matrices were formed over the first 1000 s of the file. The replicas were then combined with these sample covariances to form conventional (replica cross-correlation) and MVDR (see Sec. V for comments on MVDR processing) ambiguity functions.[40] The conventional processing does not indicate any consis-

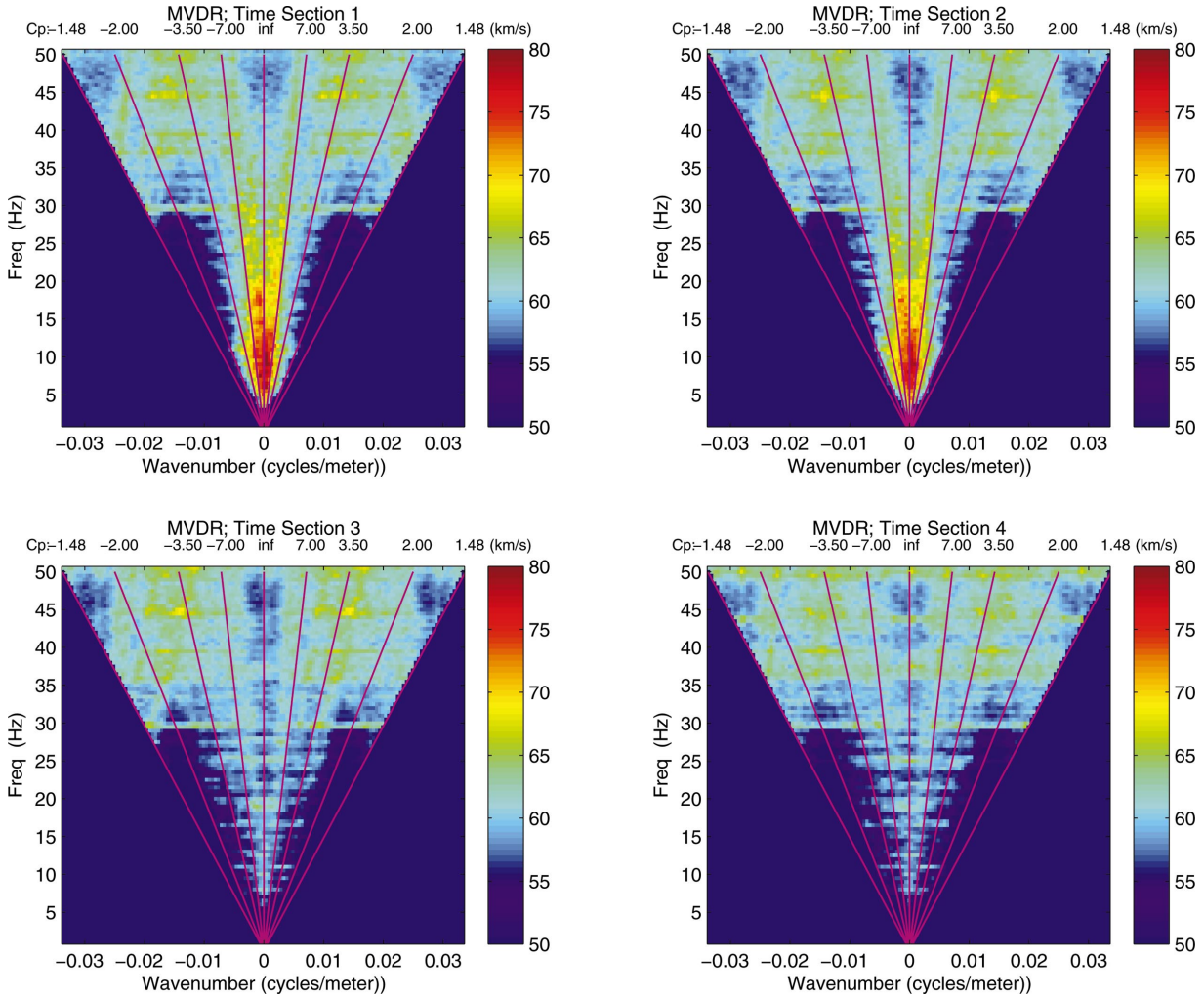


FIG. 21. Frequency wave number spectra for a transient generated by a micro-earthquake. Note that all the energy is ducted, typical of T phase propagation and extends up to 30 Hz. The source of the transient in Secs. II and III at 45 Hz is not known, but it is quite distinct.

tency, so we omit it. We show the results of the MVDR processing in a very compressed form in Fig. 23. Each horizontal stripe in the figure is the MFP output *versus* the range and depth for each processing interval of 50 s. The vertical dimension is depth 0–200 m from the surface and the horizontal axis is in the range up to 10 km. The processing interval time also goes from top to bottom, with the start time for each noted on the right. The entire set of horizontal stripes indicate the evolution of the processing intervals. At the beginning, or near the top, a maximum can be seen initially at about a 4 km range and near the surface. This arrival approaches the array until its range is less than 1.5 km at time 301 s, where it seems to fade. Sidelobes appear in the near-field, which is indicative of replica mismatch until 700 s, where the source reemerges again, focusing near the surface. It reaches a range near 4 km again at 951 s. Despite the lack of specific environmental data at the time this data was acquired, a reasonable estimate of the ship's range, except when it was close. Moreover, the total 8 km covered in the 1000 s leads to a speed of 16 kt, which is very close to the speed of 15 kt derived from the epoch times in Fig. 3. A simultaneous combination of MFP and geoinversion would most likely improve these results.[37] These ambiguity surfaces with their prominent sidelobes are fairly typical of MFP with field data.

B. Bearing and vertical path backpropagation localization

A second method obtains an estimate of the range and bearing for the marine mammal chirps that appeared in Fig. 4. It is similar to wave front curvature processing and vertical backpropagation used in ASW. We first compute the cross-correlation of all sensors at approximately equal depth on the first five arrays using the sensor on VLA 3 for a reference. This gives us a set of peaks in delay time for which we find the minimum mean square fit for range and bearing. These lead to a range of 6 ± 2 km, which is characteristic of wave front curvature processing and bearings of either $\approx$230° or 335°, with a slightly higher peak at the former. (The latter is due the NPAL arrays begin to appear nearly as a line array with a cone of uncertainty for sources distant from the center of the array.)

Using this information to narrow the search, we employed an eigenray program to compute more accurate travel

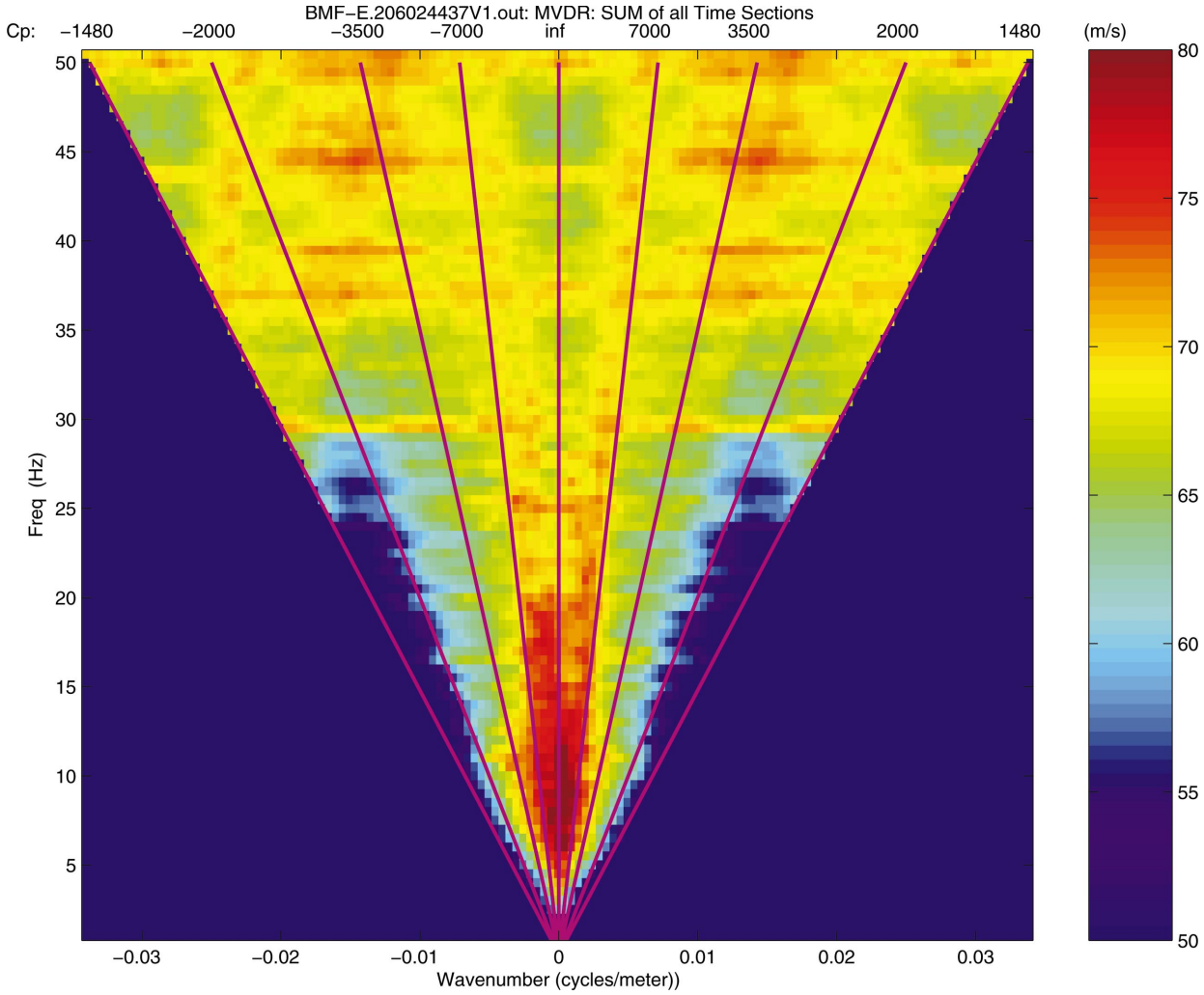


FIG. 22. Frequency wave number spectra for the entire segment (1200 s) of the micro-earthquake. Note that the ducted, low-frequency energy still dominates; however, there is also some high-frequency ducted energy as well as some omnidirectional energy at 30 Hz.

times using the average NPAL sound velocity profile for a series of ranges and directions near the previous approximate position estimates. For each source/bearing estimate, we generated a simple synthetic time series with impulses at eigenray arrival times, and then input these times series to the same correlation routine used to generate the images seen above in the format of Fig. 12. These again use the uppermost channel of VLA 6 as a reference. In Fig. 24 we display the best fit with a source estimated to be at a 5.8 km range, and a bearing of 230°. This technique is essentially a more complicated version of multipath ranging. It offers us a method to better understand the cross-correlation patterns for the NPAL data, and to match the synthetic patterns with the actual correlation pattern to get a best fit.

VIII. CONCLUSIONS

Ambient noise is usually categorized with rather coarse statistics, e.g. the average level, maximum values, and possibly variability. In fact, it is rich in many phenomena, so an observation at one time can change significantly in short times. In this paper we have concentrated on the low-frequency vertical coherence and directionality of the NPAL array deployed near Pt. Sur, CA. There are two important limits of the array design impacting our observations. First, the vertical spacing of the array elements sets an aliasing frequency of 20 Hz for vertical observations. Second, the array is aliased at 1.2 Hz in the horizontal, so very strong signals are needed to make unambiguous statements about ambient signals. (The NPAL source provides *a priori* information regarding signal direction, but absolute level measurements still require strong signals to sort out the sidelobe contamination issues.)

We make the following observations:

(i) The ambient environment at the NPAL array is dominated by shipping, whales and occasional earthquakes. Whales are present in the 16–18 Hz, and the 25–40 Hz bands well over sixty percent of the time observed. Quiescent spectra were observed only 18% on a single channel and 8% at a single VLA beam output using a signal entropy criterion.

(ii) The analysis of the kurtosis for a simple test of Gaussianity indicates that those spectral bands occupied by shipping, whales, or earthquakes are significantly non-Gaussian. Only when the levels are near quiescent are the kurtoses close to those for a Gaussian process.

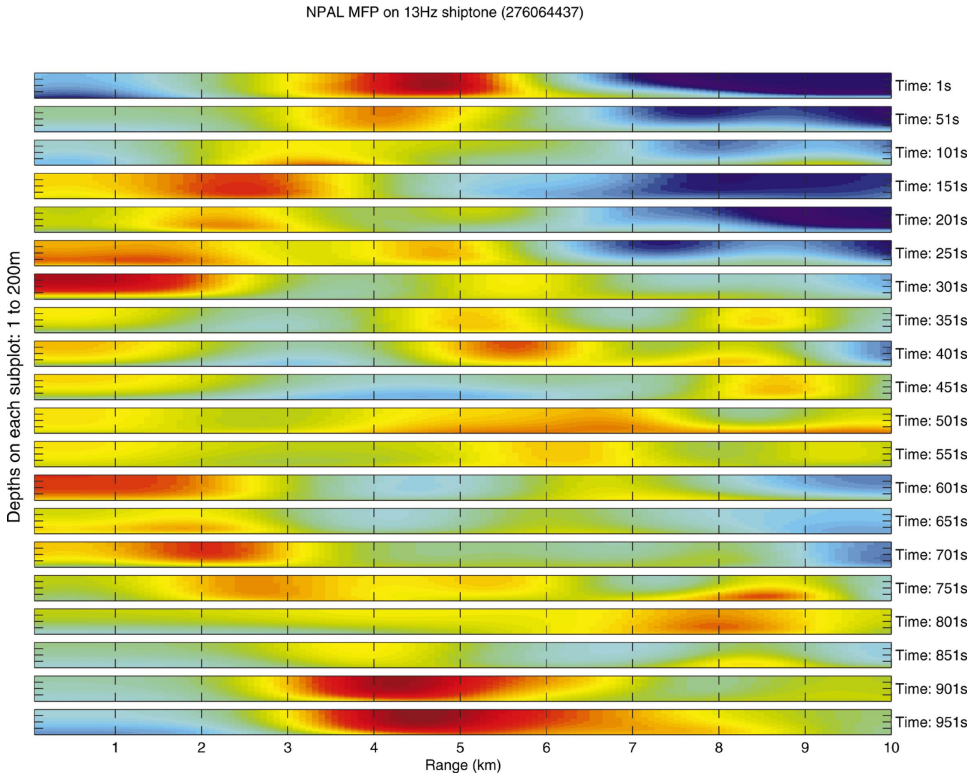


FIG. 23. Results of MFP Minimum Variance localization of 13 Hz tone from a ship passing near the NPAL arrays.

(iii) The degree of non-Gaussianity as measured by the kurtoses indicates that the lower hydrophones have higher levels. This was true on all arrays, but especially VLA 6, the deepest. The cause of this is not known, but the highpass filters in the acquisition system should eliminate cable strum that typically appears at low frequencies. In addition, the characteristic high Q and episodic nature of strumming does not appear to be present.

(iv) Standard broadband cross-coherence analysis reveals the correlated multipath for all categories of data. These extended horizontally across the full aperture in

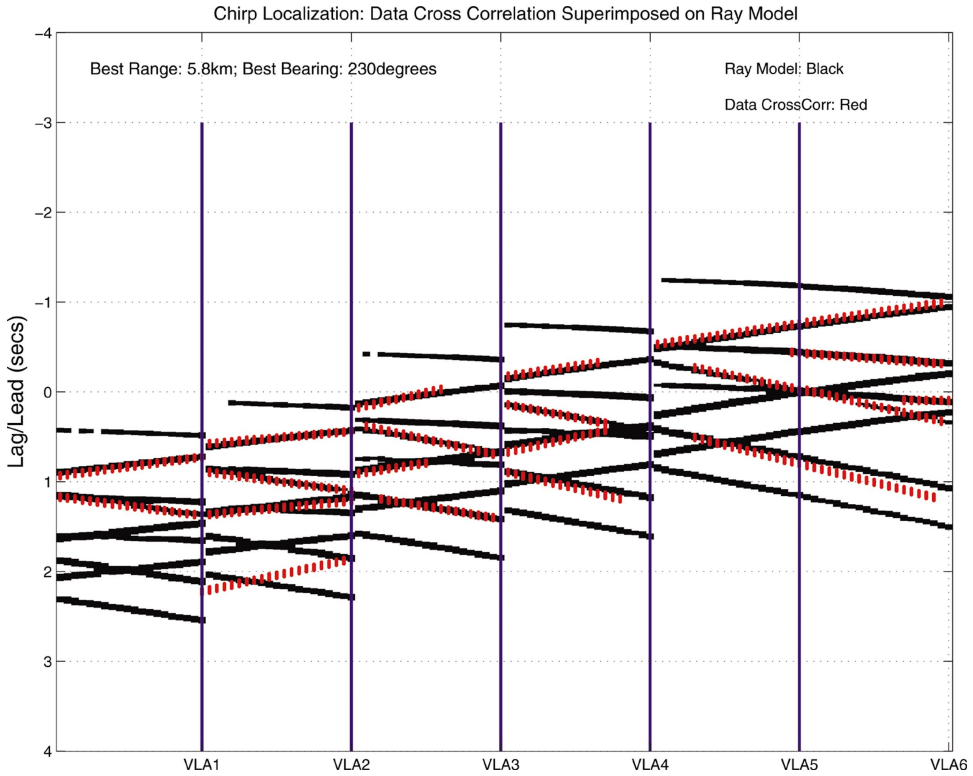


FIG. 24. Beat pattern match for chirp localization: Data (red) superimposed on the model (black).

most cases corresponding to 36, 72, 120, and 149 wavelengths for the four frequency bands analyzed.

(v) The $k-\omega$ spectra on the VLAs can be measured and indicate the vertical directionality associated with the several noise sources analyzed. By and large, most of the vertical spectra at these low frequencies are well defined with discrete lines, simplifying the interpretation in the aliased region.

(vi) Localization with matched field methods was problematic and could be done only at short ranges of less than 10 km. This is probably the result of sound speed mismatch to which matched field processing is well known to be very sensitive. Moreover, stable interarray cross-covariances could not be estimated. This was probably a combination of the long data durations required to satisfy the array transit time problem and small changes in the sensor spacing during a NPAL data segment. (This has been a problem in a number of MFP experiments.)

(vii) Localization by matching the broadband multipath structure was successful, which lends support to vertical ranging algorithms based upon this approach.

ACKNOWLEDGMENTS

This work was funded by the Ocean Modeling and Ocean Acoustics Programs of the Office of Naval Research and the Secretary of Navy/Chief of Naval Operations Chair for Ocean Science of the first author.

[1] K. B. Curtis, B. M. Howe, and J. A. Mercer, "Low-frequency ambient sound in the North Pacific: Long time series observations," J. Acoust. Soc. Am. **106**, 3189–3200 (1999).

[2] R. K. Andrew, B. M. Howe, J. A. Mercer, and M. A. Dzieciuh, "Ocean ambient sound: Comparing the 1960's with the 1990's for a receiver off the California coast," ARLO **3**, 65–72 (2002).

[3] R. J. Urick, *Principles of Underwater Sound for Engineers* (McGraw-Hill, New York, 1967); 2nd ed. (McGraw-Hill, New York, 1975).

[4] E. H. Axelrod, B. A. Schoomer, and W. A. Von Winkle, "Vertical directionality of noise in the deep ocean at a site near Bermuda," J. Acoust. Soc. Am. **37**, 77–83 (1965).

[5] Note that cross-spectra estimates across sensors can be a very subtle measurement since it is very easy to smear the phase estimates in frequency domain methods made popular by the Fast Fourier Transform. This is a very common error in forming sample spectral covariances for array processing algorithms.

[6] T. C. Yang, "A method of range and depth estimation by modal decomposition," J. Acoust. Soc. Am. **82**, 1736–1745 (1987).

[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**, 401–425 (1993).

[8] N. C. Makris, and I. Dyer, "Environmental correlates of pack ice noise," J. Acoust. Soc. Am. **79**, 1434–1440 (1986).

[9] A. J. Langley, "Acoustic emission from the Arctic ice sheet," J. Acoust. Soc. Am. **85**, 692–701 (1989).

[10] M. Dzieciuch, P. F. Worcester, and W. H. Munk, "Turning point filters: Analysis of sound propagation on a gyre-scale," J. Acoust. Soc. Am. **110**, 135–149 (2001).

[11] W. A. Kuperman and F. Ingenito, "Spatial correlation of surface generated noise in a stratified ocean," J. Acoust. Soc. Am. **67**, 1988–1996 (1990).

[12] W. M. Carey and E. C. Monahan, "Special Issue on Sea Surface-Generated Ambient Noise: 20–2000 Hz," IEEE J. Ocean. Eng. **15**, issue 4 (Oct. 1990).

[13] B. R. Kerman, *Sea Surface Sound* (Kluwer Academic, Boston, 1987).

[14] R. M. Kennedy and Teri V. Goodnow, "Measuring the vertical directional spectra caused by sea surface sound," IEEE J. Ocean. Eng. **90**, 299–310 (1990).

[15] W. M. Carey, R. B. Evans, J. A. Davis, and G. Botseas, "Deep-ocean vertical noise directionality," IEEE J. Ocean. Eng. **15**, 324–335 (1990).

[16] W. H. Hodgkiss, Jr. and F. Fisher, "Vertical directionality of ambient noise at 32° N as a function of longitude and wind speed," IEEE J. Ocean. Eng. **90**, 335–340 (1990).

[17] H. N. Oguz, "A theoretical study of low-frequency oceanic ambient noise," J. Acoust. Soc. Am. **95**, 1895–1912 (1994).

[18] T. C. Yang and K. Yoo, "Modeling the environmental influence on the vertical directionality of ambient noise in shallow water," J. Acoust. Soc. Am. **101**, 2541–2554 (1996).

[19] N. O. Booth, R. Judd, and H. Bucker, "Measurement of vertical noise directionality with a mixed polarity vertical array," IEEE J. Ocean. Eng. **28**, 537–543 (2003).

[20] R. G. Gallager, *Information Theory and Reliable Communications* (Wiley, New York, 1968).

[21] W. K. Pratt, *Digital Image Processing* (Wiley, New York, 1976).

[22] P. F. Worcester, R. C. Spindel, and the NPAL Group, "North Pacific Acoustic Laboratory," J. Acoust. Soc. Am. **117**, 1499–1510 (2005).

[23] H. L. Van Trees, *Optimum Array Processing, Part IV of Detection, Estimation and Modulation Theory* (Wiley-Interscience, New York, 2002), Chap. 5.

[24] Because of the NPAL array configuration, Sensor numbers 1, 10, and 19 on VLAs 1–4 have nearly equal depths of 400, 720, and 1030 m. VLAs 5 and 6 were configured vertically as a large aperture array. We refer to sensors 1, 10, and 19 on any VLA as "top," "middle," and "bottom." Consequently, depths for these sensors on VLA-5 were as follows: Channel 1—200 m, Channel 10—520 m, Channel 19—830 m. For VLA-6, depths were as follows: Channel 1—900 m, Channel 10—1200 m, Channel 19—1530 m. See Worcester and Spindel (Ref. 22) for details of the layout and sensor labeling.

[25] Note that the presence or absence of the NPAL signal had a very marginal impact on the noise levels because so much pulse compression gain and averaging are needed to raise its level above the ambient noise. Moreover, it was not present in the band below 60 Hz.

[26] A. B. Baggeroer, B. Sperry, K. Lashkari, C. S. Chiu, J. H. Miller, P. N. Mikhalevsky, and K. von der Heydt, "Vertical array reception of the Heard Island transmissions," J. Acoust. Soc. Am. **96**, 2395–2413 (1994).

[27] H. L. Van Trees, *Detection, Estimation and Modulation Theory, Part III* (Wiley, New York, 1971).

[28] The definition can vary, depending upon whether or not the "3" is subtracted (or "2" for complex data); e.g., see Press *et al.*

[29] J. L. Doob, *Stochastic Processes* (Wiley, New York, 1953).

[30] The sinc function is defined here by $\mathrm{sinc}(x)=\sin x/x$. Note that other definitions include a factor of π.

[31] H. Cox, "Spatial correlation of arbitrary noise fields with applications to ambient ocean noise," J. Acoust. Soc. Am. **54** (1973).

[32] A. B. Baggeroer, "Space–time processes and optimal array processing," Technical Report 506, Navy Undersea Center, San Diego, CA, December, 1976.

[33] B. F. Cron and C. H. Sherman, "Spatial correlation functions for various noise models," J. Acoust. Soc. Am. **34**, 1732–1736 (1962).

[34] A. B. Baggeroer and H. Cox, "Passive sonar limits upon nulling multiple moving ships with large aperture arrays," *Proceedings of the Asilomar 2000 Conference on Signals and Systems* (IEEE Press, Piscataway, NJ, 2000).

[35] There are four, and sometimes more, dark stripes on the displays. These correspond to "dead" or erratic sensors for the recording segment being used.

[36] The delay convention is the one normally used in signal processing. A signal with the form $y(t)=x(t-T_{\mathrm{delay}})$, where $T_{\mathrm{delay}}>0$ represents a replica of $x(t)$ occurring later in time.

[37] A. M. Thode, G. L. D'Spain, and W. A. Kuperman, "Matched field processing, source signature recovery and geoacoustic inversion of blue whale vocalizations," J. Acoust. Soc. Am. **107**, 1286–1300 (2000).

[38] P. N. Mikhalevsky (personal communication).

[39] H. Schmidt, *OASES User's Manual available at URL*: ⟨http://acoustics.mit.edu/faculty/henrik/oases.html⟩ Last viewed online 2/14/2005.

[40] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (American Institute of Physics, New York, 1994).

Assessing responses of humpback whales to North Pacific Acoustic Laboratory (NPAL) transmissions: Results of 2001–2003 aerial surveys north of Kauai

Joseph R. Mobley, Jr.[a)]
Social Sciences, University of Hawaii-West Oahu, 96-129 Ala Ike, Pearl City, Hawaii 96782

(Received 11 April 2004; revised 19 July 2004; accepted 20 July 2004)

Eight aerial surveys were flown north of the Hawaiian island of Kauai during 2001 when the North Pacific Acoustic Laboratory (NPAL) source was not transmitting, and during 2002 and 2003 when it was. All surveys were performed during the period of peak residency of humpback whales (Feb–Mar). During 2002 and 2003, surveys commenced immediately upon cessation of a 24-h cycle of transmissions. Numbers and distribution of whales observed within 40 km of the NPAL source during 2001 (source off) were compared with those observed during 2002 and 2003 (source on). A total of 75 sightings was noted during 2001, as compared with 81 and 55 during 2002 and 2003, respectively. Differences in sighting rates (sightings/km) across years were not statistically significant. Assessment of distributional changes relied upon comparisons of three measures: (a) location depths; (b) distance from the NPAL source; and (c) distance offshore. None of the distributional comparisons revealed statistically significant differences across years. Several possible interpretations are examined: (a) whales have habituated to the NPAL signal; (b) insufficient statistical power exists in the present design to detect any effects; and (c) the effects are short-lived and become undetectable shortly after the cessation of transmissions. [DOI: 10.1121/1.1854475]

PACS numbers: 43.80.Nd, 43.30.Pc, 43.60.Rw [AIT] Pages: 1666–1673

I. INTRODUCTION

The sensory capabilities of marine mammals have adapted over millions of years to exploit the enhanced properties of sound propagation in water. They rely on sound for a variety of functions including detection of predators, detection of prey, orientation, navigation, and communication with conspecifics (Tyack and Clark, 2000).

Considerable attention has been focused over the past decade on the potential for harm to marine mammals from exposure to human-made noise. The accumulated body of literature reveals a variety of responses, ranging from subtle changes in dive patterns (Frankel and Clark, 2000; Costa *et al.*, 2003) to clear avoidance responses (Malme *et al.*, 1983; 1984; 1985) to mass beachings and death (Frantzis, 1998; Balcomb and Claridge, 2001). The extent of effects varies depending on the type, frequency, and intensity of the signal, the duration of exposure, whether the source is fixed or moving, and the species involved, among other variables (see the review in Richardson *et al.*, 1995 and NRC, 2003). For baleen whales concern has focused mainly on human-made sources of low-frequency sound (LFS) (<1 kHz), since much of the sound production of these species falls in that range (Tyack and Clark, 2000). In contrast, odontocete species (toothed dolphins and whales), with the likely exception of sperm whales, are relatively insensitive to LFS, particularly below 100 Hz (Richardson *et al.*, 1995, p. 209; Au, Nachtigall, and Pawloski, 1997).

The Acoustic Thermometry of Ocean Climate (ATOC) program, which proposed to measure ocean temperature using LFS transmitters located north of the Hawaiian island of Kauai and near the Pioneer Seamount off the cost of California (Munk and Forbes, 1989; ARPA, 1995a, 1995b), included a marine mammal research program (MMRP) to determine what effects, if any, the acoustic transmissions might have on marine mammals and other marine life. For the Kauai source, the MMRP focused on endangered humpback whales as a target species since they: (a) have demonstrated sensitivities to LFS (Mobley, Herman and Frankel, 1988; Frankel, Mobley, and Herman, 1995); (b) are sufficiently numerous during the winter season to provide reasonable statistical power for experimental comparisons (Cerchio, 1998; Mobley, Bauer, and Herman, 1999); and (c) offer an extensive history of baseline data with well-established research methods.

For the Kauai ATOC MMRP, shore station observations during the first year of the ATOC source transmission (1998) showed no overt response to the signal for whales within 12 km of the source with received levels of 98–109 dB re: 1 μPa (Frankel and Clark, 2000). However, small increases in distances and times between surfacings were found which correlated with received levels of the sound. Similar responses were found in an earlier study off the coast of the island of Hawaii using scaled-amplitude playback of ATOC signals (Frankel and Clark, 1998), suggesting that these responses, although slight, were robust.

Analysis of whale distribution results from the Kauai shore station data showed no changes in distance offshore or depths of sighting locations between experimental (source on) and control (source off) conditions (Frankel and Clark, 2002). Similarly, there were no differences in sighting rates

[a)]Electronic mail: jmobley@hawaii.edu

TABLE I. Summary of 2001–2003 survey dates.

Survey no.	2001 Date	2002 Date	2003 Date	Survey no.	2001 Date	2002 Date	2003 Date
1	11 Feb.	2 Feb.	1 Feb.	5	11 Mar.	10 Mar.	21 Feb.
2	17 Feb.	6 Feb.	5 Feb.	6	17 Mar.	14 Mar.	5 Mar.
3	24 Feb.	18 Feb.	13 Feb.	7	24 Mar.	22 Mar.	17 Mar.
4	3 Mar.	2 Mar.	17 Feb.	8	31 Mar.	30 Mar.	29 Mar.

(whale pods/km^2) between experimental and control conditions, suggesting that whales were not leaving the area in response to the sound. However, there were slight but statistically significant increases in distance from the ATOC source while transmitting, as well as a slight eastward distributional change. Interpretation of these results was complicated by the fact that more whales were observed closer to the ATOC source as well as further away during transmissions relative to the control conditions. The authors noted that these short-term, small-scale behavioral changes were similar to those seen in the presence of vessels, and concluded that they were not likely of biological significance to the population.

Calambokidis (1998) performed aerial surveys of marine mammals in waters surrounding the ATOC source off the coast of California during both experimental (source on) and control (source off) conditions. He found subtle, but statistically significant, increases in distance from the source during experimental phases for both humpback and sperm whales.

The Kauai ATOC source ceased transmissions in October, 1999 and remained silent until January, 2002 when transmissions recommenced as part of the North Pacific Acoustic Laboratory program (NPAL). The mission of NPAL was to extend the earlier thermometry findings of ATOC over a longer time series to discern ocean-basin scale trends in temperature. Marine mammal monitoring under the NPAL program has focused on the use of aerial surveys to detect medium-to large-scale shifts in population abundance and/or distribution around the islands of Kauai and Niihau.

The present report summarizes the results of aerial sur-

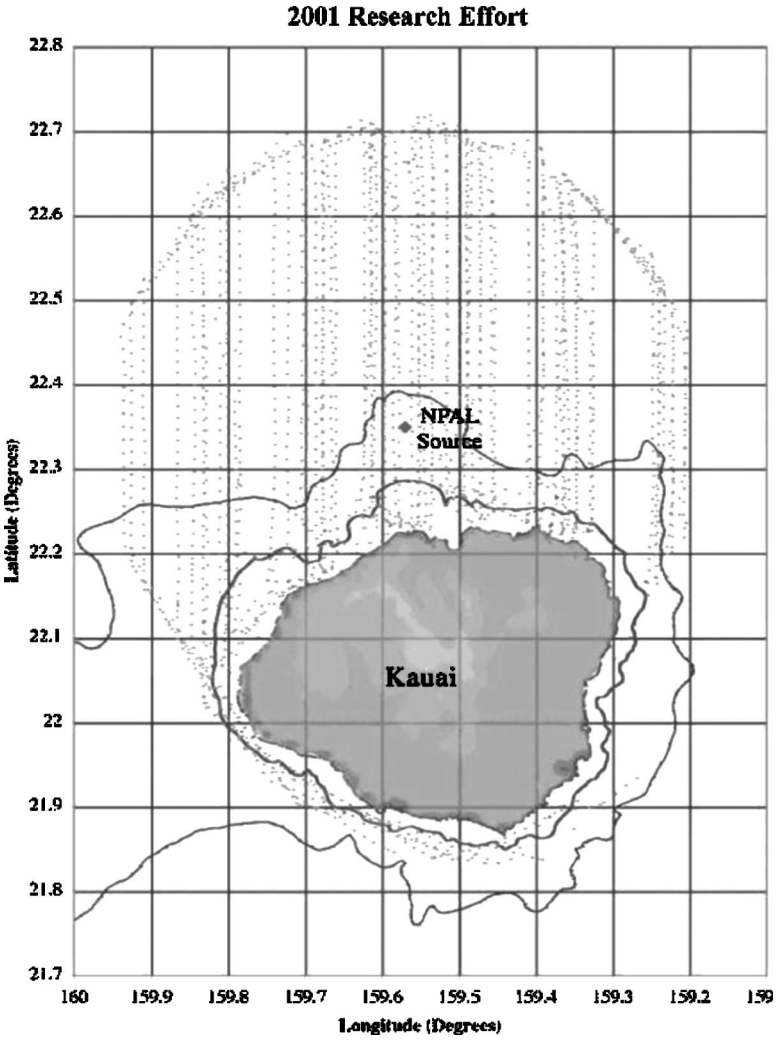


FIG. 1. Study area survey effort (dotted lines) for 2001–2003 aerial surveys (2001 shown here). For region north of Kauai, north–south tracklines were projected out to a 40-km radius from NPAL source (diamond). Random longitudinal startpoints were used so the exact trackline configuration varied on each flight. The same rules were used to generate tracklines so the effort was essentially equivalent across years. Inner and outer bathymetry lines refer to 100 and 1000 fathom contours, respectively.

veys during a baseline year (2001) when the NPAL source was not operating, as compared with those from 2002 and 2003 when it was. The mission was to identify and locate all marine mammals; however, only the results for humpback whales (representing over 80% of all sightings) are reported here. Further, this report focuses on sightings within 40 km of the NPAL source where any effects of exposure were more likely to be discernible.

II. METHODS

A total of eight weekly surveys of the waters adjoining the islands of Kauai and Niihau was performed during each study year (2001–2003) between 1 Feb. and 31 March (Table I). This period coincides with the time of peak residency of humpback whales based on the results of past surveys (Baker and Herman, 1981; Mobley, Bauer, and Herman, 1999; Mobley *et al.*, 2001). Methods were consistent across all 3 years. Survey protocol was based on distance sampling theory, which is the standard accepted approach for estimating abundance of free-ranging animal populations (Buckland *et al.*, 2001).

During the 2002 and 2003 seasons, all surveys were scheduled to coincide with NPAL transmission days. The NPAL transmissions occurred every fourth day (starting from 1 Jan.) with six 20-min transmissions per day spaced 4 hours apart. The 20-min transmission included an initial 5-min ramp-up to full power (broadband source level of 195 dB re: 1 μPa). We commenced surveying the north shore of Kauai immediately following a 24-h transmission cycle (ca. 1000 h). This was done to maximize the potential for detecting any possible distribution changes in response to the transmissions.

Surveys followed predetermined north–south tracklines spaced 13 km apart within a 40-km radius of the NPAL source and 26 km apart throughout the remainder of the area surveyed (Fig. 1). For the area north of Kauai, one or two additional lines spaced 6.5 km apart were added in the immediate vicinity of the NPAL source to permit greater sampling effort in that area. Starting longitudes were randomly chosen based on distance sampling methodology (Buckland *et al.*, 2001) so that the exact trackline configuration varied for each survey.

The survey aircraft was a twin-engine Partenavia (P68) Observer flying at a speed of 100 knots and an average altitude of 244 m. Two experienced observers made sightings of all marine mammal species, one on each side of the aircraft. Sightings were called to a data recorder who noted the species sighted, number of individuals, presence or absence of a calf, angle to the sighting (using hand-held Suunto clinometers), and any apparent reaction to the aircraft. Additionally, GPS locations and altitude were automatically recorded onto a laptop computer at 30-s intervals, as well as manually whenever a sighting was made. Environmental data (seastate, glare, and visibility) were manually recorded at the start of each transect leg and whenever conditions changed. The two data sources (manual and computer) were later merged into a single data file.

TABLE II. Summary of whale sightings and effort.

Year	Survey no.	No. sightings	Effort (km)	Sighting rate (sightings/km)
2001	1	22	399.7	0.055
	2	8	544.3	0.015
	3	3	509.0	0.006
	4	6	510.0	0.012
	5	4	521.3	0.008
	6	23	508.8	0.045
	7	7	483.5	0.014
	8	2	480.6	0.004
	Totals:	75	3957.2	0.019
2002	1	7	445.0	0.016
	2	9	496.0	0.018
	3	13	479.6	0.027
	4	0	465.5	0.000
	5	12	535.8	0.022
	6	20	542.7	0.037
	7	6	503.2	0.012
	8	14	489.3	0.029
	Totals:	81	3957.1	0.020
2003	1	9	460.5	0.020
	2	11	296.3	0.037
	3	2	388.5	0.005
	4	5	443.5	0.011
	5	8	466.8	0.017
	6	15	284.0	0.053
	7	2	503.1	0.004
	8	3	487.0	0.006
	Totals:	55	3329.7	0.017

Sighting positions were calculated trigonometrically based on the position of the plane at time of sighting as derived from GPS data, along with the altitude and angle to the sighting. All sightings included here were within 5 km of the plane's position, with 87% at 3 km or less. Although survey effort included coastal waters surrounding the islands of Kauai and Niihau (Fig. 1), only sightings within 40 km of the NPAL source are summarized here.

For depth information, sighting positions were translated into GIS format, and correlated with depth data from NOAA GEODAS dataset (representing depth soundings from ship surveys 1953 to 2003). Average distance between whale positions and depth soundings was 34.1 m.

III. RESULTS

A total of 211 humpback pods, including 366 individual whales (average pod size=1.7 whales), was observed within the study area across the 3 years surveyed involving a total linear effort of 11244 km (Table II). Peak sightings tended to occur between 5–17 March (survey #6), similar to previous reports for the Hawaiian Islands (Herman and Antinoja, 1977; Baker and Herman, 1981; Bauer, and Herman, 1999).

As is the case throughout the Hawaiian wintering grounds (Herman and Antinoja, 1977; Mobley *et al.*, 2001), the majority of humpbacks was observed in the shallow shelves surrounding the island (Fig. 2). Although depths of sighting locations varied widely (range: 28–3936 m) there was a clear preference for shallower water with a modal depth of 0–50 m (Fig. 3).

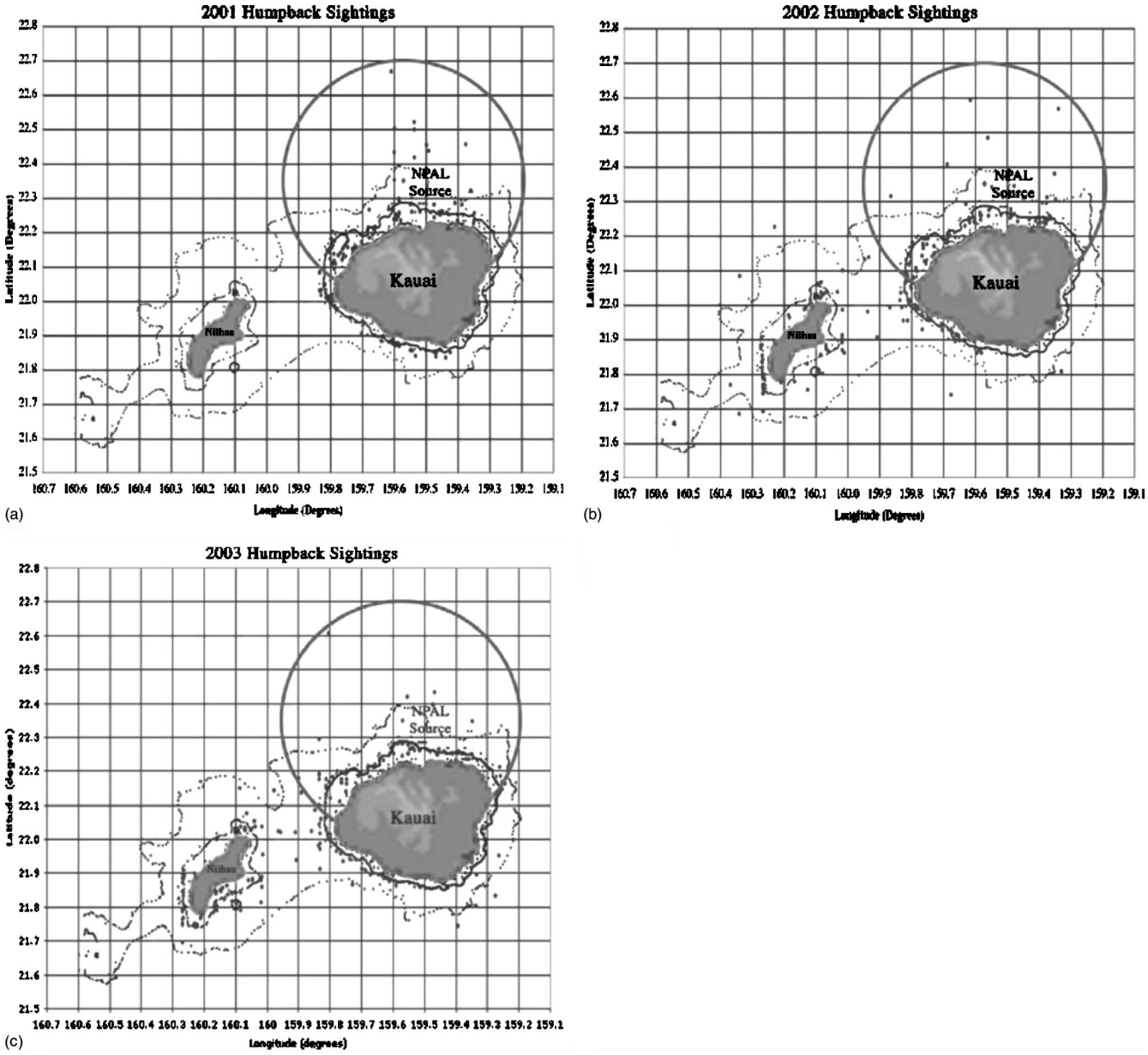


FIG. 2. Humpback whale sightings during 2001 (source off) and 2002–2003 (source on). Sighting locations are indicated by dots. Analyses described here were limited to 40-km radius (semicircle) around NPAL source (diamond). Inner and outer bathymetry lines refer to 100 and 1000 fathom contours, respectively. As shown, humpbacks generally preferred waters within 100 fathoms across all 3 years surveyed.

A. Effects of seastate

Differences in seastate affect sighting probabilities when surveying marine mammals (Buckland *et al.*, 2001). In the present study, sighting probability decreased substantially beyond a Beaufort seastate of 2 (Fig. 4). The distribution of seastates varied significantly across years, $\chi^2(10) = 49.75$, $p < 0.05$ (Fig. 5). Specifically, there was a greater relative incidence of higher seastates (i.e., Beaufort 3 and greater)

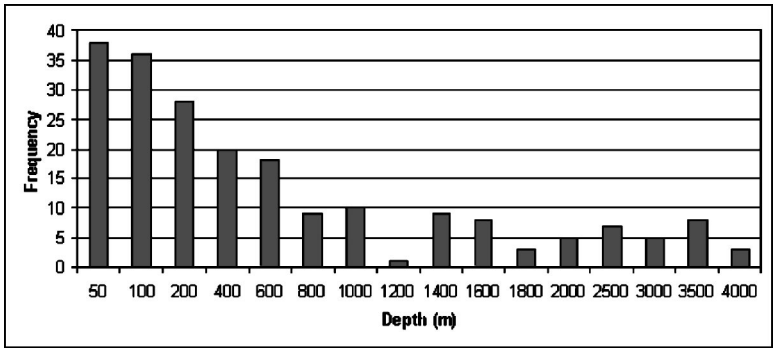


FIG. 3. Depths of sighting locations—all sightings combined. Depths of sighting locations varied widely; however, there was a clear preference for shallower depths (Note: This was not an artifact of effort since most effort occurred in depths greater than 1000 m.)

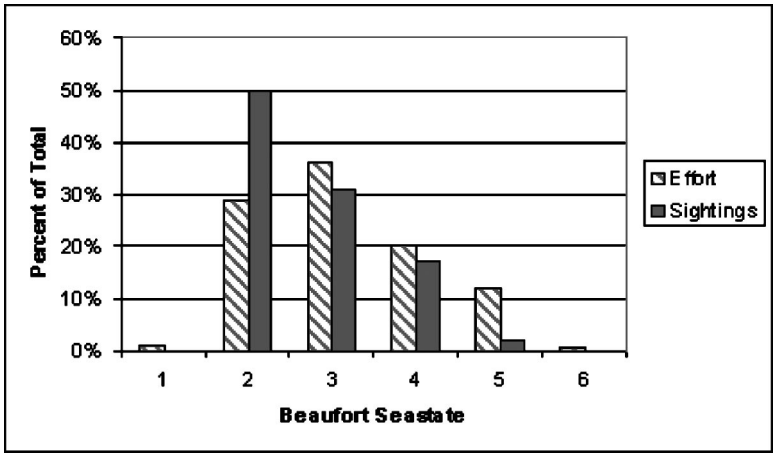


FIG. 4. Sightings and effort by seastate (all years combined). Seastate influenced sighting probability. Probability of detection decreased substantially beyond a seastate of 2.

during 2003, and a greater incidence of lower seastates during 2002. Thus, sighting conditions were most favorable during 2002 and least favorable during 2003.

B. Relative abundance

A potential effect of the NPAL transmission could be the displacement of whales from the study area (waters within 40 km of the NPAL source). To compare relative abundance across years, numbers of sightings per survey were normalized by linear effort to produce sighting rates (sightings/km) (Table II). The highest and lowest sighting rates for 2002 and 2003, respectively, correspond with the seastate differences for those years (Fig. 5). However, a comparison of sighting rates across years showed the differences to be nonsignificant (Kruskal–Wallis $KW_{3,24}=0.64$, $p>0.05$).

C. Distribution effects

Another potential effect of the NPAL transmissions could be changes in normal patterns of distribution from the baseline year (2001) to those with the NPAL source on (2002 and 2003). In order to test for such changes, three measures of distributional change were assessed: changes in depth, changes in distance from the source, and distance from shore.

1. Depth differences

The median depth of sighting locations was the same for 2001 (source off) and 2002 (source on), but was different for 2003 (source on) (Table III). When all 3 years were compared, the difference in median depths was not statistically significant ($KW_{3,211}=1.00$,$p>0.05$). Regarding the distribution of depths (Fig. 6), paired source-on vs source-off comparisons revealed no statistically significant differences (2001–2002: Kolmogorov–Smirnov $K–S_{75,81}=0.08$, $p>0.05$; 2001–2003: $KS_{75,55}=0.18$, $p>0.05$).

2. Distance from source

Based on past studies using the ATOC source (Calambokidis, 1998; Frankel and Clark, 2002), the most likely response to NPAL transmission exposure was an increase in distance from the source itself. The median distances from the source were greater when the source was operating (2002 and 2003) relative to when it was not (2001) (Table IV). However, a comparison of distances across all 3 years showed the difference to be statistically nonsignificant ($KW_{3,211}=3.11$, $p>0.05$). When the distribution of distances from the source is compared (Fig. 7), more whales were in the furthest distance category when the source was operating (2002 and 2003) than when it was not (2001). However, paired source-on vs source-off comparisons revealed no statistically significant differences in the distribution of distances (2001–2002: $K\text{-}S_{75,81}=0.17$, $p>0.05$; 2001–2003: $K\text{-}S_{75,55}=0.18$, $p>0.05$).

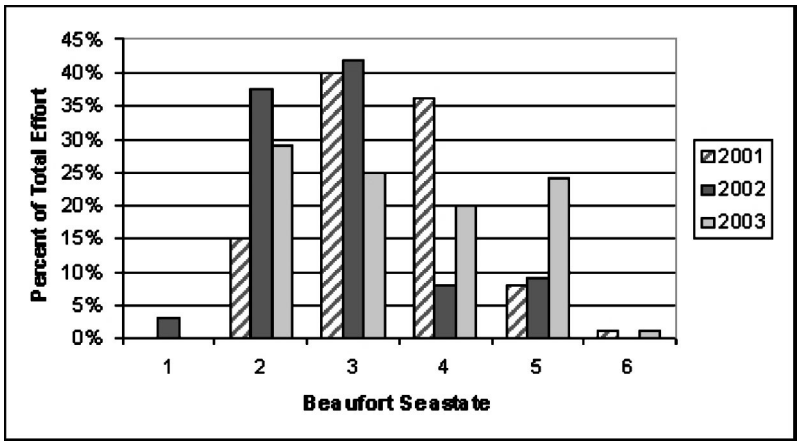


FIG. 5. Seastate by year. The distribution of seastates varied significantly across years. Lower seastates and thus, best sighting conditions, prevailed during the 2002 season. The greater prevalence of higher seastates during 2003 implies that sighting conditions were least favorable during that year.

TABLE III. Analysis of depth differences.

Year	Depth of sighting locations (m): *N*	Median	SD	KW
2001	75	195	1011	1.00
2002	81	195	928	
2003	55	352	923	

TABLE IV. Analysis of distances.

Year	Distance from source (km): *N*	Median	SD	KW	Year	Distance from shore (km): *N*	Median	SD	KW
2001	75	21.13	9.06	3.11	2001	75	4.58	9.28	1.65
2002	81	25.17	10.40		2002	81	5.21	7.74	
2003	55	26.85	10.55		2003	55	5.18	8.71	

3. Distance from shore

It is possible that exposure to the NPAL transmissions may cause the whales to move further or closer to shore. Median distances for the transmission years (2002, 2003) were somewhat further offshore than the median distance for 2001 (Table IV). When offshore distances were compared across years, the differences were not statistically significant ($KS_{3,211}=1.65$, $p>0.05$). When the distribution of offshore distances is compared, whales were found in the midrange distance categories more often during the transmission years (Fig. 8). However, paired source-on vs source-off comparisons showed no statistically significant differences in the distribution of offshore distances (2001–2002: $K\text{-}S_{75,81}=0.16$, $p>0.05$; 2001–2003: $K\text{-}S_{75,55}=0.20$, $p>0.05$).

D. Statistical power analysis

Particularly when reporting negative results, one must consider the issue of statistical power, i.e., the likelihood of detecting an effect if one exists. Ideally, statistical power should approach 0.80, which implies an 80% likelihood of detecting an actual effect (Cohen, 1988). Of all the comparisons reported here, we will focus on the statistical power of two: (a) changes in relative abundance; and (b) changes in distance from the source. The former was the chief concern of the NPAL marine mammal monitoring effort, that whales might be abandoning formerly preferred habitat. Two previous studies during ATOC transmissions off Kauai (Frankel and Clark, 2002) and California (Calambokidis, 1998) found subtle, but statistically significant, increases in distance from the source during experimental phases for humpback whales.

The changes in relative abundance across years (Table II) presented a small effect size ($\gamma=0.14$) and low power of 0.13. In order to produce the recommended level of statistical power of 0.80 (Cohen, 1988) using eight surveys, sighting rate would have to change by 47%. The latter corresponds to a medium effect size ($\gamma=0.50$). Thus, the present schedule of eight surveys is sufficient to detect at least moderate changes in sighting rate.

Changes in distance from the NPAL source for the baseline year (2001) vs the transmission years (2002, 2003) represent a low effect size ($\gamma=0.28$) and power of 0.40. In order to increase power to the recommended level of 0.80, the average distances would have to increase by an additional 7%, corresponding to a medium effect size ($\gamma=0.47$). Thus, the average number of sightings across the three years sampled ($N=70$) is sufficient to detect at least moderate changes in distance from the source.

IV. DISCUSSION

The results presented here suggest that exposure to the NPAL (formerly ATOC) source during the 2 years sampled with the source on (2002,2003) did not change the numbers of whales wintering north of Kauai, nor did it produce any discernible distributional changes as measured by distance from the source and from shore, as well as depths of sighting locations. Given previous reports of increases in distance from the ATOC source when comparing humpback whale positions during sound-on and sound-off blocks (Calambokidis, 1998; Frankel and Clark, 2002) as well as other distributional changes (Frankel and Clark, 2002), one of several possible interpretations is likely true, either: (a) whales in the study area have habituated to the sound; (b) the current survey design had insufficient statistical power to detect the small levels of distributional change; or (c) distributional changes may have occurred during transmissions, but were not sufficiently long-lasting to be detected (Note: surveys during 2002 and 2003 commenced upon the completion of a 24-h cycle of NPAL transmissions.)

Regarding the habituation hypothesis, little is known about the habituation of marine mammals to human-made sound. Watkins (1986) provided evidence of whales habitu-

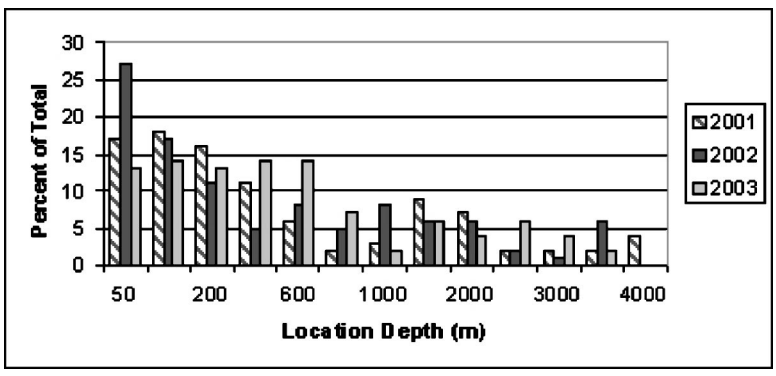


FIG. 6. Sighting location depths by year. Humpbacks generally preferred shallower water across all years surveyed. Comparisons between the source off year (2001) and source on years (2002 and 2003) revealed no statistically significant differences ($p>0.05$).

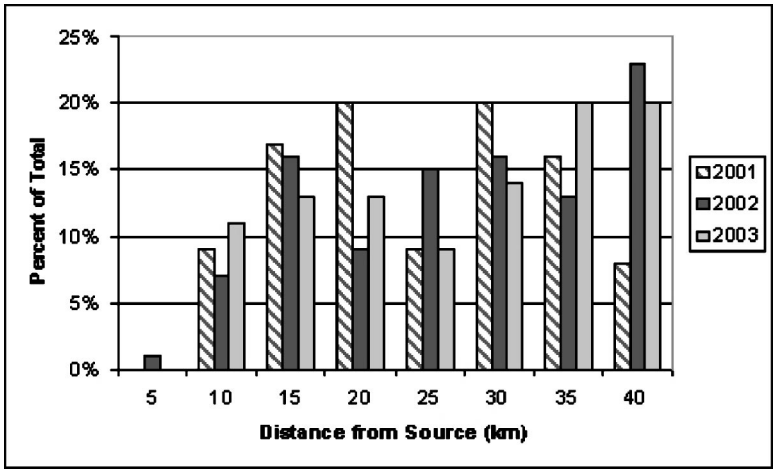


FIG. 7. Distance from source. More humpbacks occupied the furthest distance category (35–40 km) during 2002 and 2003 when the NPAL source was transmitting. However, comparisons between the sound-on years (2002 and 2003) with the control year (2001) showed no statistically significant differences ($p > 0.05$).

ating to the presence of vessels off the coast of Cape Cod over a 25-year period, with humpback whales showing the greatest reductions in level of response. Presumably, the main sensory channel mediating the response of marine mammals to boats is sound. In the present case, by the time the NPAL source first transmitted (2002), many of the same whales had likely been exposed to the same (ATOC) source during 2 previous years (1998 and 1999) and may have habituated to the sound as a result.

The low statistical power hypothesis is supported by the fact that, for both transmission years (2002, 2003), the median distances from the NPAL source and median distances offshore were remarkably similar and both greater than the baseline year (2001) (Table IV), although the differences were not statistically significant. Calambokidis (1998) found statistically significant increases in distance from the California ATOC source for humpbacks and sperm whales based on substantially larger sample sizes than those reported here ($N = 207$ humpbacks and 210 sperm whales), thus affording greater statistical power. Additionally, the mean distances from the transmission source in Calambokidis' study were much smaller (means of 14–16 km for both species) than those reported here (Table IV). This implies that the California whales were exposed to higher received levels of sound than the whales in this study, since the source levels of both the California ATOC source and the NPAL source were identical. As a result, the effect size derived from the Calambokidis results for distance from the source was considerably larger ($\gamma = 0.45$) than that derived from the present results ($\gamma = 0.28$). The latter results in a power of 0.40, as opposed to 0.90 for the Calambokidis study. With a larger number of surveys per year, and hence, greater numbers of sightings, it is possible that the small differences reported here would be statistically significant.

The hypothesis that responses to the source are short-lived is consistent with the conclusions of Frankel and Clark (2002). The authors noted that "these subtle, short-term [distributional] effects are apparently not large enough to alter the abundance and large-scale distribution of humpbacks off Kauai" (p. 660). They also noted the similarity of these responses to those seen in the presence of vessels (e.g., Bauer, 1986) where normal behavior typically resumes upon passing of the vessel. It may be that following the 24-h cycle of NPAL transmissions, any changes in distribution quickly return to baseline levels.

Regardless of which interpretation is true, it appears safe to say that after 4 years of exposure to ATOC/NPAL transmissions (i.e., 1998, 1999, 2002, and 2003) the humpback whales continue to return to their wintering grounds north of Kauai and show little changes in normal patterns of distribution. The current regime of eight surveys per year appears sufficient to detect changes in relative abundance and distribution of at least moderate effect size and has the additional advantage of monitoring much larger areas than shore-based observation (ca. 4000 km^2 for the present surveys vs 55.5 km^2 described by Frankel and Clark, 2000, 2002). Current plans call for continuing the marine mammal monitoring surveys for the remaining years of transmission.

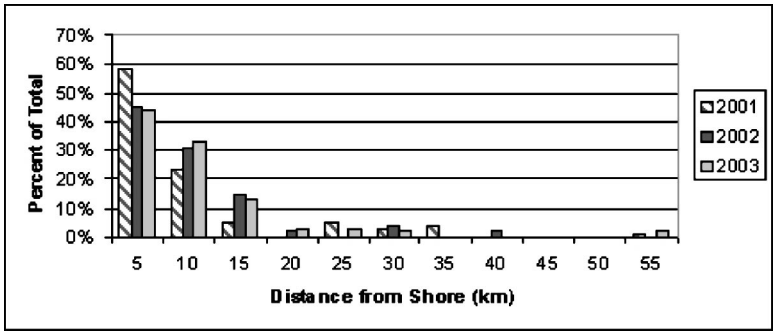


FIG. 8. Distance from shore. Humpbacks were seen slightly further from shore during 2002 and 2003 when the NPAL source was on, relative to 2001 when it was not. However, the differences were not statistically significant ($p > 0.05$).

ACKNOWLEDGMENTS

Data reported here were collected under Scientific Collecting Permit No. 642-1536-00 issued by NOAA Office of Protected Resources to the author. Funds for these surveys were provided by the North Pacific Acoustic Laboratory as sponsored by the Office of Naval Research (ONR). I would like to thank our competent crew of observers including Kim Andrews, Brian Branstetter, Marlee Breese, Alison Craig, Tori Cullen, Mark Deakos, Adam Frankel, Matthias Hoffmann-Kuhnt, Marc Lammers, Lori Mazzuca, Amy Miller, Michele Morris, Tom Norris, Adam Pack, and Scott Spitz. Thanks also to Scott Spitz and Mark Deakos, who prepared the graphics in this report, to Amy Miller for her GIS depth work, and to Jonathan Hino who assisted with data analysis. I am also grateful to Paula Mobley for her exceptional ground support and for reading earlier versions of this manuscript. Finally, ''mahalo'' to John Weiser and Howard Word for their superb piloting.

ARPA (**1995a**). ''Final environmental impact statement for the Kauai Acoustic Thermometry of Ocean Climate project and its associated Marine Mammal Research Program,'' Available from Advanced Research Projects Administration, 3701 N. Fairfax Dr., Arlington, VA 22203.

ARPA (**1995b**). ''Final environmental impact statement for the California Acoustic Thermometry of Ocean Climate project and its associated Marine Mammal Research Program,'' Available from Advanced Research Projects Administration, 3701 N. Fairfax Dr., Arlington, VA 22203.

Au, W. W. L., Nachtigall, P. E., and Pawloski, J. A. (**1997**). ''Acoustic effects of the ATOC signal (75 Hz, 195 dB) on dolphins and whales,'' J. Acoust. Soc. Am. **101**, 2973–2977.

Baker, C. S., and Herman, L. M. (**1981**). ''Migration and local movement of humpback whales (*Megaptera novaeangliae*) through Hawaiian waters,'' Can. J. Zool. **59**, 460–469.

Balcomb, III, K. C., and Claridge, D. E. (**2001**). ''A mass stranding of cetaceans caused by naval sonar in the Bahamas,'' Bahamas J. Sci. **2**, 2–12.

Bauer, G. B. (**1986**). ''The behavior of humpback whales in Hawaii and modifications of behavior induced by human interventions,'' Ph.D. thesis, University of Hawaii at Manoa, Honolulu, HI.

Buckland, S. T., Anderson, D. R., Burnham, K. P., Laake, J. L., Borchers, D. L., and Thomas, L. (**2001**). *Introduction to distance sampling* (Oxford University Press, Oxford).

Calambokidis, J. (**1998**). ''Effects of the ATOC sound source on the distribution of marine mammals observed from aerial surveys off central California,'' World Marine Mammal Conference, Monte Carlo, Monaco, 20–24 January, 1998, Abstracts, p. 22.

Cerchio, S. (**1998**). ''Estimates of humpback whale abundance off Kauai, 1989 to 1993: Evaluating biases associated with sampling the Hawaiian Islands breeding assemblage,'' Mar. Ecol.: Prog. Ser. **175**, 23–34.

Cohen, J. (**1988**). *Statistical Power for the Behavioral Sciences* (Erlbaum, Hillsdale, NJ).

Costa, D. P., Crocker, D. E., Gedamke, J., Webb, P. M., Houser, D. S., Blackwell, S. B., Waples, D., Hayes, S. A., and Le Boeuf, B. J. (**2003**). ''The effect of a low frequency sound source (ATOC) on the diving behavior of northern elephant seals, *Mirounga angustirostris*,'' J. Acoust. Soc. Am. **113**, 1155–1165.

Frankel, A. S., and Clark, C. W. (**1998**). ''Results of low-frequency playback of M-sequence noise to humpback whales, *Megaptera novaeangliae*, in Hawaii,'' Can. J. Zool. **76**, 521–535.

Frankel, A. S., and Clark, C. W. (**2000**). ''Behavioral responses of humpback whales (*Megaptera novaeangliae*) to full-scale ATOC signals,'' J. Acoust. Soc. Am. **108**, 1–8.

Frankel, A. S., and Clark, C. W. (**2002**). ''ATOC and other factors affecting distribution and abundance of humpback whales (*Megaptera novaeangliae*) off the north shore of Kauai,'' Marine Mammal Sci. **18**, 644–662.

Frankel, A. S., Mobley, J. R., Jr., and Herman, L. M. (**1995**). ''Determining auditory thresholds in free-ranging cetaceans using biologically meaningful sound,'' in *Sensory Processes of Aquatic Animals*, edited by R. A. Kastelein, J. A. Thomas, and P. Nachtigall (De Spil, Woerden, Netherlands), pp. 55–70.

Frantzis, A. (**1998**). ''Does acoustic testing strand whales?'' Nature (London) **392**, 29.

Herman, L. M., and Antinoja, R. C. (**1977**). ''Humpback whales in the Hawaiian breeding waters: Population and pod characteristics,'' Sci. Rep. Whales Res. Inst. **29**, 59–85.

Malme, C. I., Miles, P. R., Clark, C. W., Tyack, P., and Bird, J. E. (**1983**). ''Investigations on the potential effects of underwater noise from petroleum industry activities on migrating gray whale behavior,'' Report No. 5366 submitted to the Minerals Management Service, U.S. Department of the Interior, NTIS PB86-218377 (Bolt, Beranek, and Newman, Washington, DC).

Malme, C. I., Miles, P. R., Clark, C. W., Tyack, P., and Bird, J. E. (**1984**). ''Investigations on the potential effects of underwater noise from petroleum industry activities on migrating gray whale behavior. Phase II: January 1984 migration,'' Report No. 5586 submitted to the Minerals Management Service, U.S. Department of the Interior, NTIS PB86-218377 (Bolt, Beranek, and Newman, Washington, DC).

Malme, C. I., Miles, P. R., Tyack, P., Clark, C. W., and Bird, J. E. (**1985**). ''Investigations on the potential effects of underwater noise from petroleum industry activities on feeding humpback whale behavior,'' BBN Report 5851, OCS Study MMS 85-0019. Report from BBN Laboratories, Inc., Cambridge, MA, for U.S. Minerals Management Service, U.S. Department of the Interior, NTIS PB86-218385 (Bolt, Beranek, and Newman, Anchorage, AK).

Mobley, J. R., Jr., Bauer, G. A., and Herman, L. M. (**1999**). ''Changes over a ten-year period in the distribution and relative abundance of humpback whales (*Megaptera novaengliae*) wintering in Hawaiian waters,'' Aquat. Mammals **25**, 63–72.

Mobley, J. R., Jr., Herman, L. M., and Frankel, A. S. (**1988**). ''Responses of wintering humpback whales (*Megaptera novaeangliae*) to playback of recordings of winter and summer vocalizations and of synthetic sound,'' Behav. Ecol. Sociobiol. **23**, 211–223.

Mobley, J. R., Jr., Spitz, S. S., Grotefendt, R., Forestell, P. H., Frankel, A. S., and Bauer, G. A. (**2001**). ''Abundance of humpback whales in Hawaiian waters: Results of 1993–2000 aerial surveys,'' Report to the Hawaiian Islands Humpback Whale National Marine Sanctuary, 26 Nov., 2001. Available in pdf form at: http://socrates.uhwo.hawaii.edu/SocialSci/jmobley/2001sancreport.pdf

Munk, W. H., and Forbes, A. M. G. (**1989**). ''Global ocean warming: An acoustic measure?'' J. Phys. Oceanogr. **10**, 1765–1778.

National Research Council (NRC) (**2003**). *Ocean Noise and Marine Mammals* (National Academy, Washington, DC).

Richardson, W. J., Greene, C. R., Jr., Malme, C. I., and Thomson, D. H. (**1985**). *Marine Mammals and Noise* (Academic, San Diego).

Tyack, P. L., and Clark, C. W. (**2000**). ''Communication and acoustic behavior of dolphins and whales,'' in *Springer Handbook of Auditory Research; Hearing by Whales and Dolphins*, edited by W. W. L. Au, A. N. Popper, and R. R. Fay (Springer, New York), pp. 156–224.

Watkins, W. A. (**1986**). ''Whale reactions to human activities in Cape Cod waters,'' Marine Mammal Sci. **2**, 251–262.