CODEN: JASMAN

ISSN: 0001-4966

The Journal of the Acoustical Society of America

Vol. 109, No. 5, Pt. 1 May 2001

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 109, NO. 5, PT. 1, MAY 2001

CONTENTS—*Continued from preceding page*

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 109, NO. 5, PT. 1, MAY 2001

CONTENTS—*Continued from preceding page*

PSYCHOLOGICAL ACOUSTICS [66]

SPEECH PRODUCTION [70]

SPEECH PERCEPTION [71]

SPEECH PROCESSING AND COMMUNICATION SYSTEMS [72]

BIOACOUSTICS [80]

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 109, NO. 5, PT. 1, MAY 2001

CONTENTS—*Continued from preceding page*

LETTERS TO THE EDITOR

CONTENTS

ACOUSTICAL NEWS—USA

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

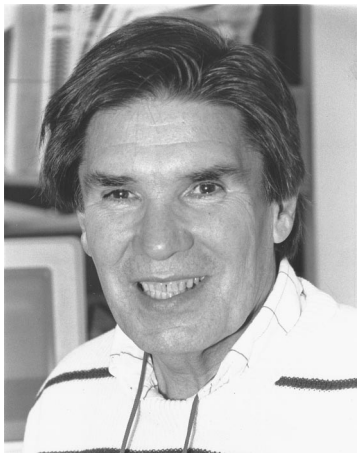
Editor's Note: *Readers of this Journal are encouraged to submit news items on awards, appointments, and other activities about themselves or their colleagues. Deadline dates for news items and notices are 2 months prior to publication.*

New Fellows of the Acoustical Society of America

Takayuki Hidaka—For contributions to acoustical design of concert and opera halls.

Christy K. Holland—For contributions to the safety of diagnostic ultrasound and to the affairs of the Society.

Christian Soize—For contributions to the understanding of the acoustic response of complex structures.



William A. Watkins—For contributions to marine mammal bioacoustics.

Michael S. Howe awarded the Per Bruel Gold Medal for Noise Control and Acoustics

ASA fellow Michael S. Howe was awarded the Per Bruel Gold Medal of the American Society of Mechanical Engineers at its 2000 International Mechanical Engineering Congress and Exposition, held 5–10 November 2000 in Orlando, FL. He was recognized for his work on problems related to the production of sound and vibration by unsteady fluid flow interacting with elastic structures. This medal bears the name of Dr. Per Bruel, the Danish engineer and industrialist, who pioneered the development of highly sophisticated noise and vibration measuring and processing equipment. It is given in recognition of eminent achievement and extraordinary merit in the field of Noise Control and Acoustics. The achievement must include useful applications of the principles of noise control and acoustics to the art of science of mechanical engineering. Previous recipients of the medal include K. Uno Ingard, Lothar Cremer, Alan Powell, Miguel Junger, David Crighton, Eric E. Ungar, Allan D. Pierce, Maurice Sevik, and John E. F. Williams. Dr. Howe is professor of aerospace and mechanical engineering at Boston University. He received the Ph.D. at Imperial College London and currently serves as Associate Editor of the *Journal of the Acoustical Society of America* for Aeroacoustics.

 0001-4966/2001/109(5)/1753/2/$18.00

Peter A. Traykovski receives ONR Young Investigator Grant

ASA member Peter A. Traykovski, Applied Ocean Physics and Engineering Department, Woods Hole Oceanographic Institution, was named a recipient of an ONR Young Investigator Grant. The Office of Naval Research (ONR) announced the award of 26 grants totaling $8.5 million as a result of the Fiscal Year 2001 ONR Young Investigator Program competition. A total of 191 proposals were submitted in response to this year's program announcement. The Young Investigator Program supports basic research by exceptional faculty at U.S. universities who received a Ph.D. or equivalent degree within the preceding five years. Grants to their institutions provide up to $100,000 per year for three years; additional funds may be made available to purchase equipment related to the investigator's research. Young Investigators are selected on the basis of prior professional achievement, the submission of a meritorious research proposal, and evidence of strong support by their respective universities. ONR's Young Investigator awards recognize exceptional young scientists and engineers. The program supports outstanding research in a wide range of science and engineering fields that are critical to the evolution of a first-rate Navy and Marine Corps. Dr. Traykovski will conduct a field and modeling study of factors controlling sandy ocean inner shelf geometry and change over time. This research can help to predict sand transport processes that can bury underwater items.

USA Meetings Calendar

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

2001

4–8 June	141st Meeting of the Acoustical Society of America, Chicago, IL, Palmer House Hilton Hotel [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org].
7–9 June	International Hearing Aid Conference VI, Novel Processing and Fitting Strategies, Iowa City, IA [Rich Tyler; Tel.: 319-356-2471, E-mail: rich-tyler @uiowa.edu, WWW: www.medicine.uiowa.edu/ otolaryngology/news/news].
9–13 July	2001 SIAM Annual Meeting, San Diego, CA [Society for Industrial and Applied Mathematics (SIAM); Tel.: 215-382-9800; Fax: 215-386-7999; E-mail: meetings@siam.org; WWW: www.siam.org/meetings/ an01/].
15–19 August	ClarinetFest 2001, New Orleans, LA [Dr. Keith Koons, ICA Research Presentation Committee Chair, Music Dept., Univ. of Central Florida, P.O. Box 161354, Orlando, FL 32816-1354; Tel.: 407-823-5116; E-mail: kkoons@pegasus.cc.ucf.edu].
19–24 August	Asilomar Conference on Implantable Auditory Prostheses, Pacific Grove, CA [Michael Dorman, Dept. of Speech and Hearing Science, Arizona State Univ., Tempe, AZ 85287-0102; Tel.: 480-965-3345; Fax: 480-965-0965; E-mail: mdorman@asu.edu].
4–6 October	Ninth Annual Conference on the Management of the Tinnitus Patient, Iowa City, IA [Rich Tyler; Tel.: 319-356-2471; E-mail: rich-tyler@uiowa.edu; WWW: www.medicine.uiowa.edu/otolaryngology/news/news].
7–10 October	2001 IEEE International Ultrasonics Symposium Joint with World Congress on Ultrasonics, Atlanta, GA [W. O'Brien, Electrical and Computer Engineering, Univ. of Illinois, 405 N. Mathews, Urbana, IL 61801; Fax: 217-244-0105; WWW: www.ieee-uffc.org/2001].
29–31 October	NOISE-CON 01, The 2001 National Conference and Exposition on Noise Control Engineering, Portland, ME [Institute of Noise Control Engineering, P.O. Box 3206 Arlington Branch, Poughkeepsie, NY 12603; Tel.: +1-914-462-4006; FAX: +1-914-462-4006; E-mail: omd@ince.org; WWW: users.aol.com/inceusa/ ince.html].
15–18 November	American Speech Language Hearing Association Convention, New Orleans, LA [American Speech-Language-Hearing Association, 10801 Rockville Pike, Rockville, MD 20852; Tel.: 888-321-ASHA; E-mail: convention@asha.org; WWW: professional.asha.org/ convention/abstracts/welcome.asp].
3–7 December	142nd Meeting of the Acoustical Society of America, Ft. Lauderdale, FL [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org]. Deadline for submission of abstracts: 3 August 2001.

2002

21–23 February	National Hearing Conservation Association Annual Conference, Dallas, TX [NHCA, 9101 E. Kenyon Ave., Ste. 3000, Denver, CO 80237; Tel.: 303-224-9022; Fax: 303-770-1812; E-mail: nhca@gwami.com; WWW: www.hearingconservation.org/index.html].
10–13 March	Annual Meeting of American Institute for Ultrasound in Medicine, Nashville, TN [American Institute of Ultrasound In Medicine, 14750 Sweitzer Lane, Suite 100, Laurel, MD 20707-5906; Tel.: 301-498-4100 or 800-638-5352; Fax: 301-498-4450; E-mail: conv_edu@aium.org; WWW: www.aium.org].
3–7 June	143rd Meeting of the Acoustical Society of America, Pittsburgh, PA [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org].
2–6 December	Joint Meeting: 144th Meeting of the Acoustical Society of America, 3rd Iberoamerican Congress of Acoustics, and 9th Mexican Congress on Acoustics, Cancun, Mexico [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org/cancun.html].

OBITUARIES

Irving Hochberg • 1934–1999

Irving Hochberg, a member of the Acoustical Society of America, died on 30 October 1999 after an extended illness. Born on 17 April 1934 in Brooklyn, NY, he graduated with a B.A. in English from Brooklyn College in 1955. In 1957 he received a Master's degree in Speech from Teachers College of Columbia University. He subsequently received a doctorate in audiology from Pennsylvania State University in 1962.

Dr. Hochberg's distinguished career in academia began immediately after his graduation from Penn State. He was associated with New York University from 1962 to 1970 as a professor in the Department of Speech Education, as a Research Scientist in the School of Dentistry, and, finally, as the Director of the Program in Speech Pathology and Audiology in the Department of Special Education. In 1970 he joined the faculty of the Department of Speech at Brooklyn College of the City University of New York. Shortly thereafter, in 1974, he was appointed Executive Officer of the Ph.D. Program in Speech and Hearing Sciences at the Graduate School of the City University of New York (CUNY), a position to which he was reappointed every three years until his death. In all, he served more than 25 years as Executive Officer. In 1979, during his tenure at the CUNY Graduate School, he established and directed the Center for Research in Speech and Hearing.

His extensive research accomplishments are documented in articles that appeared in the *Journal of the Acoustical Society of America*, the *Journal of Auditory Research*, *Archives of Otolaryngology*, *Cleft Palate, Cortex, Ear and Hearing*, and other periodicals. The books he authored and coedited were concerned with issues in audiological testing (*Interpretation of Audiometric Results*), hearing aids (*Acoustical Factors Affecting Hearing Aid Performance*), and the speech of the deaf (*Speech for the Hearing Impaired: Research, Training and Personnel*). He served as editor and editorial consultant to the *Journal of Speech and Hearing Research*, the *Journal of Speech and Hearing Disorders*, and the *Journal of Communication Disorders*.

Dr. Hochberg's activities in professional organizations included the Vice-presidency and Presidency of the New York State Speech Association (as it was known then), as well as many committee and board assignments in the New York State organization. In his work for the National Council of Graduate Programs, he championed the cause of research in the speech and hearing profession, producing influential white papers that advocated the inclusion of basic science and practical research experience in the curriculum at all levels of higher education. He was a recognized leader in the concerted effort to preserve, foster, and renew the research enterprise in the speech and hearing sciences.

His work as Executive Officer of the Ph.D. Program at CUNY resulted in a reconfiguration of the Program. When he assumed leadership, the program had a relatively small number of students and was staffed by a faculty that, although distinguished, worked only part-time at the Graduate School. The laboratory facilities were minimally adequate. By exploiting some existing initiatives and creating others, he oversaw an expansion of the program which, in many respects, can only be described as astonishing. Within a decade the bulk of the teaching and mentoring duties had been assumed by a cadre of internationally renowned researchers and teachers with full-time appointments to the Program. Moreover, many of these faculty members had associations with other research institutions such as Haskins Laboratories, Bell Laboratories, and the Clark School for the Deaf. The augmentation and quality of the faculty, and the resultant increase in the facilities available in which to do research, naturally attracted large numbers of students. More than 150 students received doctorates while he was executive officer. Virtually all of those who needed financial support received it, many from the training grants that had been successfully proposed by Dr. Hochberg.

Dr. Hochberg's view of speech, language, and hearing sciences was a unified one. He sought to foster the notion of a single, overreaching discipline that comprises a number of specialty areas. He believed that hearing scientists and audiologists should be able to communicate easily with researchers and professionals specializing in speech and language, and that speech and language scientists should be able to understand what hearing scientists and audiologists were doing. For many years, the requirements of the program reinforced his notion, and the students who met those requirements were educated in a way that in more recent years has become increasingly uncommon.

LAWRENCE J. RAPHAEL

Warren W. Denner • 1938–1999

Warren W. Denner, a member of the Acoustical Society of America, died 9 November 1999 of a heart attack in Nizhny Novgorod, Russia, while on a business trip. Dr. Denner was a physical oceanographer whose broad interests included ambient noise in the arctic, shallow water acoustics, and environmentally adaptive sonar.

He was born in Portland, OR, on 9 February 1938, grew up there, and attended Portland State University, where he received a Bachelor's Degree in Physical Science in 1961. For his graduate work, he went to Oregon State University, where he studied physical oceanography and received his M.S. in 1963. He subsequently spent a year (1963–1964) at the Douglas Marine Station of the University of Alaska, and afterwards began a teaching and research career at the Naval Postgraduate School (NPS) in Monterey, CA, in 1964. The association with this institution extended from 1964 through 1978, but he continued his graduate education with Oregon State, receiving his doctorate in physical oceanography in 1969. At NPS, he became Associate Professor of Oceanography, taught physical oceanography, and focused his research on arctic ocean processes, ice mechanics, and arctic ambient noise. The 15-year association with NPS was interrupted by a two year stint (1974–1976) with the University of Alaska where he served as Director of the Naval Arctic Research Laboratory and as Associate Professor of Marine Science. In 1978 he left NPS to assume the position (1978–1982) as the Director of the Newfoundland Institute of Cold Ocean Science at Memorial University of Newfoundland in Canada. There he also held an appointment as Professor of Physics and Ocean Engineering.

In 1982, Dr. Denner returned to Monterey to assume the position of Technical Director and Senior Scientist of the Science Applications International Corporation (SAIC). He left SAIC in 1990 to start his own company, Environmental Research Associates, with the subsequent name EOS Research Associates.

Although Dr. Denner was primarily a physical oceanographer, he is remembered for his visionary understanding of how acoustics can unify many diverse interests of ocean scientists: geological, physical, biological, and chemical. Each of these has distinct roles in the propagation of sound energy in the ocean, and acoustic techniques provide powerful means for understanding properties and processes in each of these areas of ocean science. As President of EOS Research Associates, he was instrumental in the establishment and organization of several important scientific collaborations between American and Asian Pacific Rim ocean scientists. Most notable was his vital role in initiating and coordinating the Asian Sea International Acoustics Experiment (ASIAEX) under Office of Naval Research (ONR) sponsorship. He helped assemble an excellent international team consisting of some of the best ocean scientists and collaborators in the world, and worked diligently in the overall international coordination of ASIAEX until the time of his death. He had unique diplomatic skills and was able to help facilitate large scale projects that involved ocean scientists from the United States and other Pacific Rim countries.

CHING-SANG CHIU

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).]*

Physical Acoustics: Principles and Methods, Volume XXIII—Ultrasonic Instruments and Devices I Reference for Modern Instrumentation, Techniques, and Technology

Academic Press, San Diego, CA, 1999.
xii + 482 pp. Price: $179.00 ISBN: 0124779239.

Physical Acoustics: Principles and Methods, Volume XXIV—Ultrasonic Instruments and Devices II Reference for Modern Instrumentation, Techniques, and Technology

Academic Press, San Diego, CA, 1999.
xiv + 372 pp. Price: $153.00 ISBN: 012477945X.

R. N. Thurston and Allan D. Pierce, Editors-in-chief

Emmanuel P. Papadakis, Volume editor

Combined volume available in paperback under the title:

Ultrasonic Instruments and Devices

Emmanuel P. Papadakis, editor

Academic Press, San Diego, CA, 1999.
809 pp. Price: $99.95 ISBN: 0125319517.

You may begrudge, but you cannot deny the changing environment of research and development in the field of physical acoustics. The roles of universities, industries, and government agencies have all evolved since the first volume of the series *Physical Acoustics: Principles and Methods* was published in 1964. From the beginning, the series has been a perennial favorite among those pursuing research in ultrasonics and acoustics. The series provides a foundation for our accumulated understanding and chronicles advances through contributions from the most eminent figures in the field. The earlier volumes of the series focused primarily on fundamental research, but there is a natural progression that occurs as ideas are transformed through research and development into commercial products and applications. In the present era of corporate downsizing, outsourcing, and increased global competition, companies are constantly striving to reduce their product cycle development times and have deliberately eliminated their capacity in certain technical areas. These changes, as well as the maturity of the field, have influenced the nature of research in physical acoustics. It is only fitting that as the focus of physical acoustics research changes, so too should the reference series. The purpose of the latest two volumes is to establish the current state-of-the art in the commercial development of ultrasonic devices and instruments. Topics include medical ultrasound, nondestructive testing, process control, surface acoustic waves, frequency control devices, research measurement instruments, transducer design, and ultrasonic microscopy.

The two volume set, also available in a single-volume paperback edition from Academic Press (ISBN 0125319517), chronicles the development of ultrasonic devices arising from past research efforts that have reached commercial success. This focus is an ambitious one, and it is natural to wonder how effectively the scope has been covered, or where the emphasis has been placed. The series editor, Emmanuel Papadakis, admits that the range of topics is not inclusive and, in particular, he decided to omit coverage of acoustic emission, high-intensity ultrasound, and bulk wave delay lines. The resulting selection is not justified in terms of commercial significance. In fact, the two most commonly associated applications of ultrasound, fish finders and sonar, are not discussed. Perhaps the down-selected list reflects the editor's interest in topics that are continuing to benefit significantly from ongoing research and development. As a result of this focus, however, many well-established ultrasonic applications are not covered. For example, mechanical processes such as ultrasonic cleaning, welding, milling, agglomeration, flocculation, liquid atomization, as well as ultrasonically assisted chemical and electrochemical processes are not discussed. Despite these omissions, these volumes provide the most complete discussion on ultrasonic instrumentation, techniques, and technology that I have encountered.

The two-volume set continues the tradition of excellence that has made *Physical Acoustics* the leading reference series in the field of ultrasound (UT). The ten chapters are written by 20 contributors, all leading authorities in their areas. Despite the stature of the contributors, the discussions are directed toward a general technical audience. Each chapter provides broad coverage of its topic with extensive references and illustrations. Each chapter is complete, self-contained, and can be read in isolation, yet surprisingly there is very little overlap in the material covered. And each chapter must meet the same challenge of providing unbiased historical perspective along with balanced detail on state-of-the-art developments. The list of contributors in Vol. XXIII includes Albert Goldstein and Raymond Powis on medical UT diagnostic, Emmanuel Papadakis on UT nondestructive testing, and Lawrence Lynnworth and Valentin Magori on UT industrial process control sensors. The contributors to Vol. XXIV include Emmanuel Papadakis, Clyde Oakley, Alan Selfridge, and Bruce Maxfield on UT transducer fabrication; Fred Hickernell on surface acoustic wave devices; John Vig and Arthur Ballato on UT frequency control devices; Robert Gilmore on industrial UT imaging, and Bruce Chick on UT research instruments. Additionally, there is a collection of short essays at the beginning of each volume.

Consistent with the theme of ultrasonic device commercialization, Vol. XXIII begins with a set of six unique essays on technology transfer. Oddly this content is repeated verbatim at the beginning of Vol. XXIV, placing undue emphasis on a section which is pleasant but does not deserve special mention. In the first essay, Aaron Gellman of Northwestern University gives a concise description of the elements of successful technology transfer. He highlights the role of champions, market forces, and transfer mechanisms in overcoming the omnipresent barriers to success. Next, Emmanuel Papadakis relates a collection of anecdotes on technology transfer, many of which can be attributed to his personal experience as a university professor, small business and large business scientist. Neil Goldfine provides a personal account of his company's success with meandering winding magnetometer technology. He discusses how the technology was developed from basic research, through innovation, technology transfer, and finally into profitable growth for his company Jentek Sensors, Inc. In the fourth essay, Stephen

 0001-4966/2001/109(5)/1761/6/$18.00

Ringlee laments about the specific difficulties that face innovators in the conservative NDT marketplace. These conditions include low growth rates, small fractured markets, innovation-stifling regulations, and the poor scalability of individual inspection solutions. William Lord, Satish Udpa, and Robert Harris of Iowa State University co-author an incongruous contribution that relates how faculty, student, and industrial teaming on research and development projects provides pragmatic lessons in an engineering curriculum. In the final essay we hear from the government perspective as Arthur Ballato and Richard Stern of the US Army Communication-Electronics Command review the technology transfer programs within their lab. Cooperative research and development agreements (CRADAs), patent license agreements (PLA), and small business innovation research programs (SBIR) are mentioned.

Detailed discussions of ultrasonic devices begins in earnest in Chap. 2, Medical Ultrasonic Diagnostics. Albert Goldstein and Raymond Powis do a tremendous job of carrying the reader through 30 years of development in the medical ultrasound field with a lucid, well-structured, and liberally referenced discussion. They begin by explaining the physics behind ultrasonic image formation, relating image resolution, contrast, noise, and perception to fundamental transducer and tissue parameters. Next, analog and digital imaging systems design concepts are reviewed with excellent use of illustrative images, color plates, and fundamental expressions that help the reader quickly comprehend the material. These discussions move from the historical use of analog scan converters through to the present practice of transducer arrays, digital beamformers with digital signal processing schemes, Doppler imaging, and contrast agents.

Emmanuel Papadakis provides the next chapter on (ultrasonic) Nondestructive Testing, writing with a broad, heuristic, and, at times, cavalier style. He describes the development of the modern flaw detectors, thickness gages, and large fixed installation scanning systems. Although there are numerous figures, many of the images overlap each other in content. Often they appear out of sequence with the corresponding text and are not consistently dated to allow the reader to establish their historical perspective.

Material property testing systems based on ultrasonic velocity and resonance spectroscopy are also reviewed. Despite a brief section on tube wall thickness measurement systems, there is very little mention of systems used by the field inspection service community. While it is true that some of these systems are proprietary, common themes on the use of time-of-flight diffraction, phased array beam steering and focusing, and synthetic aperture focusing have been finding their way into commercially available devices. Common service inspection areas include nuclear and fossil fuel power generation, pipeline construction and maintenance, storage tanks, pressure vessels, and other large structures. These areas represent an important part of the ultrasonic testing market and have not been addressed here or elsewhere in the two-volume set.

There are also places where the author's personal and collegial style contrasts with the historical reference authority of the *Physical Acoustic* series. In one passage the author cites a ''landmark discovery by Torre and others at General Motors (GM) that ultrasonic velocity was monotonically related to nodularity, yield strength and tensile strength in ductile (nodular) gray cast iron....'' This statement is referenced by a personal communication between the author and Torre in 1986. I could not find any references to their discovery in the open literature. Yet, the correlation between cast iron nodularity, sound velocity, and mechanical properties has been studied and reported for over 40 years by a variety of researchers on numerous continents. Other cast iron inspection experts that I consulted were also unaware of the GM group's efforts. Given the long standing and active interest in this topic, it is unfortunate that this discovery from which ''the field of ultrasonic velocity for material property measurement came into its own and spawned many instruments'' is not more effectively referenced or dated. In short, this chapter could have benefited from a more detailed exposé.

Chapter 4, the final chapter of Vol. XXIII, written by Lawrence Lynnworth and Valentin Magori, is devoted to Industrial Process Control Sensors and Systems. Both authors have distinct writing styles and there is some duplication in their content so the flow of the material is a bit choppy, but the topic coverage is excellent. They discuss the currently available ultrasound solutions for the four major process control issues: level, flow, temperature, and pressure measurements. These methods are typically applied to single-component gas and liquid environments. In addition, they discuss the use of ultrasound to determine the concentration or density of binary gas and liquid mixtures. A variety of measurement approaches are reviewed for both clamp on and wetted probes such as range gating, echo tracking, tag cross-correlation, Doppler shift, and Gauss–Chebyshev integration. Wireless surface acoustic wave sensors are also mentioned here and the discussion blends well with the more complete review of the topic given in Chap. 3 of Vol. XXIV. Practical solutions to the problems encountered in real world applications are given along with numerous examples of commercially successful products.

Chapter 2 of Vol. XXIV (Chap. 1 is a repeat of Chap. 1, Vol. XXIII) is devoted to Fabrication and Characterization of Transducers. This well-written chapter flows smoothly through four sections discussed sequentially by the authors: Emmanuel Papadakis—Monolithic Piezoelectric Plate Transducers, Clyde Oakley—Composite Transducers, Alan Selfridge—PVDF Film Transducers, and Bruce Maxfield—Electromagnetic Acoustic Transducers. The emphasis is placed on reviewing the collected research and outlining how this has been incorporated into commercial products. No mathematical foundations are given. Rather the fundamental developments are cited with a long list of relevant historical and contemporary references. Despite the emphasis on commercial practice, there is curiously little mention of the software design tools, such as PZFlex (Weidlinger Associates), PiezoCAD (SonicConcepts), PASS (R/D Tech), Imagine3D (Utex Scientific Instruments Inc.), CIVA (Centre d'Etudes Atomiques) and CAPA (WisSoft) used by many nondestructive testing and medical transducer manufacturers in the design process.

In the next chapter Fred Hickernell presents an inspired discussion on Surface Acoustic Wave Technology Macrosuccess through Microseisms. The author writes with an infectious enthusiasm, and even hints of pride, as he relates the growth and success this field has experienced in the past 30 years. Starting from 1964, when modern surface acoustic wave (SAW) devices were first proposed, the author reviews the development progress. Initially the basic interdigital transducer, processing, fabrication, and materials issues are described. Then the use of SAW devices as delay lines, filters, resonators, oscillators, sensors, correlators, and other signal processing operators is discussed. The author goes to great lengths to provide wonderfully rich reference and appendix sections that allows the reader to pursue any interest in greater detail.

Chapter 4 of Vol. XXIV, Frequency Control Devices, is written by John Vig and Arthur Ballato. They begin by motivating our interest in the subject of quartz resonators by discussing applications. Numerous civilian applications, such as spread spectrum cellular communication and hand-held global positioning systems, have benefited from the militarily driven advances in quartz resonators. After they convince us of the pervasive and important nature of these devices, they review the fundamentals of frequency control. The important operating parameters such as stability, tunability, noise, drift, as well as the impact of temperature and other physical effects are discussed in the context of various quartz crystal fundamentals, crystal cuts, oscillator circuit designs, and compensation mechanisms. In the final section they compare frequency control devices with quartz sensors, transducer and filter applications.

Robert Gilmore handles the topic Industrial Ultrasonic Imaging/Microscopy in Chap. 5 of Vol. XXIV. The author draws heavily from his previous publications to provide the reader with a consolidated and polished presentation on this topic. Both C-scan imaging and scanning acoustic microscopy are discussed, but in contrast to other treatments of these topics the author chooses to emphasize the similarities, not the differences. Simple ray trace geometry arguments are used to support a discussion on near surface and subsurface volume imaging. The effects of lens geometry, spherical aberration, spatial sampling, point spread, and material response are considered in the context of lateral and axial resolution. The chapter is concluded with a short discussion on image processing, test target fabrication, and application examples.

Bruce Chick, a cofounder of the company Ritec Inc., wrote Chap. 6, entitled Research Instruments and Systems. After a brief historical introduction, this short chapter concentrates on describing how attenuation and velocity measurements are determined using a quadrature phase detection instrument developed and marketed by his company. The chapter also mentions a few other equipment manufacturers by name, but is not exhaustive in its coverage. A pulsed laser ultrasonic system (Rudolph Technology, Inc.) and a wide-band acoustic emission measurement system (Digital Wave Corp.) are discussed, but numerous competing manufacturers such as, Sonoscan Inc., Physical Acoustics Corp. UltraOptec Inc., and SecondWave Systems, are not mentioned here or elsewhere in the two volume set.

Despite the noted omissions, these volumes successfully establish themselves as the quintessential references on commercial development of ultrasonic devices and instruments. The expansive coverage that they provide on this very broad topic are is without parallel and the volumes serve as a valuable knowledge gateway. Each chapter provides a very readable account of the selected topic, liberally referenced and illustrated. The table of contents and index are extensive, allowing one to quickly locate relevant information. I find myself reaching for this reference set on a regular basis and I encourage anyone interested in this technology area to obtain personal copies.

GRANT A. GORDON
McDermott Technology, Inc.
Lynchburg Research Center
Lynchburg, Virginia 24506

Handbook of Radiation and Scattering of Waves: Acoustic Waves in Fluids, Elastic Waves in Solids, Electromagnetic Waves

Adrianus T. de Hoop

Academic Press, August 1995.
1110 pp. Price: $209.00 hardcover ISBN: 0122086554.

This is an advanced book dealing with linear-wave radiation and scattering; it contains a wealth of useful information about these topics; it derives the equations of motion describing mechanical wave propagation in unconventional ways; and material is repeated from one part to the next making it easy to use as a handbook but annoying to read as an advanced textbook or research monograph. Almost no detailed calculations are given; the emphasis is on general theorems that help the reader formulate problems and verify the accuracy of numerical calculations. The intended readers are engineers and others who do advanced numerical or analytical calculations in linear acoustic, elastic and electromagnetic waves. The parts dealing with acoustic and elastic waves have a slight orientation towards the concerns of researchers in geoacoustics. Two appendices are included, one dealing with Cartesian tensors and one with integral transforms, so that, beyond some training in advanced calculus and linear algebra, the mathematical prerequisites are not demanding. Problems of varying difficulty appear throughout the book, all with answers. The lists of references, which appear at the ends of the chapters, are not very complete. The index is adequate and is based on key words and phrases rather than concepts. For example, looking up the phrase ''direct scattering'' will give all the needed page numbers, but the word ''embedding'' is not listed so that it is hard to find where it is first defined. The print is easily read, the diagrams are clear, but the binding is inadequate for a book this size. It is an expensive book: it should be bought by libraries and perhaps larger research groups, but the individual should purchase it only if he or she works on linear wave propagation problems.

The book consists of the three parts indicated by its subtitle. Each part can be read independently. However, this means the reader will see much the same formulations over and over again, often forced into a single pattern that is more descriptive for one field theory than another. Moreover, too many of the sentences are of the pattern-writing type. Note below the reviewer's own use of this kind of writing: ''Part X,..., is *n* pages long and ...'' is repeated three times. The book could have benefited from severe editing.

Cartesian tensors are used throughout. The notation however is not standard: $\partial_i \tau_{i,j}$ is used rather than the more usual notation $\tau_{ij,i}$ or $\partial_i \tau_{ij}$. The reader is advised to work through Appendix A of the book which deals with tensors. He or she will be rewarded: the discussion is concise but very complete, and provides an excellent introduction to the subject. As well, the reader is advised to work through Appendix B, which deals with integral transforms, to learn the various transform conventions used in the book. The author is very careful in his use of these, particularly of the temporal Laplace transform.

Part one, radiation and scattering of acoustic waves in fluids, is 283 pages long and consists of 8 chapters. Part two, radiation and scattering of elastic waves in solids, is 309 pages long and consists of 8 chapters. The two parts are quite similar. Each part can be separated into two smaller parts: the first of these gives a derivation of the equations of motion and the second deals with radiation and scattering problems, including inverse problems. The derivations of the equations of motion are somewhat different from those taught by the reviewer to engineering students. The author attempts a generality in his presentation—perhaps with the goal of including effects induced by fluid-filled porous media—at the cost of losing some simple but important physical information. For example, on page 34, equation (2.7-29), the author gives the compatibility relation $\epsilon_{i,n,k}\, \partial_n\, \partial_t \Phi_k = \epsilon_{i,n,k}\, \partial_n f_k$. The term $\partial_t \Phi_k$ is a rate of mass flux and f_k a body force. This equation is telling the reader that in an ideal fluid the flow, away from a localized body force, is irrotational. This fact is not immediately apparent. However, an experienced reader will find the presentation original and thought provoking. The chapters dealing with radiation and scattering are excellent. Therein lies the reason for reading the book. The topics covered include the following: time-dependent and time-harmonic propagation, symmetry considerations used when formulating problems, radiation and scattering in unbounded domains, applications of reciprocity, usually in combination with a free-space Green's tensor.

Chapter 15 of the book is representative of what the reader can expect. It describes elastodynamic reciprocity relations and their applications. The applications are to modeling transmission and reception by transducers, and to formulating both direct and inverse radiation and scattering problems.

How might the reader use this chapter? Imagine that he or she wants an analytic expression for the compressional-wave attenuation of material X in terms of the measured transmitted voltage. The measurement is done by coupling two wide aperture transducers, opposite to one another, to a plate made from X. The reader goes to this chapter and reads with care the section on transmission and reception. He or she finds that, by taking one reciprocating state to be the wave field in the absence of attenuation and the second to be that when attenuation is present, and by assuming that any shear wave that may be excited can be ignored, the reciprocity relation gives an expression for the compressional-wave attenuation, but that the expression contains a number of terms that refer to the particular measurement arrangement. However, the reader notes that the same measurement can be done on a plate of mild steel, whose compressional-wave attenuation is known. This latter fact is used to remove the various terms associated with the measurement arrangement and what is left is an expression for the compressional-wave attenuation in material X relative to that in mild steel. What the chapter has done has told the reader how to do this calculation correctly. The book succeeds quite well at fulfilling such a use.

Part three, radiation and scattering of electromagnetic waves, is 388 pages and consists of 14 chapters. Some of the chapters are however very short. The pattern laid out in the first two parts is again repeated. The writing style becomes slightly livelier and somewhat more attention is paid to the physical theory than in the previous parts. The chapters on wave propagation again are very thorough and perhaps somewhat more complete than in the first two parts. The very last chapter, which describes problems arising in electromagnetic compatibility, is out of place and could have been omitted.

The reader who wants to learn the subjects treated, but has little previous knowledge of them will find the book too demanding. However, if the reader is experienced then it works quite well as a handbook provided a certain initial effort is made to understand how the equations are derived and used. Moreover, because the book describes the work of a group of researchers who have very much helped to shape the subject of wave propagation, any researcher in this subject will find that his or her competence at it will improve from a careful study of how the topics of this book are considered.

JOHN G. HARRIS
2723 McDaniel Ave.
Evanston, Illinois 60201

Signal Processing for Intelligent Sensor Systems

David C. Swanson

Marcel Dekker, Inc., New York, NY, 2000.
Signal Processing and Communications Series (K. J. Ray Liu, Series Ed.) Xiii+616 pp. Price: $195 ($79.75 for classrooms) ISBN: 0-8247-9942-9.

There is a myriad of textbooks available on all aspects of signal processing, yet this new book finds a niche to fill. Intelligent sensor systems are able to collect and analyze data, adapt, self-diagnose, make predictions, and are reprogrammable. They are a new and important technology, the development of which requires a mature understanding of the capabilities and limitations of both sensor technology and signal processing.

Signal Processing for Intelligent Sensor Systems introduces the reader to the essential elements of the subject in a way that instills enthusiasm for the subject as well as an appreciation of its limitations, caveats, and subtleties. The author's own enthusiasm and insights are conveyed in a congenial and nonintimidating style. The material is presented in a consistent framework with abundant examples and references to applications.

Dr. David C. Swanson is with the Applied Research Laboratory at Pennsylvania State University and is a professor in the Graduate Program in Acoustics at that university. Many applications of intelligent sensor systems are acoustical in nature and the author has emphasized this, making the book well suited to acousticians. Only a minimal knowledge of signal processing is prerequisite to benefit from this book, yet experts in the field may enjoy the insights and practical advice it offers.

The emphasis given to motivation and insight is very valuable but it brings with it some demerits. In many places the book is more descriptive than analytic and the reader who desires greater detail or rigor is obliged to seek other references for clarification or more complete information. Fortunately, there is no shortage of such and a sampling is identified in the references and bibliographies.

The book is organized into 15 well-focused chapters that are grouped into five parts. Each chapter contains an informative introduction and summary. Problems, bibliography, and references are included at the end of each chapter. A basic index is provided and answers to the problems are included as an appendix. The references and bibliography are somewhat limited but the industrious reader can generally find further readings easily. In general, the individual sections and chapters can be read on their own with only minimal reference to other parts of the book.

The print quality is excellent, with many equations, figures, and tables. Although the quality and format of the figures is variable, they are generally well thought out and informative. The book suffers from frequent typographical and editorial errors, requiring considerable tolerance and alertness of the reader. (Apologies are owed to E. Galois, a famous French Mathematician, for misspellings of his name.) Unfortunately, neither the Table of Contents nor the Index does justice to the breadth of information in the book. Many sections contain uncharted excursions into interesting side issues. For example, a brief discussion of satellite GPS systems, as an application of multipath resolution, is neither indexed nor listed in the table of contents.

Part I covers the fundamentals of digital signal processing. The treatment is necessarily cursory, requiring either prior familiarity or supplementary reading. Nevertheless, it develops an appreciation for practical aspects of signal processing and provides the essential tools needed throughout the remainder of the book. For example, initial conditions are carefully treated and the problem of mapping poles from the analog to the digital domain is presented from the perspective of obtaining meaningful physical quantities. State variable theory and fixed gain trackers are presented in a straightforward manner and establish a foundation on which subsequent chapters build.

Part II addresses frequency domain processing with chapters on the Fourier transform, spectral density, and wavenumber transforms. There are excursions into polyspectra, the wave equation, and structural vibration. The section includes important material on error analysis and quality measures, which are topics that are often overlooked or poorly treated. Power spectral density receives brief, yet careful attention. The presentation of wavenumber transforms shows good applications and examples, but the footings in Fourier optics and role of the aperture geometry is somewhat obscure. In this part the author tries to balance descriptive and analytic styles but the result is a bit uneven, requiring careful study of the text in places.

In Part III the presentation seems to get into full stride, covering adaptive system identification and filtering. There is greater analytical clarity here and the presentation flows well across a wide range of topics. The material is presented following the unifying principles of basis sets, projection operators, and orthogonal spaces, without indulging in the tedium of mathematical rigor. The author returns consistently to these notions, covering LMS, RLS, lattice structures, Kalman filters, and recursive structures. The presentation is abbreviated in places, but has numerous examples and illustrations.

Part IV covers wavenumber sensor systems. It begins with a relatively detailed discussion of probability distributions and detection theory. The importance of error estimates and the Cramer–Rao lower bound are emphasized. An excursion into acoustic wave propagation modeling is interesting but seems out of place. In the treatment of adaptive beamforming the author includes an interesting section on the application of maximum length sequences to multipath resolution—this is an example of the broad and practical perspective of the book.

Part V is titled ''Signal Processing Applications'' but it is actually an overview of the important aspects of intelligent sensor systems. A succinct definition of intelligent sensor systems seems illusive but this final section begins with a ''starting point'' for a definition. It then touches on pattern recognition, neural networks, fuzzy logic, feature selection, and condition-based monitoring, giving the reader a flavor for the essential issues through examples and basic formalisms. Following this, the section reaches into some practical aspects of sensors and electronics with a compact and focused presentation of circuit noise reduction techniques, transducer equivalent circuit representations, and calibration. Noise cancellation is squeezed in, drawing on earlier material, and the book finishes up with a brief discussion of active noise cancellation.

The greatest strength and the greatest weakness of this book is its motivational and intuitive treatment of a broad field, omitting detail and rigor that is available elsewhere in favor of examples and insights that are valuable to students and professionals alike. This book offers the reader a maturity of knowledge that is often only achieved in the course of lengthy and difficult experience.

DAVID I. HAVELOCK
Acoustics and Signal Processing Group
National Research Council
Ottawa, ON K1A OR6, Canada

The Making of Symphony Hall—Boston *A History with Documents*

Richard Poate Stebbins

Boston Symphony Orchestra, Boston, 2000.
xi+234 pp. Price: $24.95 (hardcover) ISBN: 0-9671148-1-0.

The author of this highly readable new work, Richard P. Stebbins, has had an association with Symphony Hall that goes back to the 1920s, when he attended the Young People's Concerts of the Boston Symphony Orchestra; then later as a singer in the Harvard-Radcliffe chorus under the direction of Serge Koussevitzky; and in recent years as a volunteer in the Symphony Orchestra's Archives. He has done a remarkable job of pulling together a mass of source materials, including previously unpublished correspondence between the leading figures involved in the construction of Symphony Hall. The book is well documented by extensive notes, many of them of general interest. For example, in Note 10 concerning the Boston Music Hall, the original home of the Orchestra, Dr. Stebbins has this to say: ''Not to be overlooked in any account of the Music Hall's varied uses was its employment by Henry James for the climactic episode of his 1886 novel *The Bostonians*.''

The person most responsible for establishing the Boston Symphony Orchestra in 1881 was Henry Lee Higginson, a distinguished citizen who had a solid background in music. Its first home, the Music Hall, had been built in 1852, and was a poorly ventilated firetrap, located on what had now become valuable property in a newly desirable business location. The driving force behind the construction of a new concert hall was Higginson. He

found property in Boston that would serve as a suitable site, and then approached the celebrated architect Charles Follen McKim of McKim, Mead & White, the leading architectural firm in the nation, about becoming the architect for the project. McKim responded enthusiastically. Before he left on an extended trip to Europe, where he was to visit a number of concert halls, Higginson asked McKim to take with him an outline of those features of the old Music Hall that he considered desirable. These included seating for approximately 2600 listeners, and minimizing the number and size of windows and locating them so as to reduce street noise inside the hall. With regard to the stage, he indicated that ''the authority on sound here—Prof. Cross [of MIT]—tells me that an angular sounding board over the stage is best of all.''

As Stebbins relates, during the early part of 1893, long before actual plans for a new concert hall could be developed, municipal authorities proposed a scheme for a new rapid transit system on elevated tracks that would pass directly *through* the old Boston Music Hall. This would have required that the old hall be demolished without delay, leaving the orchestra without a home; but fortunately, the proposal was brought to a public election and defeated in a close vote, allowing adequate time for the planning and construction of the new concert hall.

The book contains a chronicle of events leading to the construction of Symphony Hall that will interest anyone seriously concerned with the performance of orchestral music in America, but it will be of particular interest to members of the Acoustical Society in the field of architectural acoustics. After all, Symphony Hall is usually considered the mother of all concert halls of acoustical excellence in this country, setting a standard against which subsequent halls have been compared for many generations. It is a hall with a rectangular plan (often called a shoebox configuration) and a highly coffered ceiling, with additional diffusion of sound provided by the irregular surfaces on the side walls and rear wall.

McKim's initial concept for the hall was a classical design with a rectangular plan, in which the rear of the hall opposite the stage was terminated by a semicircular configuration resembling an ancient Greek theater, including sculptured figures. The hall's construction was financed by a corporation set up to sell capital stock in the New Boston Music Hall, issued in a sufficient amount to enable the project to proceed; and at its first meeting on January 6, 1894, the purchase of the Huntington Avenue site was approved. Two days later, a large wood and plaster model of McKim's proposed structure was placed on display in the Public Library for the public to see.

Further plans had to be held in abeyance because of poor economic conditions, so no additional work was done for almost four years. Then, on October 27, 1898, Higginson wrote to McKim, ''While we hanker for the Greek theatre plan, [i.e., the semi-circular configuration at the rear of the rectangular hall], we think the risk too great as regards results, so we have definitely abandoned that idea. We shall therefore turn to the general plan of our Music Hall and of the halls in Vienna and Leipzig, the latter being the best of all, and Mr. Cotting will ask for a plan on those lines....We have 18 months to complete our work.''

At a meeting of the board of directors of the corporation on December 1, 1898, a serious discussion was held concerning the shape of the new hall: i.e., a rectangular plan having its length about twice its width, two balconies, and the stage at one end. Many in Boston were concerned whether the new concert hall would have acoustical properties that measured up to their expectations. The events described in this book are of special interest because of Wallace Clement Sabine's role in its ultimate design. It was Harvard President Charles W. Eliot who knew of Sabine's work and alerted Major Higginson to it. As a result, Sabine was taken on as a technical advisor to Higginson. His first meeting in this capacity was on January 25, 1899. He had a continuing role in the design.

The great American physicist Joseph Henry, during his tenure as the first Secretary (i.e., Director) of the Smithsonian Institution in Washington between 1846–1878, had observed that in a large room, the duration of the time which reverberation remained audible increases with its volume, and decreases with the amount of acoustic absorption in the room. But Henry did not develop a mathematical relation between these parameters. This formulation was later provided by the brilliant experiments of Sabine, resulting in his well-known reverberation-time formula, the technical basis for which is presented in his *Collected Papers on Acoustics*.

Because of the certain design similarities, comparisons were made between Symphony Hall and both the earlier Boston Music Hall and the Leipzig Gewandhaus. But Sabine indicated that neither of the earlier halls served as an architectural model for the new concert hall, although ''both were used...as definitions and starting points on the acoustical side of the discussion.''

George S. Hutchings, the builder of the organ (the predecessor of the present Aeolian-Skinner organ, which was installed in 1949), had a dispute with the architect concerning the height of the ceiling above the stage, wanting to increase it by 8 feet. When Sabine was asked whether this would be acceptable, he suggested it would be better not to increase the height over the stage, but to rearrange the pipes—a solution that was finally adopted. Later, Hutchings declared that it was the finest organ he had ever built.

The original design of Symphony Hall by McKim included 16 statues placed in niches that lined the walls above the second balcony—6 on each side of the hall and an additional 4 in the back. Unfortunately, the board of directors did not provide funds for the statuary. Thus, all the niches but one were unoccupied at the inauguration of Symphony Hall in 1900. Stebbins tells us that there was some objection to the appearances of the empty niches, so one of the patrons of the Symphony took it upon herself to raise money for this purpose. By the fall of 1902, statues had been placed in 14 of the 16 niches; the remaining two were filled some years later. As we know today, the inclusion of irregular bounding surfaces such as the statues, and the niches in which they are housed, provide a positive contribution to the acoustics of a concert hall by scattering sound from their surfaces and increasing the diffusion of sound in the hall. Stebbins, in Note 216, quotes Sabine in a communication to McKim assuring him that the statues ''will not in the least affect the reverberation of the hall.'' (Here Sabine must have been referring to the reverberation time in the hall rather than the general reverberation characteristics of the hall, for the additional diffusion by the statues does have the positive effect of smoothing out the decay of reverberant sound.)

The inaugural concert of Symphony Hall, the most important musical event in the history of Boston, took place on October 15, 1900 (before the introduction of the statuary in the empty niches) and featured Beethoven's *Missa Solemnis*. Music critics of two of New York's leading newspapers were highly complimentary in their judgment of the hall's acoustics, although the favorable judgments were not universal.

A substantial section of the book is devoted to documents, including copies of many of McKim Mead & White's original drawings and letters between the principal individuals associated with the Hall's construction-.There are also three appendices that many readers will find of interest: *Appendix A: Calculation in Advance of Construction*, by Wallace C. Sabine, reproduced from his 1900 paper, *Architectural Acoustics: Part I—Reverberation; Appendix B: Collateral Observations*, by Leo L. Beranek; and *Appendix C: Who Was Who*, a list of the cast of 68 individuals associated with the planning and/or construction of the building, directly or indirectly, with one or two lines about the involvement of each.

CYRIL M. HARRIS
Columbia University
New York, New York

Ruido: Fundamentos y *Control*

Samir N. Y. Gerges

Spanish Edition. (Florianopolis-SC-Brasil, 1998, contact gerges@mbox1.ufsc.br.)
555 pp. ISBN: 8590004601 x pb.

This book is the translation from Portuguese to Spanish of the book *Ruido—Fundamentos e Controle*, published in 1992. It fulfills a badly needed bibliography in Spanish. It is intended for members of the academia (students of acoustics or related matters), as well as general practitioners (engineers, safety professionals, industrial hygienists, etc).

Because some of the intended readers are students, there is a strong emphasis on theory with detailed deductions of some of the formulas, something that is not needed for the practitioners. The text is completed with illustrations, such as graphs, photos, and tables that make for relatively easy reading.

The book is divided into 13 chapters and contains 9 appendices. Each

chapter starts with an introduction, and ends with a list of references that can be used for further reading. References are relatively old; however, for the purpose of the book they can be quite useful.

Chapter 1, Acoustical Waves, is entirely theoretical, dealing with the physics of pressure, sound density, intensity, and other basic concepts such as the plane, spherical, and cylindrical waves. Concepts of sound level, as well as the sum and difference of levels, are also introduced here.

The second chapter is on effects of sound and vibrations. It starts with the description of the human ear, and deals mainly with hearing loss, although other effects from noise are also mentioned. The assessment of noise is introduced in this chapter, explaining the weighting curves as well as some criteria for annoyance and speech interference. Effects from vibrations are also treated here, although neither the physical phenomenon nor the units are explained. This is a serious omission, since a reader not familiar with this subject will have a hard time to go through it.

Chapter 3 is dedicated to measurement instruments. As expected, it covers the range of sound level meters, analyzers, dosimeters, etc. Although there is a great deal of material dealing with the different kind of signals, there is very little that a practitioner could find to help him choose the right instrument for a particular measurement. Also missing are instructions on how to perform the measurement itself and how to report the results. Particulars on the measurement of impulse noise are also missing.

A quite serious theoretical analysis of the issue of sound radiation of vibrating structures is the subject of Chap. 4. There is also an attempt to extract some practical conclusions, at the end of the chapter. However, the reader will only find general recommendations regarding the control of the vibrations.

Chapter 5 is one of the longest (50 pages). It deals with sound insulation. It provides a comprehensive review of concepts and definitions. However, there is a need for a more in-depth analysis of the sound wave impinging on a surface, so that the reader can better grasp the concepts of reflection, absorption, and transmission.

Surprisingly enough, free field sound propagation is the subject of Chap. 6, instead of being dealt with before in Chap. 5. The chapter contains the classical treatment by Maekawa as well as more recent developments by Hearing *et al.* for the calculation of sound barriers. The subject of barriers made by vegetation is examined quite extensively, something surprising since it is generally accepted that trees are not the best option for shielding of noise.

The title of Chap. 7 is Acoustics of Enclosed Spaces. One would expect the content to be close to Architectural Acoustics, which it does to some extent. However, it also contains issues not directly related, such as a study on sound sources, measurement of acoustical power, and sound absorption in reverberant chamber. There is also a section on measurements in an anechoic chamber, something not quite related to the subject of the chapter.

Another 50 pages or so are dedicated to sound absorption and absorptive silencers in Chap. 8, material complemented by a study of acoustic filters and resonators in Chap. 9. Probably those two chapters contain more practical applications than the previous seven. The same can be said regarding the remaining Chaps. 10 (Vibrations and Impact Isolation), 11 (Noise from Machinery), and 13 (Engineering Noise Control).

An entire chapter (Chap. 12) is dedicated to hearing protection devices. One would expect this subject to be dealt with right after Chap. 2 (Effects of Noise on Hearing) and not right before Engineering Noise Control, but that is the author's prerogative. The quite comprehensive treatment of this subject includes nonconventional devices such as those using active noise control as well as linear protectors.

In summary, this is a good reference book that will help fill the need for books on acoustics in Spanish. One serious drawback, however, is the quality of the translation: not only are there serious language problems, but also many technical terms and expressions are translated literally, as if the translator did not have an acoustical background. This does not help the reader, who on top of difficulties with the technical content has to fight language problems.

ALBERTO BEHAR
Institute of Biomedical Engineering
University of Toronto, Toronto, Canada

BOOKS RECEIVED

Ultrasound Imaging: Waves, Signals, and Signal Processing. Bjørn A. J. Angelsen. Emantec AS, Trondheim, Norway, 2000. 2 volumes. 1416 pp. $650.00 *pb*. ISBN: 82-995811-0-9 (www.ultrasoundbook.com).

Acoustical Imaging (Volume 24). Hua Lee (Editor). Kluwer Academic/Plenum, New York, 2000. 422 pp. $150.00 *hc*. ISBN: 0306465183.

Acoustical Imaging (Volume 25). Peter N. T. Wells and Michael Halliwell (Editors). Kluwer Academic/Plenum, New York, 2000. 580 pp. $175.00 *hc*. ISBN: 0306465167.

Fundamentals of Acoustics, 4th Edition. Lawrence E. Kinsler, Austin R. Frey, Alan B. Coppens, and James V. Sanders. Wiley, New York, 1999. 548 pp. $94.75 *hc*. ISBN: 0471847895.

Classic Papers in Shock Compression Science. James N. Johnson and Roger Cheret (Editors). Springer Verlag, New York, 1998. 524 pp. $99.95 *hc*. ISBN: 0387984100.

Introduction to Wave Propagation in Nonlinear Fluids and Solids. D. S. Drumheller. Cambridge U. P., New York, 1998. 536 pp. *hc* (534 pp. *pb*). $140.00 *hc* ($52.95 *pb*). ISBN: 0521583136 *hc* (0521587468 *pb*).

Signal Processing for Active Control. Stephen Elliott. Academic, San Diego, CA, 2000. 511 pp. $89.95 *hc*. ISBN: 0122370856.

Intermittency in Turbulent Flows. J. C. Vassilicos (Editor). Cambridge U. P., New York, 2000. 288 pp. $74.95 *hc*. ISBN: 0521792215.

Music and Memory. Bob Snyder. MIT, Cambridge, MA, 2001. 370 pp. $75.00 *hc* ($30.00 *pb*). ISBN: 0262194414 *hc* (0262692376 *pb*).

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*
DAVID PREVES, *Songbird Hearing, Inc., 5 Cedar Brook Drive, Cranbury, New Jersey 08512*
CARL J. ROSENBERG, *Acentech Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
KEVIN P. SHEPHERD, *M.S. 463, NASA Langley Research Center, Hampton, Virginia 23681*
WILLIAM THOMPSON, JR., *601 Glenn Road, State College, Pennsylvania 16803*

6,046,963

43.30.Yj DEPLOYABLE HULL ARRAY SYSTEM

Daniel M. Glenning, assignor to the United States of America as represented by the Secretary of the Navy
4 April 2000 (Class 367/188); filed 11 September 1998

This patent describes a planar, rectangular shaped, hydroacoustic, sensor array which is mounted on a rigid panel that can be stored in the space between the inner and outer hulls of a double-hull surface vessel or submarine. The edges of the panel are supported by guide rails and an exit slot exists in the outer hull to allow the panel to slide out from between the two hulls, by means of some motorized drive mechanism, to a deployed position in the surrounding water.—WT

6,050,361

43.30.Yj CAVITATION-RESISTANT SONAR ARRAY

Anthony A. Ruffa and Thomas R. Stottlemyer, assignors to the United States of America as represented by the Secretary of the Navy
18 April 2000 (Class 181/108); filed 17 September 1998

The cavitation resistance of an array of closely spaced, low-frequency, transducer elements (preferably class IV flextensional transducers) is said to be improved by filling the usual free-flooded interstices between the elements with rho-c rubber.—WT

6,076,629

43.30.Yj LOW FREQUENCY FLEXTENSIONAL ACOUSTIC SOURCE FOR UNDERWATER USE

Rune Tengham, assignor to Unaco Systems AB
20 June 2000 (Class 181/102); filed in Norway 30 April 1996

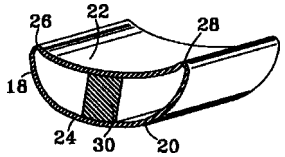
A standard class IV flextensional transducer, consisting of a shell of elliptical cross section with an unspecified drive unit mounted within the shell, is described. Novelty seems to revolve around the fact that the shell is formed from two curved plates, connected together with hinge supports, and in how the drive unit is supported within the shell.—WT

6,076,630

43.30.Yj ACOUSTIC ENERGY SYSTEM FOR MARINE OPERATIONS

Loran D. Ambs, assignor to Western Atlas International, Incorporated
20 June 2000 (Class 181/110); filed 4 February 1999

An acoustic source for seismic exploration consists of a flexible closed housing **18** containing an actuator **30**. This assembly is usually part of a larger apparatus that can be towed by a surface ship, or be remotely controlled, or be itself an autonomous vehicle. The actuator **30** may be a mechanical, electromechanical, pneumatic, hydraulic, electric, or any device



suitable for moving the flexible housing **18**. Upon excitation, the housing expands or contracts resulting in a net volume displacement of the surrounding water. A number of different embodiments of the housing shape and activation mechanism are discussed.—WT

6,088,299

43.30.Yj VERTICAL HYDROPHONE ARRAY

Louis W. Erath and Phillip Sam Bull, assignors to Syntron, Incorporated
11 July 2000 (Class 367/154); filed 4 December 1998

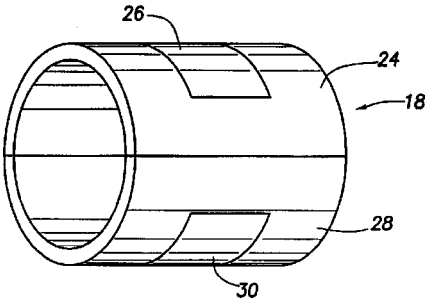
A number of long line arrays of hydrophones are each supported on vertical cables that extend downward from surface floats that also serve as communication ports to a towing vessel. The outputs of the several hydrophones in each line are sequentially time delayed so that the output signals combine coherently for a vertically propagating acoustic wave. Because the individual line arrays are not bottom anchored, the whole assembly can be readily towed to another region of the ocean.—WT

6,151,277

43.30.Yj HYDROPHONE WITH FERROELECTRIC SENSOR

Louis W. Erath and Gary Craig, assignors to Syntron, Incorporated
21 November 2000 (Class 367/173); filed 16 April 1999

A seismic-streamer line array comprises a number of coaxially positioned, cylindrical elements, each of which is fabricated from two half-cylinders **24** and **28** that are attached together around a central cable, not shown. Each half-cylinder has a machined-out region that houses a hydrophone **26** or **30**. The piezoelectric sensors may be conformal in shape, as



suggested by the figure, or simply flat discs that fit into the recessed regions. The two diametrically opposite hydrophones in each pair are electrically combined to form an acceleration-canceling sensor. The entire assembly is covered with a suitable, acoustically transparent, boot.—WT

6,029,113

43.38.Fx DIFFERENTIAL HYDROPHONE ASSEMBLY

Roger L. Woodall, assignor to the United States of America as represented by the Secretary of the Navy
22 February 2000 (Class 702/1); filed 21 December 1998

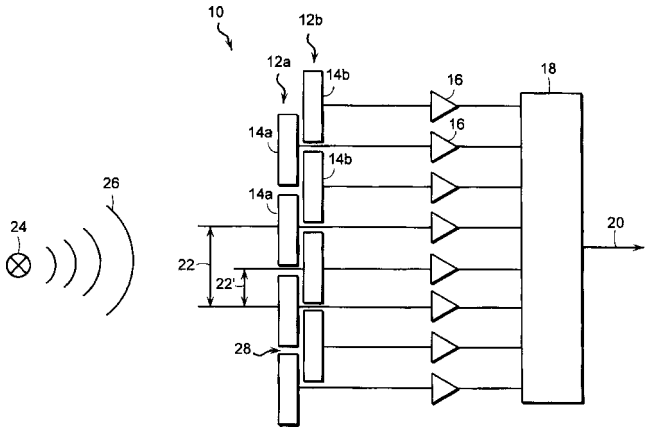
The device consists of two piezoceramic hemispheres, electroded on both inner and outer spherical surfaces, and oppositely polarized. The hemispheres are bonded together to form a sphere. Electrical leads are brought from each of the four hemispherical electrode surfaces as two pairs of twisted shielded leads. The four leads are suitably connected to a differential summing amplifier so that the signals from the two halves generated by some incident acoustic wave are additive while common mode noise signals are nulled.—WT

6,046,961

43.38.Fx MULTI-LAYER TILED ARRAY

Maurice J. Griffin *et al.*, assignors to the United States of America as represented by the Secretary of the Navy
4 April 2000 (Class 367/153); filed 16 December 1996

A sonar sensor array comprises two parallel planar sheets of PVDF material, **12a** and **12b**, each configured as a grid of square elements **14a** and **14b**. The elements of one layer are laterally displaced by one-half the center-to-center spacing **22′**, in two orthogonal directions, from those of the other layer. The signal from each element is conditioned by an individual preamplifier **16** and passed to a beamformer **18**. It is assumed that the outer



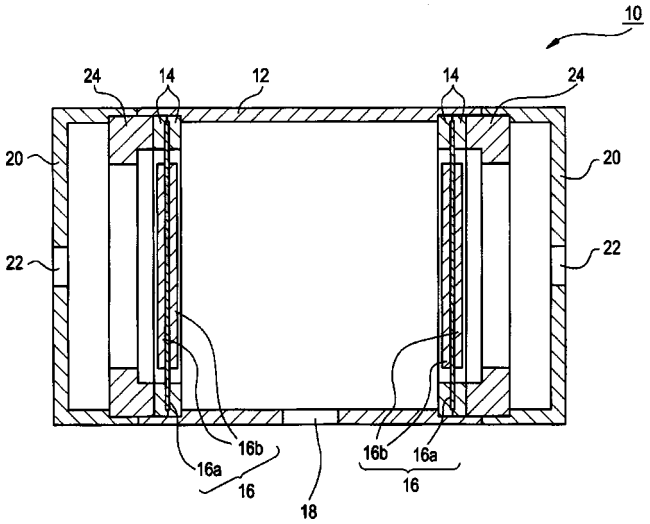
layer **12a** is sufficiently acoustically transparent that an incident wave **26** readily propagates through the layer **12a** to elements **14b**. Because the cross-sectional area of each element is greater than if both sets were coplanar, the capacitance of each element is greater with attendant improvement in the signal-to-noise ratio.—WT

6,108,429

43.38.Ja SPEAKER ADAPTED FOR USE AS A CENTER WOOFER IN 3-DIMENSIONAL SOUND SYSTEM

Takeshi Nakamura and Yoshiaki Heinouchi, assignors to Murata Manufacturing Company, Limited
22 August 2000 (Class 381/152); filed in Japan 15 April 1997

The inventors have miniaturized JBL's triple-chamber bandpass design and fitted it with piezoelectric transducers. Whether small or large, it is



still a variant of the original Bose multiple porting loudspeaker system disclosed in United States Patent 4,549,631.—GLA

6,111,970

43.38.Ja SUSPENSION FOR HIGH POWER PLEATED RIBBON TRANSDUCER

Alexander Voishvillo and Laszlo Megyeri, assignors to Cerwin-Vega, Incorporated
29 August 2000 (Class 381/398); filed 7 November 1997

In the past year or so there has been a revival of interest in Heil's ''Air Motion Transformer,'' which uses a meandering conductive ribbon bonded to a pleated diaphragm. The original design was notoriously sensitive to heat

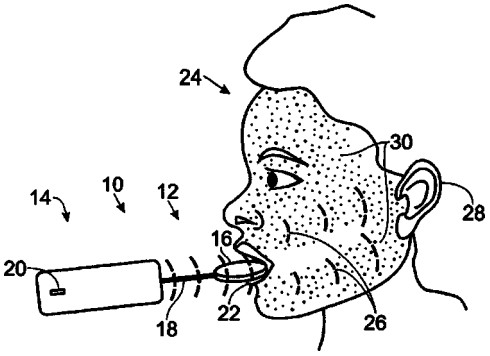
buildup. At moderate power levels, by today's standards, the diaphragm assembly would expand and buckle, changing sensitivity and response characteristics. This clearly written patent describes several modifications that maintain precise diaphragm geometry and provide improved heat dissipation.—GLA

6,115,477

43.38.Ja DENTA-MANDIBULAR SOUND TRANSMITTING SYSTEM

Andrew S. Filo and David G. Capper, assignors to Sonic Bites, LLC
5 September 2000 (Class 381/151); filed in Germany 23 January 1995

This acoustic lollipop generates audible signals via bone conduction. It



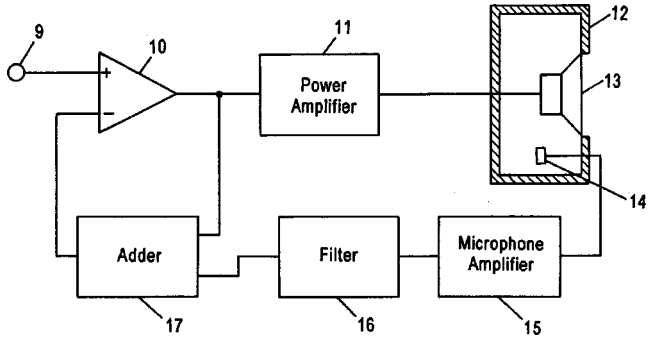
can be actuated by an electromechanical transducer or a music box mechanism.—GLA

6,122,385

43.38.Ja SOUND REPRODUCTION APPARATUS WITH STABLE FEEDBACK

Fumiyasu Konno *et al.*, assignors to Matsushita Electric Industrial Company, Limited
19 September 2000 (Class 381/96); filed in Japan 16 July 1996

It has long been known that the far-field response of a closed or vented box loudspeaker can be derived from a pressure microphone inside the box. The invention utilizes this arrangement in a feedback loop. The patent argues that this provides better stability over a greater bandwidth than near-



field pickup at the front of the cone. For a small box this may be true. However, it has also been known for more than 50 years that sound pressure is uniform inside the box only at frequencies whose wavelengths are more than eight times the largest box dimension.—GLA

6,127,919

43.38.Ja REARVIEW VEHICLE MIRROR WITH AUDIO SPEAKERS

James P. Wylin, assignor to DaimlerChrysler Corporation
3 October 2000 (Class 340/425.5); filed 10 March 1999

A combination rear view mirror and loudspeaker is proposed for an automobile. One or two tweeters are hidden from passengers' view, pointing toward the windshield.—KPS

6,091,828

43.38.Kb DYNAMIC MICROPHONE

Hiroshi Akino and Shioto Okita, assignors to Kabushiki Kaisha Audio-Technica
18 July 2000 (Class 381/355); filed in Japan 26 December 1997

This clever microphone design combines a conventional microphone capsule and a vibration detecting assembly. The vibration detecting apparatus provides both electrical and acoustical feedback to minimize noise caused by external vibrations.—GLA

6,111,966

43.38.Kb CAPACITOR MICROPHONE

Raimund Staat, Burgwedel, and Claus-Peter Hinke, Burgdorf, both of Germany
29 August 2000 (Class 381/174); filed in Germany 11 April 1997

Moisture is the enemy of capacitor microphones. Tiny, concealed microphones as used in musical productions can easily be contaminated by sweat. The design patented here incorporates a second, lighter diaphragm that serves as a moisture barrier.—GLA

6,122,389

43.38.Kb FLUSH MOUNTED DIRECTIONAL MICROPHONE

Steven R. Grosz, assignor to Shure, Incorporated
19 September 2000 (Class 381/361); filed 20 January 1998

This cardioid microphone assembly uses side-entry waveguides to allow flush mounting. The patent is short, clearly written, and includes measured performance curves.—GLA

6,088,460

43.38.Lc EQUALIZER AND AUDIO DEVICE USING THE SAME

Hiroyuki Funahashi and Ichiro Yokomizo, assignors to Rohm Company, Limited
11 July 2000 (Class 381/101); filed in Japan 8 November 1994

An analog audio equalizing circuit is described which is intended for use with portable, battery-operated stereo systems. In comparison with prior art, the circuit is said to provide higher audio quality at relatively low supply voltages.—GLA

6,124,542

43.38.Md WAVEFUNCTION SOUND SAMPLING SYNTHESIS

Avery L. Wang, assignor to ATI International SRL
26 September 2000 (Class 84/603); filed 8 July 1999

Digital waveform synthesizers change pitch by changing the sampling frequency. In most cases this requires a fairly complicated interpolation scheme, involving trade-offs between sound quality and computational cost. This invention uses waveform synthesis instead of pulse code modulation to overcome many prior art limitations. Difficult interpolative computations are performed off-line, greatly simplifying real time operations.—GLA

6,124,895

43.38.Md FRAME-BASED AUDIO CODING WITH VIDEO/AUDIO DATA SYNCHRONIZATION BY DYNAMIC AUDIO FRAME ALIGNMENT

Louis Dunn Fielder, assignor to Dolby Laboratories Licensing Corporation
26 September 2000 (Class 348/515); filed 17 October 1997

The patent gives a thorough review of the problems inherent in frame-based editing and splicing. The invention combines several interesting techniques to provide ''near-perfect'' splicing.—GLA

6,142,094

43.38.Pf DEPTH SENSITIVE MECHANICAL ACOUSTIC SIGNAL GENERATING DEVICE

Edmund J. Sullivan and Robert W. Gauthier, assignors to the United States of America as represented by the Secretary of the Navy
7 November 2000 (Class 116/70); filed 17 May 1999

A simple mechanical device for producing a one-time, transient, acoustic signal is described. The device consists of a spring-supported piston mounted in a closed housing. The piston is also held in place by a shear plug set to rupture at a predetermined depth in the ocean because of the pressure differential between the exterior and interior of the housing. Consequently, the piston is pulled inward because of the pre-tensioned support spring, whence it bottoms against a stop in the housing thereby creating the transient acoustic signal.—WT

6,104,824

43.38.Si HEADPHONE DEVICE WITH HEADBAND ARRANGED AROUND OCCIPITAL REGIONAL OF THE HEAD

Tomohiro Ito, assignor to Sony Corporation
15 August 2000 (Class 381/381); filed in Japan 13 March 1997

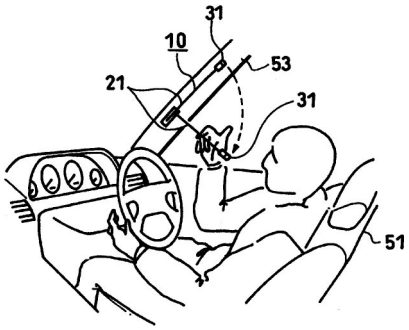
This redesigned headphone/headband assembly is intended to be easy to wear and comfortable. Moreover, it automatically maintains proper relationships between driver units and auricles.—GLA

6,111,964

43.38.Si MICROPHONE DEVICE NAVIGATION SYSTEM, COMMUNICATION DEVICE, AUDIO DEVICE, AND MOTOR VEHICLE

Yasuhiro Ishibashi, assignor to Sony Corporation
29 August 2000 (Class 381/86); filed in Japan 27 September 1996

A microphone in an automobile is proposed for use in controlling a navigation system or stereo, or for use with a mobile telephone. The microphone is mounted on the end of a rotatable arm, the base of which is



attached to the driver's side A-pillar. In contrast to other mounting arrangements such as the dashboard or sun visor, this arrangement is said to produce less interference with the driver's field of view.—KPS

6,134,456

43.38.Si INTEGRATED MOBILE-PHONE HANDSFREE KIT COMBINING WITH VEHICULAR STEREO LOUDSPEAKERS

Stephen Chen, assignor to E. Lead Electronic Company, Limited
17 October 2000 (Class 455/569); filed 15 June 1998

A system that allows hands-free operation of a mobile telephone in an automobile is described which relies on the audio components already present in the car. An interesting aspect of this patent is the inclusion of signal processing which allows the creation of a sound field in which the voice of the caller appears to be coming from in front of the driver, thus encouraging concentration on the road while in conversation.—KPS

5,982,903

43.38.Vk METHOD FOR CONSTRUCTION OF TRANSFER FUNCTION TABLE FOR VIRTUAL SOUND LOCALIZATION, MEMORY WITH THE TRANSFER FUNCTION TABLE RECORDED THEREIN, AND ACOUSTIC SIGNAL EDITING SCHEME USING THE TRANSFER FUNCTION TABLE

Ikuichiro Kinoshita and Shigeaki Aoki, assignors to Nippon Telegraph and Telephone Corporation
9 November 1999 (Class 381/18); filed in Japan 26 September 1995

In virtual reality systems, convincing out-of-head localization via headphones can be achieved by filtering left and right signals through suitable pinnae transfer functions. These can be retrieved from lookup tables. Unfortunately, different listeners have different ears and different pinnae functions. One feature of the invention is a method for deriving representative functions from a large number of subjects. The patent includes useful background information and more than 30 well prepared illustrations.—GLA

6,108,430

43.38.Vk HEADPHONE APPARATUS

Hirofumi Kurisu, assignor to Sony Corporation
22 August 2000 (Class 381/310); filed in Japan 3 February 1998

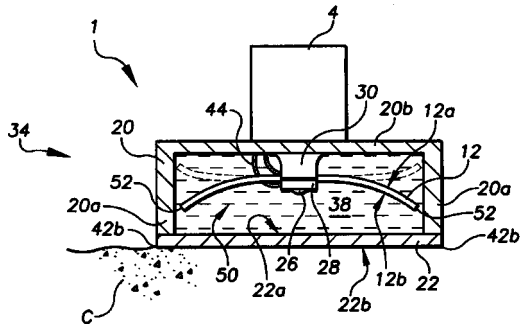
Virtual reality systems require a device to track the listener's head movements and elaborate electronic trickery to generate virtual sound sources in space. In effect, headphones are replaced by out-of-head-phones. The patent describes a relatively simple circuit which combines digital filters, delays, and level adjustments, all controlled by a microcomputer.—GLA

6,075,310

43.40.Rj ACOUSTIC TRANSDUCER WITH LIQUID-IMMERSED PRE-STRESSED PIEZOELECTRIC ACTUATOR IN ACOUSTIC IMPEDANCE MATCHED TRANSDUCER HOUSING

Richard P. Bishop, assignor to Face International Corporation
13 June 2000 (Class 310/328); filed 17 April 1998

A hand-operated tool for finishing the exposed surface of a layer of concrete comprises a working head **1** that houses one or more piezoelectric vibrators **12** each positioned within an oil bath **38**. The transfer of vibratory energy from the transducer to the layer of concrete **C** is facilitated by the use of an intermediate layer **22**, the actual working face of the tool, which is constructed as a fiber reinforced resin layer by the ''pultrusion'' technique



so that a gradient in the number of fibers exists from the top to the bottom surfaces of this layer. Hence there is an impedance gradient through the thickness of the layer **22** whereby one can attempt to match the impedance at the face **22a** to that of the oil bath **38** and the impedance at face **22b** to that of the concrete.—WT

6,080,924

43.50.Gf ACOUSTICAL REFLECTOR

Norman Cowen, Van Nuys, and Christopher Weik, Studio City, both of California
27 June 2000 (Class 84/453); filed 2 July 1999

This device has holes in two rotating disks. The holes can be aligned (or not) to allow more (or less) of the sound from a musical instrument to pass through the device. When the holes are aligned, the sound passes unhampered for transmission to the audience. When the holes are not aligned, the sound is partially reflected back to the performer, to enhance his balance and ensemble with other musicians.—CJR

6,113,193

43.50.Gf APPARATUS AND METHOD FOR AUTOMATICALLY REDUCING ENGINE EXHAUST NOISE

Joseph D. Kunzeman, assignor to Caterpillar Incorporated
5 September 2000 (Class 298/1H); filed 2 February 1999

The engine exhaust gases on a dump truck are often routed through ductwork in the truck bed in an attempt to reduce the tendency of material to adhere to the bed's surface in cold weather. When the bed is rotated to dump the load, the exhaust is routed through a second duct and muffler. This patent proposes the elimination of this second duct and muffler by reducing the engine idle speed in response to a sensor which determines when the bed has been rotated from its horizontal position.—KPS

6,119,490

43.50.Gf MATERIAL FOR SOUND-ABSORBENT AND HEAT-INSULATING LINING OF AN AUTOMOTIVE ENGINE COMPARTMENT

Claus Schierz *et al.*, assignors to Asglawo GmbH—Stoffe Zum Dammen und Verstarken
19 September 2000 (Class 66/170); filed in Germany 16 October 1996

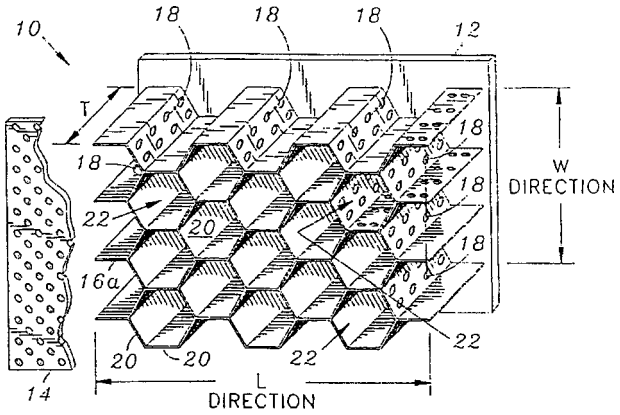
A material system is described which has both thermal and acoustical insulation properties, and is intended for use in engine compartments of automobiles. Thermal insulation is provided by a layer of carbon fibers supported by a rigid backing layer of thermoplastic fibers. Acoustically absorbent material is placed between these two layers. Details of the nonwoven stitching arrangement are given.—KPS

6,135,238

43.50.Gf EXTENDED REACTION ACOUSTIC LINER FOR JET ENGINES AND THE LIKE

Noe Arcas *et al.*, assignors to Northrop Grumman Corporation
24 October 2000 (Class 181/292); filed 9 September 1996

This is an extension to United States Patent 5,923,003. A typical acoustic liner for use in aircraft turbofan engines consists of a honeycomb layer sandwiched between an inner, acoustically porous layer and an outer impervious layer. This patent proposes, in contrast to typical practice, that neighboring honeycomb cells be in fluid/acoustical communication. This



introduces an additional acoustical absorption mechanism due to viscous losses between the cells. Also, layers of honeycomb, separated by porous septa, are described in which at least one of the layers has cells which are in fluid communication.—KPS

6,135,541

43.50.Gf AUTOMOBILE DOOR TO PROVIDE HIGH-QUALITY CLOSING SOUND

Lawrence Geise *et al.*, assignors to Honda Giken Kogyo Kabushiki Kaisha
24 October 2000 (Class 296/188); filed 12 August 1998

''An acoustical vibration damping system for a car door utilizes the side impact beam and connects intermediate areas of the beam to the inside of the outer skin by means of at least one intermediate bracket. An absorbing adhesive is used in the connection.''—KPS

6,112,167

43.50.Lj TIRE TREAD NOISE TREATMENT

Paul Zakelj, assignor to Bridgestone/Firestone Incorporated
29 August 2000 (Class 702/191); filed 8 January 1998

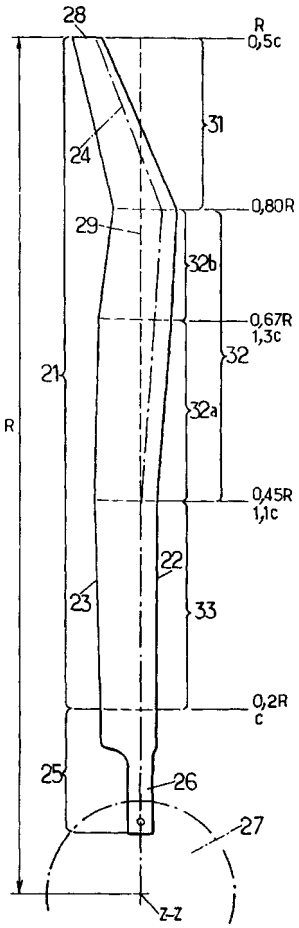
A method to assess the spectral content of a tire tread pattern is described. A typical tire consists of a small number of differing tread patterns of varying lengths concatenated to form the circumferential pattern. The essence of this patent is to use spectrograms from short-time Fourier transforms of the geometrical pattern, thus enabling the identification of circumferential acoustical ''hot spots.'' A MATLAB code to accomplish this is included.—KPS

6,116,857

43.50.Nm BLADE WITH REDUCED SOUND SIGNATURE, FOR AIRCRAFT ROTATING AEROFOIL, AND ROTATING AEROFOIL COMPRISING SUCH A BLADE

Wolf R. Splettstoesser *et al.*, assignors to Onera
12 September 2000 (Class 416/228); filed in France 10 September 1997

The dominant noise source for a helicopter in its landing phase is frequently so-called blade vortex interaction (BVI) noise, in which a rotor blade collides with the vortex generated by a previous blade passage. The blade planform described here has leading and trailing edge camber which varies along the outer half of the blade. It is designed to reduce BVI noise



through the reduction of vortex intensity and by modifying the geometry of the interaction of the vortex line and the blade leading edge. Relative to a blade having a rectangular planform and the same tip speed, this blade design yields a 6 dB(A) reduction in BVI noise. Furthermore, reductions of 4–5 dB(A) are achieved in level flight with reduced power consumption.—KPS

6,123,170

43.50.Nm NOISE REDUCING CONNECTION ASSEMBLY FOR AIRCRAFT TURBINE HOUSINGS

Alain Porte and Robert Andre, assignors to Aerospatiale Societe Nationale Industrielle
26 September 2000 (Class 181/214); filed in France 19 August 1997

This patent concerns the deployment of acoustical treatment in the inlet of aircraft turbofan engines. A means of mechanically connecting the inlet nacelle to the remainder of the engine is described which allows the duct area in the vicinity of the connection to be acoustically treated.—KPS

6,119,807

43.55.Dt SOUND ABSORBING ARTICLE AND METHOD OF MAKING SAME

Vernon C. Benson, Jr. and Glenn E. Freeman, assignors to PPG Industries Ohio, Incorporated
19 September 2000 (Class 181/208); filed 13 January 1997

This is an extension of United States Patent 5,796,055 and addresses the design of automobile glazing for noise and vibration reduction. Multilayer systems are described, with various combinations of single and double constrained damping layers, and ranges of glass and damping layer thick-

nesses. Details of materials selected for testing are given, along with the acoustical performance of some of the designs.—KPS

6,090,478

43.55.Ev SOUND ABSORBING/SHIELDING AND ELECTRIC WAVE ABSORBING PLASTIC SHEET CONTAINING ENCAPSULATED MAGNETIC FLUID, AND SOUND ABSORBING/SHIELDING AND ELECTRIC WAVE ABSORBING PLASTIC PANEL

Akihiko Nishizaki *et al.*, assignors to Nitto Boseki Company, Limited; Matsumoto Yushi-Seiyaku Company, Limited
18 July 2000 (Class 428/297.4); filed 11 September 1996

The patent describes a light-weight sound absorbing and insulating plastic sheet material apparently effective at low and medium sonic frequencies. The sheet material can be used in building materials. The sheet includes magnetic fluid-enclosing capsules (fine magnetic particles in an organic solvent) so that the sheet (and panels that use it) are useful in absorbing sound and also in electro-magnetic wave shielding. The insulting plastic sheet might be combined with a fiber-reinforced plastic sheet to form the insulating panel.—CJR

6,109,388

43.55.Ev SOUND ABSORBING MECHANISM USING A POROUS MATERIAL

Kouji Tsukamoto *et al.*, assignors to Mitsubishi Electric Home Appliance Company, Limited; Mitsubishi Denki Kabushiki Kalsha
29 August 2000 (Class 181/286); filed in Japan 31 August 1994

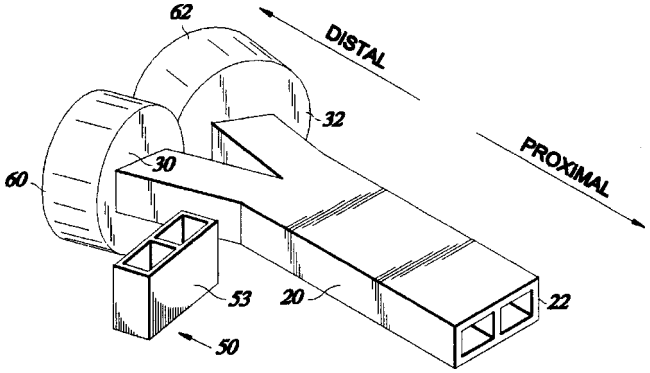
This is a sound absorbing porous plate material supported over a back air space. The panels would be placed around a noise generating source. The porous material forms resonating cavities in the back air spaces.—CJR

6,119,521

43.55.Ev APPARATUS AND METHOD FOR MEASURING THE ACOUSTIC PROPERTIES OF ACOUSTIC ABSORBERS

Belur Shivashankara *et al.*, assignors to Northrop Grumman Corporation
19 September 2000 (Class 73/589); filed 20 April 1998

This device essentially consists of two conventional impedance tubes placed side by side. Absorption properties of the test specimen placed at **22** can be determined for a range of incidence angles by phasing of the signals



to the acoustical drivers **60** and **62**. This method is of particular interest for test specimens which have extended rather than locally reacting impedance characteristics.—KPS

6,123,171

43.55.Ev ACOUSTIC PANELS HAVING PLURAL DAMPING LAYERS

Christopher P. McNett and John C. McNett, both of Indianapolis, Indiana
26 September 2000 (Class 181/290); filed 24 February 1999

This patent describes panels for sound absorption that have several layers in order to minimize transmission of sound, perhaps for aircraft or vehicle sound isolation. The panel design has rigid outer layers, but a greater number of flexible inner layers. The alternating rigid and flexible foam layers thus increase sound attenuation without increasing weight.—CJR

6,082,490

43.55.Pe MODULAR ANECHOIC PANEL SYSTEM AND METHOD

Chris W. Rowland, Austin, Texas
4 July 2000 (Class 181/295); filed 15 July 1997

These modular panels incorporate the necessary sound absorbing wedges as well as structural members to form both a self-supporting shell of an anechoic chamber and at the same time the materials for transmission loss. The panels have compression clip mountings and the result is claimed to be easier for assembly, repair, and replacement of damaged wedge tips.—CJR

6,082,489

43.55.Rg SOUND ISOLATION PLATE STRUCTURE

Keijiro Iwao and Yuji Shimpo, assignors to Nissan Motor Company, Limited
4 July 2000 (Class 181/286); filed in Japan 7 March 1997

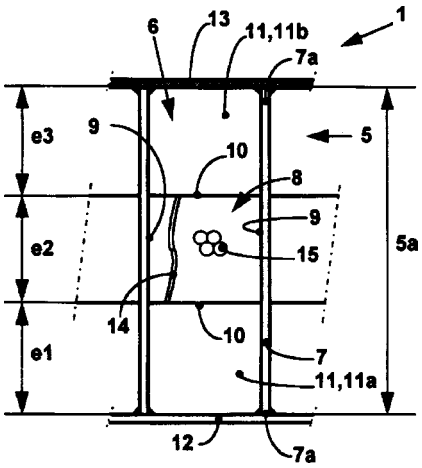
This patent describes a sound isolation plate, most likely for use on the undercarriage of a vehicle. The plate comprises a sheet of perforated board with lots of holes and cylinders behind the holes. This permeability allows the panels to both reduce noise and dissipate heat.—CJR

6,085,865

43.55.Rg SOUNDPROOFING PANEL AND METHOD OF PRODUCING SAID PANEL

Osmin Regis Delverdier *et al.*, assignors to Societe Nationale d'Etude et de Construction de Moteurs d'Aviation; Ateca; Hispano-Suiza Aerostructures
11 July 2000 (Class 181/292); filed in France 26 February 1998

The soundproofing panel has a honeycomb core sandwiched between a solid skin and a porous skin. The cells of the honeycomb are divided into at least two resonant cavities by an internal partition, which is traversed by



passages interlinking the cavities. The partitions are formed by hollow microbeads that have porous walls.—CJR

6,116,375

43.55.Rg ACOUSTIC RESONATOR

Frederick A. Lorch, Ashland, Massachusetts *et al.*
12 September 2000 (Class 181/224); filed 16 November 1995

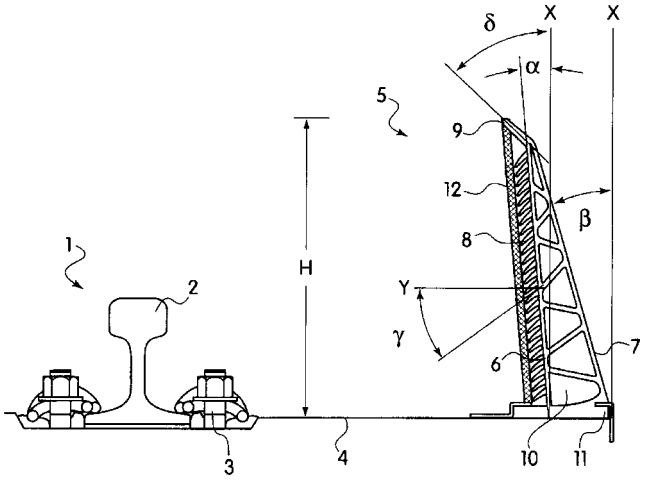
This resonator provides attenuation of sound within a conduit or HVAC duct. The resonating chambers are designed for specific frequencies.—CJR

6,112,849

43.55.Ti SOUND PROTECTION DEVICE FOR RAILROAD TRACK SYSTEM

Norbert Garbers *et al.*, assignors to Phoenix Aktiengesellschaft
5 September 2000 (Class 181/210); filed in Germany 19 February 1997

It is claimed that this sound protection device will reduce the propagation of airborne sound from a rail track. The device includes an elastically deformable, low sound-protecting wall made of rubber or plastic alongside



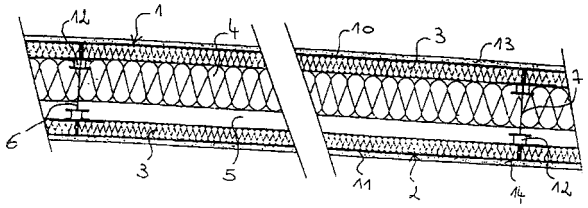
(parallel with) a rail. The wall is designed so that there will be no material damage in the event of any collision. The low wall has inner chambers to increase absorption.—CJR

6,122,867

43.55.Ti ACOUSTIC BUILDING STRUCTURE

Alain Leconte, assignor to Isover Saint-Gobain
26 September 2000 (Class 52/144); filed in France 2 December 1997

The structure emphasizes separation of different masses with insulation in the airspace, thus creating a mass–spring–mass system for better isolation. These rigid outer panels, or trays, are lined on the inside with



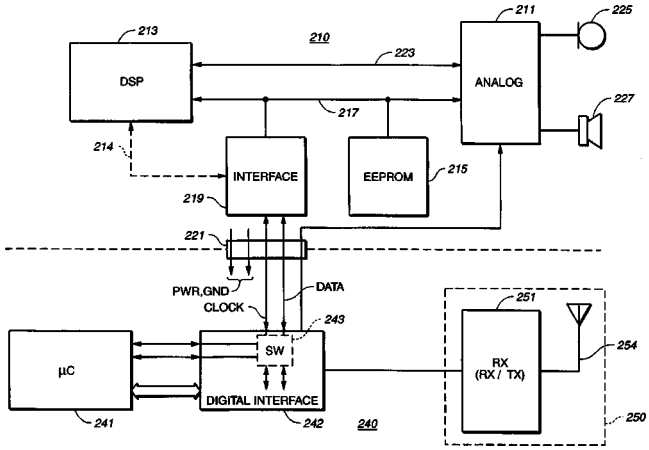
mineral wool, and can be joined together to form a continuous panel for floors, partitions, and ceilings.—CJR

6,144,748

43.66.Ts STANDARD-COMPATIBLE, POWER EFFICIENT DIGITAL AUDIO INTERFACE

Robert Q. Kerns, assignor to Resound Corporation
7 November 2000 (Class 381/312); filed 31 March 1997

Instrumentation is described for interfacing a hearing aid to a direct auditory input auxiliary device and to achieve communication with a remote control. Bi-directional controlling data and audio information are exchanged between the hearing aid and the auxiliary device. In one embodiment, the auxiliary device receives an analog audio signal from a tape player and



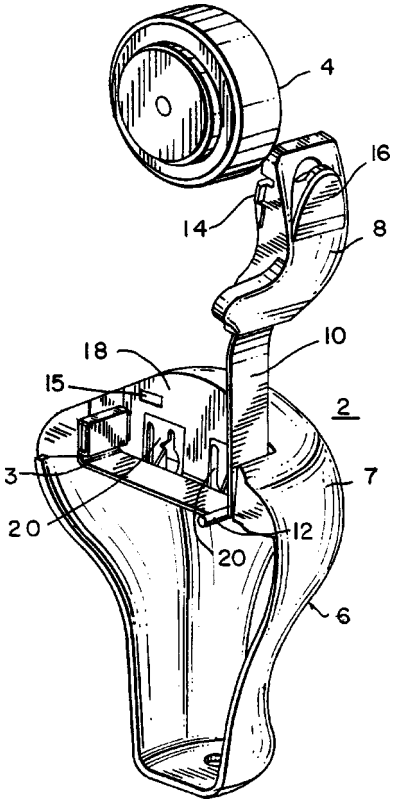
sends it to the hearing aid. In another embodiment, a rf transponder or receiver is connected to a microcontroller-based auxiliary device with input/output information being transmitted digitally, for example to establish a bi-directional communication link to a cellular telephone.—DAP

6,144,749

43.66.Ts HEARING AID FACEPLATE AND BATTERY COMPARTMENT

Brian Fideler, assignor to Starkey Laboratories, Incorporated
7 November 2000 (Class 381/323); filed 8 April 1998

Hearing aid wearers with limited dexterity frequently have difficulty changing hearing aid batteries. A small enclosure system is described that permits easier insertion and removal of batteries by hearing aid wearers. A flexible strip forms a living hinge to connect the battery door with the



hearing aid faceplate. When the door is removed from the faceplate, the flexible strip creates a sling. When the door is closed, the flexible strip fits around the circumference of the battery and folds up.—DAP

6,148,087

43.66.Ts HEARING AID HAVING TWO HEARING APPARATUSES WITH OPTICAL SIGNAL TRANSMISSION THEREBETWEEN

Raimund Martin, assignor to Siemens Augiologische Technik GmbH
14 November 2000 (Class 381/327); filed in Germany 4 February 1997

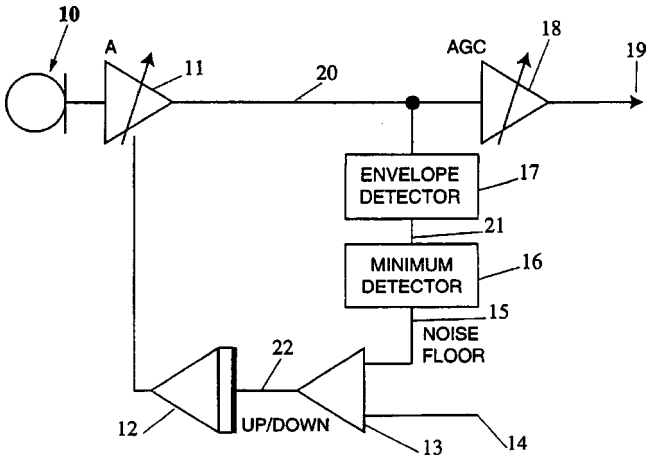
In a hearing aid fitting known as Contralateral Routing of Signals or CROS configuration, the signal from one side of the head is routed to the other side by hard wire, inductive, or rf means. This patent describes an optical method for routing the signals across the head. One possible implementation transmits infra-red energy through a light waveguide contained in an eyeglass frame.—DAP

6,151,400

43.66.Ts AUTOMATIC SENSITIVITY CONTROL

Peter Seligman, assignor to Cochlear Limited
21 November 2000 (Class 381/317); filed in Australia 24 October 1994

Some AGC or compression systems for hearing aids and other auditory prosthetics have distracting audible ''breathing'' sounds caused by sudden gain increases during quiet periods in which no environmental noise is present. A compression system is described in which a minimum detector



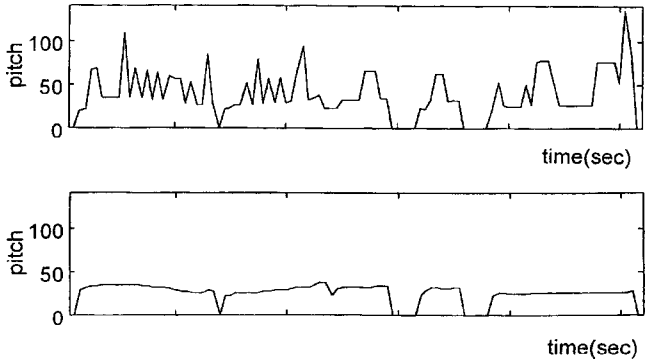
senses and essentially saves the lowest signal amplitude as the noise floor. Thereafter, amplifier gain slowly decreases or increases if the noise floor changes to be greater or less, respectively, than that previously sensed.—DAP

6,119,081

43.72.Ar PITCH ESTIMATION METHOD FOR A LOW DELAY MULTIBAND EXCITATION VOCODER ALLOWING THE REMOVAL OF PITCH ERROR WITHOUT USING A PITCH TRACKING METHOD

Yong-Duk Cho and Moo-young Kim, assignors to Samsung Electronics Company, Limited
12 September 2000 (Class 704/207); filed in Republic of Korea 13 January 1998

The multiband excitation vocoder is known for having good voice output quality over a wide range of bitrate settings. However, it suffers from a long delay time required for pitch processing. This patent describes a method for improving the pitch estimates using a short speech window. An error function is computed in the normal manner over a range of pitch



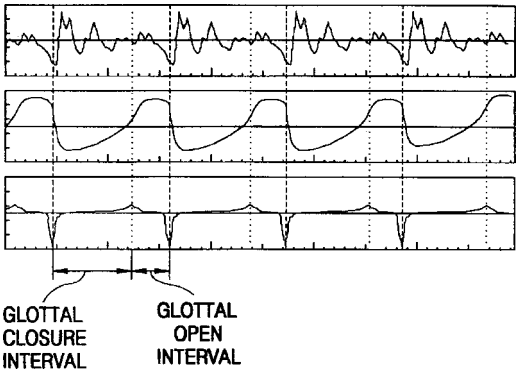
candidate values. This error is then multiplied by a normalized covariance function, greatly reducing the tendency for pitch doubling and pitch halving which tend to occur in short windows. The figures compare normal and modified pitch tracks for 1 s of female speech.—DLR

6,125,344

43.72.Ar PITCH MODIFICATION METHOD BY GLOTTAL CLOSURE INTERVAL EXTRAPOLATION

Dong Gyu Kang *et al.*, assignors to Electronics and Telecommunications Research Institute
26 September 2000 (Class 704/207); filed in Republic of Korea 28 March 1997

A method is disclosed for speech pitch modification based on a determination of the glottal closure interval. Separate vocal tract resonance spectra are measured corresponding to closed glottis and open glottis regions.



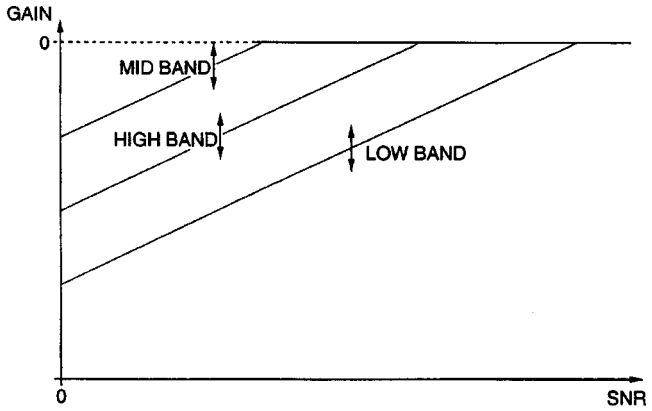
These spectral parameters are extrapolated through the time period of interest to generate the desired pitch structure and glottal closure intervals. The method depends on an electroglottograph signal, as shown in the center figure, recorded simultaneously with the input voice signal.—DLR

6,122,384

43.72.Dv NOISE SUPPRESSION SYSTEM AND METHOD

Anthony P. Mauro, assignor to Qualcomm, Incorporated
19 September 2000 (Class 381/94.3); filed 2 September 1997

This noise reduction system is very similar to previously patented systems which estimate noise levels within individual frequency bands. When no speech signal is detected, a background noise level estimate is recorded and the gain is reduced to a minimum. Speech sounds are filtered



into a number of bands and SNR values are estimated in each band to provide a gain setting for the band. The adjusted bands are recombined to produce the speech output.—DLR

6,119,086

43.72.Gy SPEECH CODING VIA SPEECH RECOGNITION AND SYNTHESIS BASED ON PRE-ENROLLED PHONETIC TOKENS

Abraham Ittycheriah *et al.*, assignors to International Business Machines Corporation
12 September 2000 (Class 704/267); filed 28 April 1998

This patent for a phonetic vocoder joins the list of several similar patents issued recently. It is not clear how this one differs from the others. In this case, the preferred transmission unit is the lefeme, defined as a context-dependent acoustic phone. A preference is noted for the use of an HMM-based recognizer and some variety of *n*-gram waveform segment synthesizer. There is also some discussion of the fact that speaker-dependent approaches at the sending and receiving ends are independent, the choices of which provide some flexibility. There is also a mention of including background noises in the synthesis waveforms, although it is not clear how this would be done.—DLR

6,122,609

43.72.Gy METHOD AND DEVICE FOR THE OPTIMIZED PROCESSING OF A DISTURBING SIGNAL DURING A SOUND CAPTURE

Pascal Scalart and André Gilloire, assignors to France Telecom
19 September 2000 (Class 704/226); filed in France 9 June 1997

Operating in both time and frequency domains, this speech noise reduction processor is said to be able to remove both echoes and uncorrelated noises. The system is designed primarily for use in the interior of a vehicle for cleaning up the speech signal of a hands-free telephone system. Using both correlations and adaptive comparisons of spectral density, the device forms estimates of both the speech signal and the disturbing noise, and adaptively filters the latter from the former.—DLR

6,122,616

43.72.Ja METHOD AND APPARATUS FOR DIPHONE ALIASING

Caroline G. Henton, assignor to Apple Computer, Incorporated
19 September 2000 (Class 704/258); filed 21 January 1993

This patent presents a method of creating or modifying diphone segments dynamically for use by a diphone synthesizer. The two primary applications are to fill in for diphones missing due to memory limitations and to provide synthesis with a flexible time base for use in video lip synch productions. The set of phonemes is characterized using a traditional phonological feature set. When a new diphone is needed, the feature sets of each half-diphone element are used to search for similar half-diphone elements. A set of rules is used in the search process to impose constraints on the types of mismatch allowable in the substitution.—DLR

6,119,085

43.72.Ne RECONCILING RECOGNITION AND TEXT TO SPEECH VOCABULARIES

James R. Lewis and Kerry A. Ortega, assignors to International Business Machines Corporation
12 September 2000 (Class 704/260); filed 27 March 1998

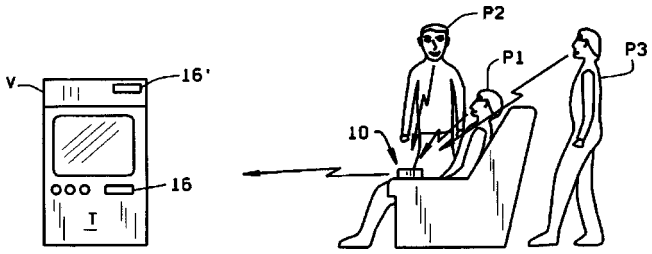
This patent describes a process of reconciling the pronunciations of words used by both recognition and text-to-speech systems, particularly in the case where the two systems may have been provided by different manufacturers. The scheme would involve playing back the recognizer vocabulary items using the TTS system. There is no mention of the basic issue that most such systems would use completely incompatible forms of vocabulary representation.—DLR

6,119,088

43.72.Ne APPLIANCE CONTROL PROGRAMMER USING VOICE RECOGNITION

Gary Ciluffo, Vincennes, Indiana
12 September 2000 (Class 704/275); filed 3 March 1998

This voice-controlled remote unit is designed to operate a device such as a TV or VCR. It is distinguished from earlier voice-operated remotes in the emphasis on speaker identification. Several speakers may be enrolled for



access to the unit and any of their voices will be accepted, while a nonenrolled speaker will be ignored. There is a hierarchical protocol to resolve commands given simultaneously by multiple enrolled speakers.—DLR

6,119,089

43.72.Ne AURAL TRAINING METHOD AND APPARATUS TO IMPROVE A LISTENER'S ABILITY TO RECOGNIZE AND IDENTIFY SIMILAR SOUNDS

Athanassios Protopapas, assignor to Scientific Learning Corporation
12 September 2000 (Class 704/278); filed 20 March 1998

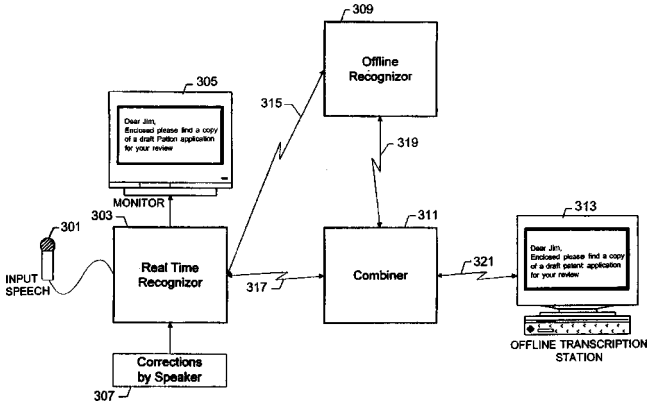
The assignee of this patent has recently been working with individuals having a type of perceptual defect in which similar normal speech sounds cannot be distinguished, but if slowed or otherwise exaggerated, can be. The patent describes a software method of linearly stretching a speech sound in any of several ways in a transform space. During use, the system would present a sequence of sounds constructed with progressively decreasing amounts of stretch, based on feedback from the subject.—DLR

6,122,613

43.72.Ne SPEECH RECOGNITION USING MULTIPLE RECOGNIZERS (SELECTIVELY) APPLIED TO THE SAME INPUT SAMPLE

James K. Baker, assignor to Dragon Systems, Incorporated
19 September 2000 (Class 704/235); filed 30 January 1997

This patent describes a speech recognition system organized to provide the highest possible accuracy in a situation where extra resources may be allocated to the task. Specifically, two or more recognizers may operate on the same speech input. The results are combined by forming a union of the



recognized words and a weighted average of the respective word scores. These results are in a form suitable for use by a human transcriptionist, who would verify the recognition accuracy.—DLR

6,122,614

43.72.Ne SYSTEM AND METHOD FOR AUTOMATING TRANSCRIPTION SERVICES

Jonathan Kahn *et al.*, assignors to Custom Speech USA, Incorporated
19 September 2000 (Class 704/235); filed 20 November 1998

This multi-user speech recognition and transcription system maintains a proficiency level for each user. A dictated voice file to be transcribed is processed according to the user's level. The voice file for a newly enrolled user is sent directly to a human transcriber and the resulting text is sent back to the user and is also used with the voice to train the recognition system. A file from a user-in-training is checked by the human and may be used for further training. Once a certain proficiency level is reached, the recognizer output is returned directly without human intervention. The routing is automatic, based on the user's history and the current recognition results.—DLR

6,122,615

43.72.Ne SPEECH RECOGNIZER USING SPEAKER CATEGORIZATION FOR AUTOMATIC REEVALUATION OF PREVIOUSLY-RECOGNIZED SPEECH DATA

Kenji Yamamoto, assignor to Fujitsu Limited
19 September 2000 (Class 704/252); filed in Japan 19 November 1997

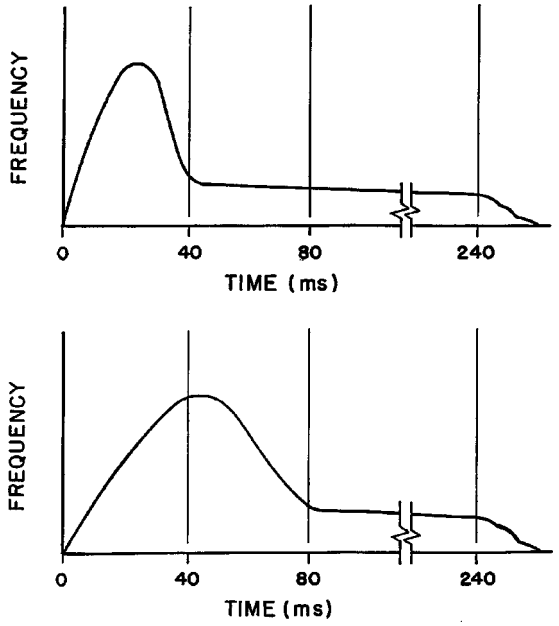
The speech recognition system described here consists of a number of well-known, prior-art recognition components. It is not clear that any of the final arrangements is any more novel than any of the components. The input speech is stored, allowing multiple accesses by the system. In addition to basic speech recognition, a speaker characterization step is also performed, consisting of at least gender identification. These results are fed back to the recognizer, providing better recognition results on a second pass.—DLR

6,123,548

43.72.Ne METHOD AND DEVICE FOR ENHANCING THE RECOGNITION OF SPEECH AMONG SPEECH-IMPAIRED INDIVIDUALS

Paula Anne Tallal *et al.*, assignors to The Regents of the University of California; Rutgers, The State University of New Jersey
26 September 2000 (Class 434/185); filed 8 December 1994

A number of patents have issued recently dealing with ways to help overcome a type of perceptual defect often known as specific language impairment (SLI). The perceptual defect results in the lack of ability to identify sounds with rapid frequency transitions. This patent presents a com-



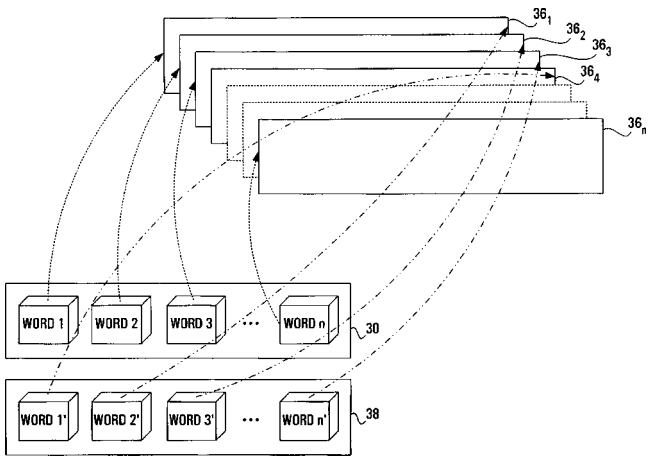
puter system for modifying a speech signal, such as shown in the first figure, to a new signal with slower transitions, as in the second figure. It is said that the process can also help with certain age-related hearing problems and with some adult foreign language learning situations.—DLR

6,125,341

43.72.Ne SPEECH RECOGNITION SYSTEM AND METHOD

Henry F. Raud and Paul M. Brennan, assignors to Nortel Networks Corporation
26 September 2000 (Class 704/8); filed 19 December 1997

This speech recognition system is organized such that a fairly large number of recognition vocabularies can be stored in memory and easily activated for the recognizer. The vocabularies may contain similar sets of



words but spoken with different dialects or different languages. Specific details of the procedure for selecting a vocabulary are not discussed, but it appears that the idea would be to test similar words from multiple vocabularies, pursuing some type of search procedure in order to find the one which most closely matches a sample of input speech.—DLR

6,119,084

43.72.Pf ADAPTIVE SPEAKER VERIFICATION APPARATUS AND METHOD INCLUDING ALTERNATIVE ACCESS CONTROL

Tracy Roberts and Craig A. Will, assignors to Nortel Networks Corporation
12 September 2000 (Class 704/246); filed 29 December 1997

The use of speaker verification techniques over the telephone system has been limited due to the known problems of accuracy in the case of varying speaker environments. The method presented in this patent allows an alternative method of identification, such as touch-tone entry, in the event of voice ID failure. When the caller uses such an alternate, the collected voice record is added to the verification database to improve future accuracy.—DLR

Investigating trends in acoustics research from 1970–1999

John A. Viator
Beckman Laser Institute and Medical Clinic, 1002 Health Sciences Road East, University of California, Irvine, California 92612

F. Michael Pestorius[a)]
Office of Naval Research, International Field Office, 223 Old Marylebone Road, London NW1 5TH, United Kingdom

(Received 23 February 2001; accepted for publication 27 February 2001)

Text data mining is a burgeoning field in which new information is extracted from existing text databases. Computational methods are used to compare relationships between database elements to yield new information about the existing data. Text data mining software was used to determine research trends in acoustics for the years 1970, 1980, 1990, and 1999. Trends were indicated by the number of published articles in the categories of acoustics using the *Journal of the Acoustical Society of America* (JASA) as the article source. Research was classified using a method based on the Physics and Astronomy Classification Scheme (PACS). Research was further subdivided into world regions, including North and South America, Eastern and Western Europe, Asia, Africa, Middle East, and Australia/New Zealand. In order to gauge the use of JASA as an indicator of international acoustics research, three subjects, underwater sound, nonlinear acoustics, and bioacoustics, were further tracked in 1999, using all journals in the INSPEC database. Research trends indicated a shift in emphasis of certain areas, notably underwater sound, audition, and speech. JASA also showed steady growth, with increasing participation by non-US authors, from about 20% in 1970 to nearly 50% in 1999.
[DOI: 10.1121/1.1366711]

PACS numbers: 43.10.Ln [ADP]

I. INTRODUCTION

Text data mining (TDM) is a nascent field of research where computational methods are used to analyze large text databases, yielding previously unknown information.[1–6] Distinct from information retrieval (IR), which is the process of filtering needed information from large, intractable databases, TDM produces new information by discovering relationships between seemingly disparate elements of such databases. Techniques from computational linguistics and statistical and adaptive algorithms are used to extract information that can be used for decision making by managers of science and technology. Specifically, TDM can be used to discover trends and patterns in large data sets for use in decision making.[1,3,7] A TDM software package, TECH OASIS (Search Technology, Inc, Atlanta, GA), is used to evaluate acoustic research trends for the years 1970, 1980, 1990, and 1999. These years are representative of the US–Soviet Union Cold War, since the intention of this research is to determine the changes in the focus of acoustic research over the period of the Cold War. The Journal of the Acoustical Society of America (JASA) is used to determine acoustic research trends over the last 30 years. The total number of JASA articles is fewer than 700 per year and evaluating a single journal provides continuity in this type of analysis. The representative years are chosen to include the period of the Cold War (1970, 1980), the Cold War transition period corresponding to the fall of the Soviet Union (1990), and the most recent full year since the end of the Cold War (1999). No data prior to 1969 is available from the citation database.

A. Limitations on database searches

Limitations in using the scientific citation databases are primarily monetary cost and computational burden. A scientific database search for journal articles, conference proceedings papers, patents, etc. may result in thousand of citations, each costing about $2. For instance, a search for acoustics articles published since 1969 results in 120 145 citations from INSPEC, a citation database for physics, electrical engineering, and computing. A similar search results in 79 911 citations from Ei Compendex, a database that concentrates on engineering topics. Limiting the acoustics search to the years 1999 and 2000, INSPEC yields 6184 citations and Ei Compendex yields 3956, with only a 15% overlap between the two. Thus, the cost of a comprehensive analysis of all acoustics journal articles in INSPEC or Ei Compendex is up to a quarter of a million dollars. The cost of a single year's acoustic citations limited to one of the aforementioned databases is well over $10 000. Additionally, the computational burden associated with the TECH OASIS software is formidable. By focusing the research exclusively on JASA the data sets are limited to fewer than 1000 citations. A single Pentium computer is then capable of the required computation in a reasonable time.

II. MATERIALS AND METHODS

A scientific database search is used to get acoustic citations for the years 1970, 1980, 1990, and 1999, with specific

[a)]Electronic mail: mpestorius@onrifo.navy.mil

data on acoustic subspecialty, country affiliation, institutional affiliation, author name, etc. Each year's publications from JASA are analyzed to specify whether the country affiliation was from US or non-US origin. Additionally, each year's results are analyzed to determine which of seven world regions is the location of the originator. The regions specified are North America (Canada, US, Mexico), South America, Western Europe (non-Soviet bloc European countries), Eastern Europe (including the Soviet Union), Asia, Middle East, Africa, and Australia/New Zealand. The acoustic subspecialties are classified according to PACS, though the PACS are slightly condensed in order to reduce redundant areas. Finally, in order to gauge the usefulness of using JASA as an indicator of worldwide acoustics research, the acoustic subspecialties nonlinear acoustics, bioacoustics, and underwater sound are searched for in the entire INSPEC database for the year 1999.

A. Literature database

The INSPEC database is used for the searches on articles in acoustics. INSPEC is a scientific database established in 1967 by the Institution of Electrical Engineers (IEE). The INSPEC database covers areas of physics and engineering, with specific classifications in physics, electrical and electronic engineering, computers and control, and information technology. All searches are restricted to the Journal of the Acoustical Society of America, except for certain specified areas in 1999 in which the entire database was searched.

B. Physics and astronomy classification scheme (PACS)

The areas of acoustics research are delineated by the Physics and Astronomy Classification Scheme (PACS) prepared by the American Institute of Physics (AIP) and International Council on Scientific and Technical Information (ICSTI). PACS classifies all area of physics and astronomy in a hierarchical scheme for use in journal classifications, among other uses. JASA uses the PACS codes in the journal, though the PACS have changed during the period 1970–1999, with the two most recent versions promulgated in 1991 and 1999.

The AIP included an Acoustics Appendix to PACS in 1975, with updates accepted from ASA from time to time. For the purpose of classifying acoustic subspecialties with respect to their relative importance in the field of acoustics, the PACS codes are slightly condensed to account for certain redundancies. For instance, the PACS code for Underwater Sound in the Acoustics Appendix, A4330, is condensed with the PACS code A9210V, for Underwater Sound in the section Geophysics, Astronomy, and Astrophysics. In most cases these substitutions are obvious, with the exception of the year 1970 when the Acoustics Appendix to PACS did not yet exist. In this case, the acoustics subspecialties are condensed in a manner most likely to correspond with the later years.

Additionally, in order to prevent extremely small acoustics subspecialties (fewer than two or three citations per year) from interfering with the analysis, the analysis is restricted to approximately 90% of journal articles, in favor of the more prolific subspecialties.

C. Data mining software

The analysis is performed by the data mining software, TECH OASIS, developed by Search Technology and the Georgia Tech Research Corporation (GTRC). It is commercially known as Vantage Point (Search Technologies, Inc., Atlanta, GA). TECH OASIS derives from TOAK, a UNIX-based software system that utilizes advanced search algorithms to assist in evaluating and analyzing large text databases for technology forecasting.[8,9] TECH OASIS is designed for technical managers to determine trends and technological and research interdependencies in order to make technology forecasts, which in turn may assist in reaching funding decisions or in optimizing research collaborations. Additionally, such text data mining may be used for literature-based discovery, in which relationships between seemingly unrelated text database elements result in new hypotheses. For example, Swanson *et al.* were successful in finding such a relationship in Raynaud's Syndrome, a vascular disease of the body's extremities, and the use of fish oil as treatment.[1,10–12]

Although TECH OASIS is capable of analyzing a large number of text databases comprising technical journals, conference proceedings, patents, news services, etc., the analysis is restricted to the INSPEC database. The search results are analyzed and the individual journal citations are divided into more primitive field elements, such as titles, keywords, authors, country of origin, PACS, etc. The primitive field elements are then cross correlated to track acoustical research by acoustic subcategory and geographic origin.

Four processing stages for the raw data consisting of JASA citations for the four representative years are performed. First, the affiliations of all papers in each year are determined to be either of US or non-US origin. Second, the number of articles published according to acoustic subspecialty, based on the PACS codes, is determined. Third, the number of papers in each acoustic subspecialty according to

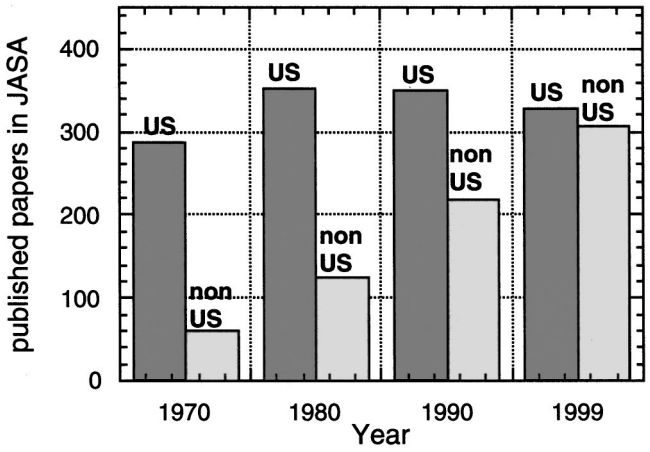


FIG. 1. The numbers of articles in JASA for the years 1970, 1980, 1990, and 1999 are presented here, broken down by US and non-US sources. While the number of US sources remains nearly constant, the percentage of non-US sources increases steadily, from about 17% of total articles in 1970 to 48% in 1999.

TABLE I. The breakdown of acoustic subspecialties for the years 1970, 1980, 1990, and 1999 is shown here. The percentage values indicate the relative number of papers published in JASA in the given subspecialities for that year.

Subspecialty	1970	1980	1990	1999
Audition	21.4	15.4	22.6	24.5
Speech	5.1	9.8	8.1	14.4
Underwater sound	11.8	21.0	18.8	13.5
General linear acoustics	23.7	19.4	11.5	7.4
Structural acoustics	7.8	4.0	5.7	6.0
Acoustic noise	3.6	1.7	1.9	4.9
Ultrasonics, quantum, physical	3.4	1.2	3.6	4.7
Measurement	1.4	2.7	2.9	4.1
Bioacoustics	2.2	0.0	3.8	3.6
Transduction	5.5	2.7	2.2	3.4
Music and musical instruments	0.8	1.7	1.9	3.3
Nonlinear acoustics	2.8	5.2	4.8	3.3
Acoustic signal processing	0.0	7.1	3.8	2.3
Architectural acoustics	2.5	1.5	1.5	2.0
Aeoroacoustics and atmospheric sound	1.0	3.3	2.9	1.4
General	1.9	2.7	3.4	0.0

world region is obtained. These world regions are listed earlier in the Materials and Methods section of this paper. Finally, the entire INSPEC database for the year 1999 is searched in each of the three subspecialties, bioacoustics, nonlinear acoustics, and underwater sound.

III. RESULTS

A. US versus non-US affiliations

The results of the analysis showing publication by US and non-US authors is shown in Fig. 1. The graph shows total number of JASA articles in both categories for the years 1970, 1980, 1990, and 1999.

B. Research areas by year

The breakdown of acoustics articles by research area for the years 1970, 1980, 1990, and 1999 is shown in Table I. The research areas are classified by a method based on PACS.

C. Research areas by world region

The breakdown of acoustics articles by research area for the years 1980 and 1999 are shown in Tables II and III. The ten acoustics subspecialties with the most articles are shown for each year. Less popular subspecialties are not shown and accounted for fewer than ten percent of the total publications for each year.

D. Breadth of coverage of JASA in three acoustics areas in 1999

The number of journal articles in the INSPEC database in 1999 for the acoustics research areas underwater sound, nonlinear acoustics, and bioacoustics is shown in Table IV.

IV. DISCUSSION

The TDM methods reveal information about the shifting nature of acoustics research in the last 30 years, as well as showing the increasingly international status of JASA. While the results from 1970 are suspect, due to the significantly different classification scheme of acoustic subjects, the subsequent years have enough continuity in classifications that a trend in worldwide research is obtained. The caveats for interpretation of the results presented in this paper include the efficacy of using a single journal, JASA, for trend analysis, the adaptation of the PACS for year-by-year comparisons, and, more fundamentally, the use of quantity of published articles as a measure of research activity. While the use of quantity of published articles may be justified, as it is a simple model lending itself to the most deterministic analysis, the use of JASA and the modified PACS schemes must be kept in mind for any interpretation of these results.

A. US versus non-US affiliations

The two most obvious trends uncovered in this paper are shown in Fig. 1. The number of papers published in JASA increases steadily in the period shown. Additionally, the percentage of non-US affiliations also increases steadily from 17.3% in 1970 to 48.2% in 1999. The number of US sources remains nearly constant at about 300–350 per year, and the world region analysis shows that most of the increase comes from Western Europe and Asia, primarily the former. This increase comes about even with the existence of a prominent European acoustics journal, Acta Acustica, and the Chinese Journal of Acoustics.

B. Research areas by year

Table I shows trends for several acoustic categories, though most areas have too few contributions to make any

TABLE II. The breakdown of acoustic subspecialities by world region in 1980.

Subspecialty	Total	North America	South America	Western Europe	Eastern Europe	Asia	Australia N. Zealand	Africa	Middle East
Underwater sound	101	86	1	7		3	4		
General linear acoustics	93	70		17		3	1		2
Audition	74	62		10			2		
Speech	47	43		4					
Acoustic Signal Processing	34	27		2		5			
Nonlinear acoustics	25	16		3	1	5			
Structural acoustics	19	6	1	3		8			1
Atmospheric sound	16	10		3		2		1	
Measurement	13	10			1	2			
General	12	10		2					

TABLE III. The breakdown of acoustic subspecialties by world region in 1999.

Subspecialty	Total	North America	South America	Western Europe	Eastern Europe	Asia	Australia N. Zealand	Africa	Middle East
Audition	156	92		53	2	4	4		1
Speech	92	49	2	33	1	3	3		1
Underwater sound	85	64		17	1	3			
General linear acoustics	47	19	1	18	1	5	2		1
Structural acoustics	38	16		10		9	2		1
Acoustic noise	31	14		10		3	4		
Ultra., Quant., Physical	30	19		5	2	4			
Measurement	26	16		4		5	1		
Bioacoustics	23	19		3			1		
Transduction	22	12		4		6			
Music	19	6		9		3	1		

statistically significant conclusions. Audition increases from 1980 to 1999, while speech shows a large increase in 1999. This increase corresponds to a commensurate decrease in underwater sound from 1980 to 1999. The anomalously low percentage in underwater sound in 1970 may be due to the different classification scheme. General linear acoustics shows a steady decline over the entire 30-year period. This phenomenon may be due to the depletion of research areas in linear acoustics. The nearly indiscernible increases in other subjects, such as bioacoustics and ultrasonics, quantum, and physical acoustics, may indicate increasing diversity in acoustic research, away from linear acoustics.

C. Research area by region

The results for the years 1970 and 1990 are not tabulated, as the results for 1970 are difficult to reconcile with the later years which use the Acoustics Appendix for PACS. The results for 1990 are not shown as they differ little from the trends shown by the differences between 1980 and 1999. The small numbers of articles in South America, Eastern Europe, Africa, the Middle East, and Australia/New Zealand make firm conclusions about any trends in these regions highly unreliable. The only conclusion is that these regions show an increase in submission to JASA, corresponding to the increase in non-US sources shown in Fig. 1. The trend in Western Europe, Asia, and North America also points to increased international participation in JASA and additional information about acoustic categories confirms the trends about shifting research focus areas.

TABLE IV. The relative number of publications in three acoustic subspecialties in 1999. IEEE, AIP, and Elsevier are journal publishers and the indicated results for these entries are probably from more than one journal, though INSPEC did not report the individual journal titles in these cases.

Journal title	Underwater sound	Bioacoustics	Nonlinear acoustics
JASA	85	23	21
Chinese J. of Acoustics	26	0	9
IEEE	31	41	5
Acoustical Physics (MAIK)	28	1	17
AIP	0	3	16
Acta Acustica (European J. Acoustics)	6	0	2
Elsevier	6	9	18

D. Breadth of coverage in acoustics research by JASA

All results in this paper hinge on using JASA as an indicator of worldwide acoustics research. In 1999, nearly half of all articles are from non-US sources, indicating that JASA is a reliable indicator, though several international acoustics journals exist, including Acta Acustica, Acoustical Physics (Russia), and the Chinese Journal of Acoustics. Additionally, the American Institute of Physics (AIP) and the Institute for Electrical and Electronics Engineers (IEEE) publish journals with significant acoustics components. Table IV shows that for the acoustic categories of nonlinear acoustics and underwater sound, JASA published more articles than any other source. For bioacoustics, IEEE published more, though the number was included in more than one IEEE journal.

V. CONCLUSIONS

Keeping in mind the caveats of TDM and of the use of JASA as an indicator of international acoustic research trends, JASA is shown to be an international journal of acoustics research. A shift in research focus away from underwater sound and general linear acoustics and toward speech, audition, and, to a lesser extent, bioacoustics is documented over the period 1980 to 1999. The slightly shifting focus of other research areas is also documented, though a more comprehensive, and hence more expensive, analysis is required to make meaningful pronouncements about these areas.

ACKNOWLEDGMENTS

The help of Webb Meyers, ONR-IFO, London is acknowledged for his expertise in scientific database searches. Captain D. Gray and LCDR E. S. Meadows of the Office of Naval Research are also thanked for their support of this project. Insightful conversation with D. W. Oard and R. N. Kostoff were important in the writing of this paper. P. Frey of Search Technology, Inc, kindly supplied us with his expertise in the design and use of TECH OASIS. ONR Naval Reserve Science and Technology Program (Program 38 Research) funded this project.

[1] M. A. Hearst, ''Untangling text data mining,'' Proceedings of the ACL '99, 37th Annual Meeting of the Association for Computational Linguistics, University of Maryland (1999).
[2] R. N. Kostoff, H. J. Eberhart, and D. R. Toothman, ''Hypersonic and supersonic flow roadmaps using bibliometrics and database tomography,'' J. Am. Soc. Info. Sci. **50**, 427–447 (1999).
[3] R. N. Kostoff, T. Braun, A. Schubert, D. R. Toothman, and J. A. Humenik, ''Fullerene data mining using bibliometrics and database tomography,'' J. Chem. Inf. Comp. Sci. **40**, 19–39 (2000).
[4] R. N. Kostoff and R. A. DeMarco, ''Science and technology data mining,'' Anal. Chem. (submitted by invitation).
[5] R. N. Kostoff, D. R. Toothman, H. J. Eberhart, and J. A. Humenik, ''Text mining using database tomography and bibiometrics: A review,'' Tech. Forecasting Soc. Change (submitted).
[6] P. Losiewicz, D. W. Oard, and R. N. Kostoff, ''Textual data mining to support science and technology management,'' J. Intel. Info. Syst. **15**, 99–119 (2000).
[7] T. Braun, A. P. Schubert, and R. N. Kostoff, ''Growth and trends of fullerene research as reflected in its journal literature,'' Chem. Rev. **100**, 23–37 (2000).
[8] R. J. Watts and A. L. Porter, ''Innovation forecasting,'' Tech. Forecasting Soc. Change **56**, 25–47 (1997).
[9] A. L. Porter and M. J. Detampel, ''Technology opportunities analysis,'' Tech. Forecasting Soc. Change **49**, 237–255 (1995).
[10] D. R. Swanson, ''Fish oil, Raynaud's syndrome, and undiscovered public knowledge,'' Perspect. Biol. Med. **30**, 7–18 (1986).
[11] D. R. Swanson and N. R. Smalheiser, ''An interactive system for finding complementary literatures: a stimulus to scientific discovery,'' Artif. Intel. **91**, 183–203 (1997).
[12] M. D. Gordon and R. K. Lindsay, ''Toward discovery support systems: a replication, re-examination, and extension of Swanson's work on literature-based discovery of a connection between Raynaud's and fish oil,'' J. Am. Soc. Info. Sci. **47**, 116–128 (1996).

Scattering of elastic waves in heterogeneous media with local isotropy

Joseph A. Turner and Phanidhar Anugonda
Department of Engineering Mechanics, W317.4 Nebraska Hall, University of Nebraska–Lincoln, Lincoln, Nebraska 68588-0526

(Received 14 August 2000; revised 20 February 2001; accepted 23 February 2001)

The scattering of elastic waves in heterogeneous media is discussed. Explicit expressions are derived for the attenuation of longitudinal and transverse elastic waves in terms of the statistics of the density and Lamé parameter fluctuations. The derivation is based upon diagrammatic methods with the problem posed in terms of the Dyson equation. The Dyson equation is solved for the mean field response. The results are given here in a straightforward manner, in which the attenuations reduce to simple integrals on the unit circle. The medium is assumed statistically homogeneous and statistically isotropic. This model, with assumed local isotropic properties, is expected to apply to many materials. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1367245]

PACS numbers: 43.20.Bi, 43.20.Gp, 43.35.Cg [DEC]

I. INTRODUCTION

The study of wave propagation and scattering of elastic waves in heterogeneous media is related to nondestructive testing, materials characterization, acoustic emission, and seismic wave analysis. An improved understanding of the effects of scattering provides insight into the microstructure of a variety of materials such as polycrystalline metals and ceramics, composites, geophysical materials, and concrete. Elastic waves which propagate through such media lose energy due to scattering from the heterogeneous structure of the material. The scattering effects may be characterized by the attenuation. Previous research on scattering problems of this nature has been dominated by studies of polycrystalline materials.[1–4] In these models, it is assumed that density fluctuations are negligible and that the material is locally anisotropic. The grains are usually assumed to be randomly oriented, such that the medium is statistically isotropic. More general cases have also been examined[5–7] in which the grains have some prescribed alignment (texture) such that the medium is statistically anisotropic. This research, in which expressions for attenuation were derived, was also extended to derivations of elastic radiative transfer equations (RTE) and diffusion equations which describe the evolution of diffuse elastic energy.[3,4,8,9] More recent developments include the derivation of the attenuations and elastic RTE using an asymptotic approach.[10,11] In those articles, a different microstructural model was used. The elastic moduli were assumed to be locally isotropic and the density was considered to vary spatially. The medium was assumed to be statistically homogeneous and statistically isotropic as well. Such a model is expected to be reasonable for geophysical materials and concrete. They derived the differential scattering cross sections, elastic radiative transfer equations, and the elastic diffusion equation.

In this article, the same microstructural model based on local material isotropy is used to derive elastic wave attenuations. The derivation is based upon the diagrammatic approach[3,12] in which the mean response is governed by the Dyson equation. The Dyson equation is easily solved in spatial Fourier transform domain within the limits of the first-order smoothing approximation (FOSA),[12] or Keller[13] approximation. A further approximation is also made which restricts the results to frequencies below the high-frequency geometric optics limit. This high-frequency limit, in which refracted ray analysis must be used,[2] is above the range of most ultrasonic experiments. With this approximation, the attenuations for the longitudinal and transverse elastic waves reduce to simple integrations on the unit circle. The results here are in basic agreement with those of Ryzhik *et al.*[10] Therefore, their asymptotic method is presumed to be equivalent to the FOSA with the additional frequency limitation.

In the next section, the theoretical model is presented in terms of the Dyson equation. The Dyson equation is solved and expressions for the attenuations derived. Then expressions for the mean free paths and elastic diffusivity are presented. Finally, further assumptions of the form of the fluctuations are made and example calculations are presented.

II. MEAN RESPONSE

The equation of motion for the elastodynamic response of a linear, elastic material to deformation is given in terms of the Green's dyadic by

$$\left\{-\delta_{jk}\rho(\mathbf{x})\frac{\partial^2}{\partial t^2}+\frac{\partial}{\partial x_i}C_{ijkl}(\mathbf{x})\frac{\partial}{\partial x_l}\right\}G_{k\alpha}(\mathbf{x},\mathbf{x}';t) = \delta_{j\alpha}\delta^3(\mathbf{x}-\mathbf{x}')\,\delta(t), \tag{1}$$

where $\delta^3(\mathbf{x}-\mathbf{x}')$ is the three-dimensional spatial Dirac delta function. The Green's dyadic, $G_{ij}(\mathbf{x},\mathbf{x}';t)$, is the response at location $\mathbf{x}$ in the ith direction due to an impulsive force applied at $\mathbf{x}'$ in the jth direction. In Eq. (1), $\rho(\mathbf{x})$ and $C_{ijkl}(\mathbf{x})$ define the material density and elastic modulus tensor, respectively. These material properties are assumed to vary spatially.

A spatio-temporal Fourier transform pair is defined as

$$\tilde{f}(\mathbf{p},\omega)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}f(\mathbf{x},t)e^{i\omega t}e^{-i\mathbf{x}\cdot\mathbf{p}}\,d^3x\,dt, \tag{2}$$

$$f(\mathbf{x},t)=\frac{1}{(2\pi)^4}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\tilde{f}(\mathbf{p},\omega)e^{-i\omega t}e^{i\mathbf{x}\cdot\mathbf{p}}\,d^3p\,d\omega. \tag{3}$$

This transform pair defines the relation between space–time variables ($\mathbf{x}$ and t) and wave vector-angular frequency variables ($\mathbf{p}$ and ω).

The temporal transform of the equation of motion, Eq. (1), is then

$$\left\{\omega^2\delta_{jk}\rho(\mathbf{x})+\frac{\partial}{\partial x_i}C_{ijkl}(\mathbf{x})\frac{\partial}{\partial x_l}\right\}G_{k\alpha}(\mathbf{x},\mathbf{x}';\omega)=\delta_{j\alpha}\delta^3(\mathbf{x}-\mathbf{x}'). \tag{4}$$

The material properties of the medium are assumed to vary spatially. The density is written

$$\rho(\mathbf{x})=\overline{\rho}(1+\delta\rho(\mathbf{x})), \tag{5}$$

where $\overline{\rho}$ is the average density and $\delta\rho(\mathbf{x})$ is a dimensionless measure of the density fluctuations.

Previous wave propagation studies of polycrystalline materials have used a locally anisotropic model which accounts for the crystal anisotropy.[2,3,7] Here, the elastic modulus tensor C_{ijkl} is assumed to be locally isotropic. It is written as

$$C_{ijkl}(\mathbf{x})=\overline{\lambda}(1+\delta\lambda(\mathbf{x}))\delta_{ij}\delta_{kl}+\overline{\mu}(1+\delta\mu(\mathbf{x}))\times(\delta_{ik}\delta_{jl}+\delta_{il}\delta_{jk}), \tag{6}$$

where $\overline{\lambda}$ and $\overline{\mu}$ are the average Lamé parameters. The elastic moduli fluctuations are defined by the dimensionless measures $\delta\lambda(\mathbf{x})$ and $\delta\mu(\mathbf{x})$. The material properties are assumed to be centered random processes such that $\langle\delta\rho(\mathbf{x})\rangle=\langle\delta\lambda(\mathbf{x})\rangle=\langle\delta\mu(\mathbf{x})\rangle=0$, where the brackets $\langle\,\rangle$ denote an ensemble average.

The derivation which follows also requires second-order statistics of the fluctuations. The covariance of the density is defined as

$$R_{\rho\rho}(\mathbf{y}-\mathbf{z})=\langle\delta\rho(\mathbf{y})\,\delta\rho(\mathbf{z})\rangle. \tag{7}$$

Similar definitions for the covariance (autocorrelation) of Lamé parameters and the cross correlations between different parameters are also made. The average medium is assumed statistically isotropic and statistically homogeneous. These assumptions imply that the correlation functions depend only on the magnitude of the difference of the two positions. Mathematically, these assumptions imply that $R(\mathbf{y}-\mathbf{z})=R(r)$, where $r=|\mathbf{y}-\mathbf{z}|$.

The bare Green's dyadic, $\mathbf{G}^0$, is defined as the solution to Eq. (4) when the fluctuations of all material properties are zero. It is the solution to

$$\left\{\omega^2\delta_{jk}\overline{\rho}+(\overline{\lambda}+\overline{\mu})\frac{\partial}{\partial x_j}\frac{\partial}{\partial x_k}+\overline{\mu}\delta_{jk}\frac{\partial^2}{\partial x_l^2}\right\}G^0_{k\alpha}(\mathbf{x},\mathbf{x}';\omega)=\delta_{j\alpha}\delta^3(\mathbf{x}-\mathbf{x}'). \tag{8}$$

A spatial Fourier transform, as defined in Eq. (2), allows Eq. (8) to be reduced to

$$\{\hat{p}_j\hat{p}_k(\overline{\rho}\omega^2-p^2(\overline{\lambda}+2\overline{\mu}))+(\delta_{jk}-\hat{p}_j\hat{p}_k)\times(\overline{\rho}\omega^2-p^2\overline{\mu})\}G^0_{k\alpha}(\mathbf{p})=\delta_{j\alpha}. \tag{9}$$

The solution for $\mathbf{G}^0$ is given by inspection as

$$\mathbf{G}^0(\mathbf{p})=g^0_L(p)\hat{\mathbf{p}}\hat{\mathbf{p}}+g^0_T(p)(\hat{\mathbf{p}}_2\hat{\mathbf{p}}_2+\hat{\mathbf{p}}_3\hat{\mathbf{p}}_3), \tag{10}$$

for propagation in the $\hat{\mathbf{p}}$ direction. The unit vectors $\hat{\mathbf{p}}_2$ and $\hat{\mathbf{p}}_3$ are transverse to the direction $\hat{\mathbf{p}}$, and form an orthonormal basis with $\hat{\mathbf{p}}$. The bare longitudinal and transverse propagators, which appear in Eq. (10), are

$$g^0_L(p)=[\overline{\rho}\omega^2-p^2(\overline{\lambda}+2\overline{\mu})]^{-1}=[\overline{\rho}\omega^2-p^2\overline{\rho}c_L^2]^{-1}, \tag{11}$$

$$g^0_T(p)=[\overline{\rho}\omega^2-p^2\overline{\mu}]^{-1}=[\overline{\rho}\omega^2-p^2\overline{\rho}c_T^2]^{-1}, \tag{12}$$

where average wave speeds are defined in terms of the average material properties ($\overline{\rho}c_L^2=\overline{\lambda}+2\overline{\mu}$, $\overline{\rho}c_T^2=\overline{\mu}$). The imaginary parts of the propagators, which are used below, are given by

$$\mathrm{Im}\,g^0_L(p)=-\frac{\pi}{\overline{\rho}}\mathrm{sgn}(\omega)\,\delta(\omega^2-p^2c_L^2), \tag{13}$$

$$\mathrm{Im}\,g^0_T(p)=-\frac{\pi}{\overline{\rho}}\mathrm{sgn}(\omega)\,\delta(\omega^2-p^2c_T^2). \tag{14}$$

Studies of wave propagation in heterogeneous materials do not lend themselves to solution by perturbation methods. Solutions of this sort do not converge.[12] Instead, Frisch used diagrammatic methods for solution of the mean response. The mean response, $\langle\mathbf{G}\rangle$, is governed by the Dyson equation which is given by[3,7,12]

$$\langle G_{i\alpha}(\mathbf{x},\mathbf{x}')\rangle=G^0_{i\alpha}(\mathbf{x},\mathbf{x}')+\int\int G^0_{i\beta}(\mathbf{x},\mathbf{y})M_{\beta j}(\mathbf{y},\mathbf{z})\times\langle G_{j\alpha}(\mathbf{z},\mathbf{x}')\rangle d^3y\,d^3z. \tag{15}$$

In Eq. (15), the quantity $\mathbf{G}^0$ is the bare Green's dyadic which was defined in Eq. (10). The second-rank tensor $\mathbf{M}$ is the mass or self-energy operator.[12] The Dyson equation, Eq. (15), is easily solved in Fourier transform space under the assumption of statistical homogeneity. The spatial Fourier transform pair for $\mathbf{G}^0$ is given by

$$G^0_{i\alpha}(\mathbf{p})\,\delta^3(\mathbf{p}-\mathbf{q})=\frac{1}{(2\pi)^3}\int\int G^0_{i\alpha}(\mathbf{x},\mathbf{x}')\times e^{-i\mathbf{p}\cdot\mathbf{x}}e^{i\mathbf{q}\cdot\mathbf{x}'}\,d^3x\,d^3x', \tag{16}$$

$$G^0_{i\alpha}(\mathbf{x},\mathbf{x}')=\frac{1}{(2\pi)^3}\int\int G^0_{i\alpha}(\mathbf{p})\,\delta^3(\mathbf{p}-\mathbf{q})\times e^{i\mathbf{p}\cdot\mathbf{x}}e^{-i\mathbf{q}\cdot\mathbf{x}'}\,d^3p\,d^3q. \tag{17}$$

The Fourier transforms which define $\langle\mathbf{G}(\mathbf{p})\rangle$ and $\tilde{\mathbf{M}}(\mathbf{p})$ are given by expressions similar to that defining $\mathbf{G}^0(\mathbf{p})$. The assumption of statistical homogeneity ensures that they are functions of a single wave vector in Fourier space.

The Dyson equation, Eq. (15), is then spatially Fourier transformed and solved for $\langle\mathbf{G}(\mathbf{p})\rangle$. The result is

$$\langle\mathbf{G}(\mathbf{p})\rangle=[\mathbf{G}^0(\mathbf{p})^{-1}-\tilde{\mathbf{M}}(\mathbf{p})]^{-1}. \tag{18}$$

The Dyson equation is exact and describes the mean response of the medium. The main difficulty in the solution of Eq. (18) is the representation of $\tilde{\mathbf{M}}$. Approximations of $\tilde{\mathbf{M}}$ are often necessary for closed-form solutions of Eq. (18) to be obtained. The self-energy, $\mathbf{M}$, can be written as an expansion in powers of material property fluctuations. Approximation of $\mathbf{M}$ can then be made to first order using the first term in such an expansion. Frisch discusses the equivalence of this technique, which he called the first-order smoothing approximation (FOSA),[12] and the Keller approximation.[13] Such an approximation is valid if the fluctuations of the material properties are small ($\delta\rho(\mathbf{x}) \ll 1$, $\delta\lambda(\mathbf{x}) \ll 1$, $\delta\mu(\mathbf{x}) \ll 1$). To this level of approximation, the self-energy is given by

$$M_{\beta j}(\mathbf{y},\mathbf{z}) \approx \langle \mathfrak{L}^1_{\beta\gamma}(\mathbf{y}) G^0_{\gamma k}(\mathbf{y},\mathbf{z}) \mathfrak{L}^1_{kj}(\mathbf{z}) \rangle, \tag{19}$$

where $\mathfrak{L}^1$ is the first-order operator.[12]

The first-order operator for the problem studied here follows from Eqs. (1), (5), and (6). It is given by

$$\mathfrak{L}^1_{\beta\gamma}(\mathbf{y}) = \omega^2 \delta_{\beta\gamma} \overline{\rho} \delta\rho(\mathbf{y}) + \overline{\lambda} \frac{\partial}{\partial y_\beta} \delta\lambda(\mathbf{y}) \frac{\partial}{\partial y_\gamma} + \overline{\mu} \frac{\partial}{\partial y_\gamma} \delta\mu(\mathbf{y}) \frac{\partial}{\partial y_\beta} + \overline{\mu} \delta_{\beta\gamma} \frac{\partial}{\partial y_l} \delta\mu(\mathbf{y}) \frac{\partial}{\partial y_l}. \tag{20}$$

The spatial transform of the self-energy and the mean Green's dyadic will have the same form as the bare Green's dyadic. They are written

$$\tilde{\mathbf{M}}(\mathbf{p}) = m_L(p) \hat{\mathbf{p}}\hat{\mathbf{p}} + m_T(p)(\hat{\mathbf{p}}_2\hat{\mathbf{p}}_2 + \hat{\mathbf{p}}_3\hat{\mathbf{p}}_3), \tag{21}$$

and

$$\langle \tilde{\mathbf{G}}(\mathbf{s}) \rangle = g_L(s) \hat{\mathbf{s}}\hat{\mathbf{s}} + g_T(s)(\hat{\mathbf{s}}_2\hat{\mathbf{s}}_2 + \hat{\mathbf{s}}_3\hat{\mathbf{s}}_3). \tag{22}$$

The propagators for the mean response are then given by the solution of the Dyson equation, Eq. (18), as

$$g_\beta(\mathbf{p}) = [g^0_\beta(\mathbf{p})^{-1} - m_\beta(\mathbf{p})]^{-1} = [\overline{\rho}\omega^2 - p^2 \overline{\rho} \overline{c}^2_\beta - m_\beta(\mathbf{p})]^{-1}, \tag{23}$$

for each wave type, β (L or T).

The expressions for the propagators of the mean response define the phase velocity and attenuation of each wave type. The solution of

$$\overline{\rho}\omega^2 - p^2 \overline{\rho} \overline{c}^2_\beta - m_\beta(\mathbf{p}) = 0, \tag{24}$$

for the wave vector p, is required, given $\tilde{\mathbf{M}}$ defined in Eq. (19).

The inverse Fourier transform of $\langle \mathbf{G}(\mathbf{p}) \rangle$ will be dominated by the zeros of the propagators. The phase velocity is given by the real part of p and the attenuation by the imaginary part. Such solutions of Eq. (24) are often done numerically using root finding techniques.[2] However, explicit expressions for the attenuation can be determined using an approximation valid below the high-frequency geometric optics limit. In this case, the wave vector, $\mathbf{p}$, within the self-energy is approximated as being equal to the bare wave vector. Such an approximation, $m(\mathbf{p}) \approx m(\omega\hat{\mathbf{p}}/c_\beta)$, is sometimes called a Born approximation.[2,3] In essence, the phase velocity is assumed to remain unchanged by the heterogeneities. This approximation allows the imaginary part of $\mathbf{p}$ to be calculated directly from Eq. (24). The attenuation of each wave type, to this level of approximation, is given by

$$\alpha_\beta(\hat{\mathbf{p}}) = -\frac{1}{2\overline{\rho}\omega c_\beta} \operatorname{Im} m_\beta\left(\frac{\omega}{c_\beta}\hat{\mathbf{p}}\right), \tag{25}$$

where the subscript β refers to either wave type (L or T). Thus the determination of the attenuations requires the imaginary part of the components of the spatial transform of the self-energy. The self-energy, given by Eq. (19), is a product of the bare Green's dyadic, $\mathbf{G}^0$, and two multiples of the first-order operator, $\mathfrak{L}^1$, defined in Eq. (20). This operator contains terms related to the density and Lamé parameter fluctuations. Therefore, the entire self-energy is a sum of six terms—a $\rho\rho$ term and five others ($\lambda\lambda$, $\mu\mu$, $\rho\lambda$, $\rho\mu$, and $\lambda\mu$).

The spatial double Fourier transform of the self-energy is given by

$$\tilde{M}_{\beta j}(\mathbf{p})\,\delta^3(\mathbf{p}-\mathbf{q}) = \frac{1}{(2\pi)^3} \int\int d^3y\, d^3z e^{-i\mathbf{p}\cdot\mathbf{y}} \langle \mathfrak{L}^1_{\beta\gamma}(\mathbf{y}) \times G^0_{\gamma k}(\mathbf{y},\mathbf{z}) \mathfrak{L}^1_{kj}(\mathbf{z}) \rangle e^{i\mathbf{q}\cdot\mathbf{z}}. \tag{26}$$

Example calculations for two of the terms of $\tilde{\mathbf{M}}$ are given here explicitly. The other terms follow from similar derivations. The first term comes from the density terms in each of the first-order operators. It is given by

$$\tilde{M}^{\rho\rho}_{\beta j}(\mathbf{p})\,\delta^3(\mathbf{p}-\mathbf{q}) = \frac{\overline{\rho}^2\omega^4}{(2\pi)^3} \int\int d^3y\, d^3z e^{-i\mathbf{p}\cdot\mathbf{y}} G^0_{\beta j}(\mathbf{y},\mathbf{z}) \times \langle \delta\rho(\mathbf{y})\,\delta\rho(\mathbf{z}) \rangle e^{i\mathbf{q}\cdot\mathbf{z}}. \tag{27}$$

The spatial Fourier transforms of correlation functions (the power spectra of the fluctuations) are defined as

$$\tilde{R}(\mathbf{p}) = \frac{1}{(2\pi)^3} \int d^3r R(\mathbf{r}) e^{-i\mathbf{p}\cdot\mathbf{r}}. \tag{28}$$

This definition allows the $\rho\rho$ term of the self-energy, Eq. (27), to reduce to

$$\tilde{M}^{\rho\rho}_{\beta j}(\mathbf{p}) = \overline{\rho}^2\omega^4 \int d^3s \tilde{G}^0_{\beta j}(\mathbf{s}) \tilde{R}_{\rho\rho}(\mathbf{p}-\mathbf{s}). \tag{29}$$

Thus, the terms of the transform of the self-energy may be written as convolutions between the bare Green's dyadic and the power spectra.

The term related to the $\lambda\lambda$ covariance is given by

$$\tilde{M}^{\lambda\lambda}_{\beta j}(\mathbf{p})\,\delta^3(\mathbf{p}-\mathbf{q}) = \frac{\overline{\lambda}^2}{(2\pi)^3} \int\int d^3y\, d^3z e^{-i\mathbf{p}\cdot\mathbf{y}} \times \left\langle \frac{\partial}{\partial y_\beta} \delta\lambda(\mathbf{y}) \frac{\partial}{\partial y_\gamma} G^0_{\gamma k}(\mathbf{y},\mathbf{z}) \frac{\partial}{\partial z_j} \delta\lambda(\mathbf{z}) \frac{\partial}{\partial z_k} \right\rangle e^{i\mathbf{q}\cdot\mathbf{z}}. \tag{30}$$

Integration by parts yields

$$\tilde{M}^{\lambda\lambda}_{\beta j}(\mathbf{p}) = \overline{\lambda}^2 \int d^3s \tilde{G}^0_{\gamma k}(\mathbf{s}) \tilde{R}_{\lambda\lambda}(\mathbf{p}-\mathbf{s}) p_\beta p_j s_\gamma s_k. \tag{31}$$

The remaining calculations follow in a similar manner. The total self-energy is finally expressed as

$$\tilde{M}_{\beta j}(\mathbf{p}) = \int d^3 s \tilde{G}^0_{\gamma k}(\mathbf{s})\{\overline{\rho}^2\omega^4\delta_{jk}\delta_{\beta\gamma}\tilde{R}_{\rho\rho}(\mathbf{p}-\mathbf{s}) + \overline{\lambda}^2\tilde{R}_{\lambda\lambda}(\mathbf{p}-\mathbf{s})p_\beta p_j s_\gamma s_k + \overline{\mu}^2(s_\beta s_j p_\gamma p_k + s_\beta p_\gamma p_l s_l \delta_{jk} + \delta_{\beta\gamma}p_l s_l s_j p_k$$
$$+ \delta_{\beta\gamma}\delta_{jk}p_l s_l p_m s_m)\tilde{R}_{\mu\mu}(\mathbf{p}-\mathbf{s}) + \overline{\rho\lambda}\omega^2(\delta_{\beta\gamma}p_j s_k + \delta_{jk}p_\beta s_\gamma)\tilde{R}_{\rho\lambda}(\mathbf{p}-\mathbf{s}) + \overline{\rho\mu}\omega^2[\delta_{\beta\gamma}(s_j p_k + p_l s_l \delta_{jk})$$
$$+ \delta_{jk}(s_\beta p_\gamma + p_l s_l \delta_{\beta\gamma})]\tilde{R}_{\rho\mu}(\mathbf{p}-\mathbf{s}) + \overline{\lambda\mu}(p_\beta s_j p_k s_\gamma + p_\beta s_\gamma p_l s_l \delta_{jk} + s_\beta p_j s_k p_\gamma + p_j s_k p_l s_l \delta_{\beta\gamma})\tilde{R}_{\lambda\mu}(\mathbf{p}-\mathbf{s})\}. \quad (32)$$

The imaginary part of this final expression for the transform of the self-energy allows the attenuations to be determined. The forms for the self-energy, Eq. (21), and the bare Green's dyadic, Eq. (10), are used in simplifying the attenuations. These expressions Eqs. (13)–(14) are substituted into Eq. (32). Appropriate inner products of the resulting equation allow the required components, $m_L(\mathbf{p})$ and $m_T(\mathbf{p})$, to be determined. The required quantity for the longitudinal attenuation is given by

$$\mathrm{Im}\, m_L(\mathbf{p}) = -\frac{\pi}{\overline{\rho}}\frac{1}{2\omega}\hat{p}_\beta\hat{p}_j\int d^3 s\left(\frac{1}{c_L}\delta(s-\omega/c_L)\hat{s}_\gamma\hat{s}_k + \frac{1}{c_T}\delta(s-\omega/c_T)(\hat{s}_{2_\gamma}\hat{s}_{2_k} + \hat{s}_{3_\gamma}\hat{s}_{3_k})\right)\{\overline{\rho}^2\omega^4\delta_{jk}\delta_{\beta\gamma}\tilde{R}_{\rho\rho}(\mathbf{p}-\mathbf{s})$$
$$+ \overline{\lambda}^2 p_\beta p_j s_\gamma s_k \tilde{R}_{\lambda\lambda}(\mathbf{p}-\mathbf{s}) + \overline{\mu}^2(s_\beta s_j p_\gamma p_k + s_\beta p_\gamma p_l s_l \delta_{jk} + \delta_{\beta\gamma}p_l s_l s_j p_k + \delta_{\beta\gamma}\delta_{jk}p_l s_l p_m s_m)\tilde{R}_{\mu\mu}(\mathbf{p}-\mathbf{s})$$
$$+ \overline{\rho\lambda}\omega^2(\delta_{\beta\gamma}p_j s_k + \delta_{jk}p_\beta s_\gamma)\tilde{R}_{\rho\lambda}(\mathbf{p}-\mathbf{s}) + \overline{\rho\mu}\omega^2[\delta_{\beta\gamma}(s_j p_k + p_l s_l \delta_{jk}) + \delta_{jk}(s_\beta p_\gamma + p_l s_l \delta_{\beta\gamma})]\tilde{R}_{\rho\mu}(\mathbf{p}-\mathbf{s})$$
$$+ \overline{\lambda\mu}(p_\beta s_j p_k s_\gamma + p_\beta s_\gamma p_l s_l \delta_{jk} + s_\beta p_j s_k p_\gamma + p_j s_k p_l s_l \delta_{\beta\gamma})\tilde{R}_{\lambda\mu}(\mathbf{p}-\mathbf{s})\}. \quad (33)$$

For the transverse attenuation, the required result is

$$\mathrm{Im}\, m_T(\mathbf{p}) = -\frac{\pi}{\overline{\rho}}\frac{1}{4\omega}(\hat{p}_{2_\beta}\hat{p}_{2_j} + \hat{p}_{3_\beta}\hat{p}_{3_j})\int d^3 s\left(\frac{1}{c_L}\delta(s-\omega/c_L)\hat{s}_\gamma\hat{s}_k + \frac{1}{c_T}\delta(s-\omega/c_T)(\hat{s}_{2_\gamma}\hat{s}_{2_k} + \hat{s}_{3_\gamma}\hat{s}_{3_k})\right)$$
$$\times\{\overline{\rho}^2\omega^4\delta_{jk}\delta_{\beta\gamma}\tilde{R}_{\rho\rho}(\mathbf{p}-\mathbf{s}) + \overline{\lambda}^2\tilde{R}_{\lambda\lambda}(\mathbf{p}-\mathbf{s})p_\beta p_j s_\gamma s_k + \overline{\mu}^2(s_\beta s_j p_\gamma p_k + s_\beta p_\gamma p_l s_l \delta_{jk} + \delta_{\beta\gamma}p_l s_l s_j p_k$$
$$+ \delta_{\beta\gamma}\delta_{jk}p_l s_l p_m s_m)\tilde{R}_{\mu\mu}(\mathbf{p}-\mathbf{s}) + \overline{\rho\lambda}(\delta_{\beta\gamma}p_j s_k + \delta_{jk}p_\beta s_\gamma)\omega^2\tilde{R}_{\rho\lambda}(\mathbf{p}-\mathbf{s}) + [\delta_{\beta\gamma}(s_j p_k + p_l s_l \delta_{jk})$$
$$+ \delta_{jk}(s_\beta p_\gamma + p_l s_l \delta_{\beta\gamma})]\overline{\rho\mu}\omega^2\tilde{R}_{\rho\mu}(\mathbf{p}-\mathbf{s}) + \overline{\lambda\mu}(p_\beta s_j p_k s_\gamma + p_\beta s_\gamma p_l s_l \delta_{jk} + s_\beta p_j s_k p_\gamma + p_j s_k p_l s_l \delta_{\beta\gamma})\tilde{R}_{\lambda\mu}(\mathbf{p}-\mathbf{s})\}. \quad (34)$$

The two longitudinal attenuations, α_{LL} and α_{LT}, are determined first using Eqs. (33) and (25). The total longitudinal attenuation, $\alpha_L = \alpha_{LL} + \alpha_{LT}$. The first term from $\mathbf{G}^0$ involving $\hat{\mathbf{s}}\hat{\mathbf{s}}$ gives α_{LL}, while the second term containing $(\hat{\mathbf{s}}_2\hat{\mathbf{s}}_2 + \hat{\mathbf{s}}_3\hat{\mathbf{s}}_3)$ gives α_{LT}. The frequency-limiting approximation implies that $\mathbf{p} \approx \omega\hat{\mathbf{p}}/c_L$ in Eq. (33). The integral over the magnitude of the wave number is trivial, leaving only an integration over the unit sphere. The result for the LL attenuation is

$$\alpha_{LL} = \frac{\pi\omega^4}{4c_L^4}\int d^2\hat{s}\left\{(\hat{\mathbf{p}}\cdot\hat{\mathbf{s}})^2\tilde{R}^{LL}_{\rho\rho} + \frac{(c_L^2-2c_T^2)^2}{c_L^4}\tilde{R}^{LL}_{\lambda\lambda}\right.$$
$$+ \frac{4c_T^4}{c_L^4}(\hat{\mathbf{p}}\cdot\hat{\mathbf{s}})^4\tilde{R}^{LL}_{\mu\mu} + (\hat{\mathbf{p}}\cdot\hat{\mathbf{s}})\frac{2(c_L^2-2c_T^2)}{c_L^2}\tilde{R}^{LL}_{\rho\lambda}$$
$$\left. + (\hat{\mathbf{p}}\cdot\hat{\mathbf{s}})^3\frac{4c_T^2}{c_L^2}\tilde{R}^{LL}_{\rho\mu} + (\hat{\mathbf{p}}\cdot\hat{\mathbf{s}})^2\frac{4c_T^2(c_L^2-2c_T^2)}{c_L^4}\tilde{R}^{LL}_{\lambda\mu}\right\}. \quad (35)$$

The notation used in Eq. (35) for the $\tilde{R}$ terms is defined by

$$\tilde{R}^{ij}_{\lambda\lambda} = \tilde{R}_{\lambda\lambda}\left(\left|\frac{\omega}{c_i}\hat{\mathbf{p}} - \frac{\omega}{c_j}\hat{\mathbf{s}}\right|\right), \quad (36)$$

where the superscripts i and j refer to the possible wave types, L or T. The correlation functions depend only upon the cosine of the angle between the incident and scattered directions, $\chi = \hat{\mathbf{p}}\cdot\hat{\mathbf{s}}$, because of the assumption of statistical isotropy. Thus, the integration in azimuthal angle is trivial. The final result for the LL attenuation is

$$\alpha_{LL} = \frac{\pi^2\omega^4}{2c_L^4}\int_{-1}^{+1} d\chi\left\{\chi^2\tilde{R}^{LL}_{\rho\rho}(\chi) + \frac{(c_L^2-2c_T^2)^2}{c_L^4}\tilde{R}^{LL}_{\lambda\lambda}(\chi)\right.$$
$$+ \frac{4c_T^4}{c_L^4}\chi^4\tilde{R}^{LL}_{\mu\mu}(\chi) + \chi\frac{2(c_L^2-2c_T^2)}{c_L^2}\tilde{R}^{LL}_{\rho\lambda}(\chi)$$
$$\left. + \chi^3\frac{4c_T^2}{c_L^2}\tilde{R}^{LL}_{\rho\mu}(\chi) + \chi^2\frac{4c_T^2(c_L^2-2c_T^2)}{c_L^4}\tilde{R}^{LL}_{\lambda\mu}(\chi)\right\}. \quad (37)$$

This result is consistent with that given by Ryzhik *et al.*[10]

The mode conversion attenuation, α_{LT}, is found in a similar manner to be

$$\alpha_{LT} = \frac{\pi\omega^4}{4c_L c_T^3}\int d^2\hat{s}\left\{((\hat{\mathbf{p}}\cdot\hat{\mathbf{s}}_2)^2 + (\hat{\mathbf{p}}\cdot\hat{\mathbf{s}}_3)^2)\tilde{R}^{LT}_{\rho\rho}\right.$$
$$+ 4(\hat{\mathbf{p}}\cdot\hat{\mathbf{s}})^2\frac{c_T^2}{c_L^2}((\hat{\mathbf{p}}\cdot\hat{\mathbf{s}}_2)^2 + (\hat{\mathbf{p}}\cdot\hat{\mathbf{s}}_3)^2)\tilde{R}^{LT}_{\mu\mu}$$
$$\left. + 4(\hat{\mathbf{p}}\cdot\hat{\mathbf{s}})((\hat{\mathbf{p}}\cdot\hat{\mathbf{s}}_2)^2 + (\hat{\mathbf{p}}\cdot\hat{\mathbf{s}}_3)^2)\frac{c_T}{c_L}\tilde{R}^{LT}_{\rho\mu}\right\}. \quad (38)$$

The medium considered here is statistically isotropic. In this type of problem, the differential scattering cross-section (the integrand) must be a function of only $\hat{\mathbf{p}}\cdot\hat{\mathbf{s}}$. It can be easily shown that

$$(\hat{\mathbf{p}}\cdot\hat{\mathbf{s}}_2)^2+(\hat{\mathbf{p}}\cdot\hat{\mathbf{s}}_3)^2=1-(\hat{\mathbf{p}}\cdot\hat{\mathbf{s}})^2. \tag{39}$$

Thus, the mode conversion attenuation may be written

$$\alpha_{LT}=\frac{\pi\omega^4}{4c_T^3c_L}\int d^2\hat{s}(1-(\hat{\mathbf{p}}\cdot\hat{\mathbf{s}})^2) \times\left\{\tilde{R}_{\rho\rho}^{LT}+\frac{4c_T^2}{c_L^2}(\hat{\mathbf{p}}\cdot\hat{\mathbf{s}})^2\tilde{R}_{\mu\mu}^{LT}+\frac{4c_T}{c_L}(\hat{\mathbf{p}}\cdot\hat{\mathbf{s}})\tilde{R}_{\rho\mu}^{LT}\right\}. \tag{40}$$

After the azimuthal integration we have

$$\alpha_{LT}=\frac{\pi^2\omega^4}{2c_T^3c_L}\int_{-1}^{+1}(1-\chi^2) \times\left\{\tilde{R}_{\rho\rho}^{LT}(\chi)+4\frac{c_T^2}{c_L^2}\chi^2\tilde{R}_{\mu\mu}^{LT}(\chi)+4\frac{c_T}{c_L}\chi\tilde{R}_{\rho\mu}^{LT}(\chi)\right\}d\chi, \tag{41}$$

where $\chi=\hat{\mathbf{p}}\cdot\hat{\mathbf{s}}$. The term in brackets corresponds to σ_{ps} given by Ryzhik *et al.*, Eq. (4.56). The complete expression agrees with that given by Papanicolaou *et al.*,[11] Eq. (A2).

The calculation of the transverse wave attenuations proceeds along a similar line. The two transverse attenuations, α_{TL} and α_{TT}, are determined from Eqs. (34) and (25). The total transverse attenuation, $\alpha_T=\alpha_{TL}+\alpha_{TT}$. The first term of $\mathbf{G}^0$ involving $\hat{\mathbf{s}}\hat{\mathbf{s}}$ gives α_{TL}, while the second term containing $(\hat{\mathbf{s}}_2\hat{\mathbf{s}}_2+\hat{\mathbf{s}}_3\hat{\mathbf{s}}_3)$ gives α_{TT}. The frequency-limiting approximation implies that $\mathbf{p}\approx\omega\hat{\mathbf{p}}/c_T$ in Eq. (34). The integral over the magnitude of the wave number is trivial, leaving only an integration over the unit sphere. The result for the TL attenuation is

$$\alpha_{TL}=\frac{\pi\omega^4}{8c_L^3c_T}\int d^2\hat{s}\Big\{((\hat{\mathbf{s}}\cdot\hat{\mathbf{p}}_2)^2+(\hat{\mathbf{s}}\cdot\hat{\mathbf{p}}_3)^2)\tilde{R}_{\rho\rho}^{TL} +4(\hat{\mathbf{p}}\cdot\hat{\mathbf{s}})^2((\hat{\mathbf{s}}\cdot\hat{\mathbf{p}}_2)^2+(\hat{\mathbf{s}}\cdot\hat{\mathbf{p}}_3)^2)\frac{c_T^2}{c_L^2}\tilde{R}_{\mu\mu}^{TL} +4(\hat{\mathbf{p}}\cdot\hat{\mathbf{s}})((\hat{\mathbf{s}}\cdot\hat{\mathbf{p}}_2)^2+(\hat{\mathbf{s}}\cdot\hat{\mathbf{p}}_3)^2)\frac{c_T}{c_L}\tilde{R}_{\rho\mu}^{TL}\Big\}. \tag{42}$$

Again, the differential scattering cross sections (the integrands) must be only functions of $\hat{\mathbf{p}}\cdot\hat{\mathbf{s}}$. It can be shown that

$$(\hat{\mathbf{s}}\cdot\hat{\mathbf{p}}_2)^2+(\hat{\mathbf{s}}\cdot\hat{\mathbf{p}}_3)^2=1-(\hat{\mathbf{p}}\cdot\hat{\mathbf{s}})^2. \tag{43}$$

Thus, we have

$$\alpha_{TL}=\frac{\pi\omega^4}{8c_L^3c_T}\int d^2\hat{s}(1-(\hat{\mathbf{p}}\cdot\hat{\mathbf{s}})^2) \times\left\{\tilde{R}_{\rho\rho}^{TL}+\frac{4c_T^2}{c_L^2}(\hat{\mathbf{p}}\cdot\hat{\mathbf{s}})^2\tilde{R}_{\mu\mu}^{TL}+\frac{4c_T}{c_L}(\hat{\mathbf{p}}\cdot\hat{\mathbf{s}})\tilde{R}_{\rho\mu}^{TL}\right\}, \tag{44}$$

or, after azimuthal integration,

$$\alpha_{TL}=\frac{\pi^2\omega^4}{4c_L^3c_T}\int_{-1}^{+1}(1-\chi^2) \times\left\{\tilde{R}_{\rho\rho}^{TL}(\chi)+\frac{4c_T^2}{c_L^2}\chi^2\tilde{R}_{\mu\mu}^{TL}(\chi)+\frac{4c_T}{c_L}\chi\tilde{R}_{\rho\mu}^{TL}(\chi)\right\}d\chi. \tag{45}$$

As a check of self-consistency, the above expression for α_{TL} satisfies the required relation

$$\alpha_{TL}=\frac{1}{2}\left(\frac{c_T}{c_L}\right)^2\alpha_{LT}. \tag{46}$$

Finally, the TT attenuation may be reduced to

$$\alpha_{TT}=\frac{\pi\omega^4}{8c_T^4}\int d^2\hat{s}\{((\hat{\mathbf{p}}_2\cdot\hat{\mathbf{s}}_2)^2+(\hat{\mathbf{p}}_3\cdot\hat{\mathbf{s}}_2)^2+(\hat{\mathbf{p}}_2\cdot\hat{\mathbf{s}}_3)^2+(\hat{\mathbf{p}}_3\cdot\hat{\mathbf{s}}_3)^2)\tilde{R}_{\rho\rho}^{TT}+\tilde{R}_{\mu\mu}^{TT}[(\hat{\mathbf{s}}\cdot\hat{\mathbf{p}}_2)^2(\hat{\mathbf{p}}\cdot\hat{\mathbf{s}}_2)^2+(\hat{\mathbf{s}}\cdot\hat{\mathbf{p}}_3)^2(\hat{\mathbf{p}}\cdot\hat{\mathbf{s}}_2)^2 +(\hat{\mathbf{s}}\cdot\hat{\mathbf{p}}_2)^2(\hat{\mathbf{p}}\cdot\hat{\mathbf{s}}_3)^2+(\hat{\mathbf{s}}\cdot\hat{\mathbf{p}}_3)^2(\hat{\mathbf{p}}\cdot\hat{\mathbf{s}}_3)^2+2(\hat{\mathbf{p}}\cdot\hat{\mathbf{s}})((\hat{\mathbf{p}}_2\cdot\hat{\mathbf{s}}_2)(\hat{\mathbf{p}}\cdot\hat{\mathbf{s}}_2)(\hat{\mathbf{s}}\cdot\hat{\mathbf{p}}_2)+(\hat{\mathbf{p}}_3\cdot\hat{\mathbf{s}}_2)(\hat{\mathbf{s}}\cdot\hat{\mathbf{p}}_3)(\hat{\mathbf{p}}\cdot\hat{\mathbf{s}}_2))+2(\hat{\mathbf{p}}\cdot\hat{\mathbf{s}})((\hat{\mathbf{p}}_2\cdot\hat{\mathbf{s}}_3)(\hat{\mathbf{s}}\cdot\hat{\mathbf{p}}_2) \times(\hat{\mathbf{p}}\cdot\hat{\mathbf{s}}_3)+(\hat{\mathbf{p}}_3\cdot\hat{\mathbf{s}}_3)(\hat{\mathbf{s}}\cdot\hat{\mathbf{p}}_3)(\hat{\mathbf{p}}\cdot\hat{\mathbf{s}}_3))+(\hat{\mathbf{p}}\cdot\hat{\mathbf{s}})^2((\hat{\mathbf{p}}_2\cdot\hat{\mathbf{s}}_2)^2+(\hat{\mathbf{p}}_3\cdot\hat{\mathbf{s}}_2)^2+(\hat{\mathbf{p}}_2\cdot\hat{\mathbf{s}}_3)^2+(\hat{\mathbf{p}}_3\cdot\hat{\mathbf{s}}_3)^2)]+2\tilde{R}_{\rho\mu}^{TT}[(\hat{\mathbf{p}}_2\cdot\hat{\mathbf{s}}_2)(\hat{\mathbf{s}}\cdot\hat{\mathbf{p}}_2) \times(\hat{\mathbf{p}}\cdot\hat{\mathbf{s}}_2)+(\hat{\mathbf{p}}_3\cdot\hat{\mathbf{s}}_2)(\hat{\mathbf{s}}\cdot\hat{\mathbf{p}}_3)(\hat{\mathbf{p}}\cdot\hat{\mathbf{s}}_2)+(\hat{\mathbf{p}}_2\cdot\hat{\mathbf{s}}_3)(\hat{\mathbf{s}}\cdot\hat{\mathbf{p}}_2)(\hat{\mathbf{p}}\cdot\hat{\mathbf{s}}_3)+(\hat{\mathbf{p}}_3\cdot\hat{\mathbf{s}}_3)(\hat{\mathbf{s}}\cdot\hat{\mathbf{p}}_3)(\hat{\mathbf{p}}\cdot\hat{\mathbf{s}}_3)+(\hat{\mathbf{p}}\cdot\hat{\mathbf{s}})((\hat{\mathbf{p}}_2\cdot\hat{\mathbf{s}}_2)^2+(\hat{\mathbf{p}}_3\cdot\hat{\mathbf{s}}_2)^2 +(\hat{\mathbf{p}}_2\cdot\hat{\mathbf{s}}_3)^2+(\hat{\mathbf{p}}_3\cdot\hat{\mathbf{s}}_3)^2)]\}. \tag{47}$$

Once again, the differential scattering cross sections must depend only on $\hat{\mathbf{p}}\cdot\hat{\mathbf{s}}$. The above combinations of inner products can be reduced considerably. The necessary identities are

$$(\hat{\mathbf{p}}_2\cdot\hat{\mathbf{s}}_2)^2+(\hat{\mathbf{p}}_3\cdot\hat{\mathbf{s}}_2)^2+(\hat{\mathbf{p}}_2\cdot\hat{\mathbf{s}}_3)^2+(\hat{\mathbf{p}}_3\cdot\hat{\mathbf{s}}_3)^2=1+(\hat{\mathbf{p}}\cdot\hat{\mathbf{s}})^2, \tag{48}$$

$$(\hat{\mathbf{p}}_2\cdot\hat{\mathbf{s}}_2)(\hat{\mathbf{p}}\cdot\hat{\mathbf{s}}_2)(\hat{\mathbf{s}}\cdot\hat{\mathbf{p}}_2)+(\hat{\mathbf{p}}_3\cdot\hat{\mathbf{s}}_2)(\hat{\mathbf{s}}\cdot\hat{\mathbf{p}}_3)(\hat{\mathbf{p}}\cdot\hat{\mathbf{s}}_2)+(\hat{\mathbf{p}}_2\cdot\hat{\mathbf{s}}_3)(\hat{\mathbf{s}}\cdot\hat{\mathbf{p}}_2)(\hat{\mathbf{p}}\cdot\hat{\mathbf{s}}_3)+(\hat{\mathbf{p}}_3\cdot\hat{\mathbf{s}}_3)(\hat{\mathbf{s}}\cdot\hat{\mathbf{p}}_3)(\hat{\mathbf{p}}\cdot\hat{\mathbf{s}}_3)=(\hat{\mathbf{p}}\cdot\hat{\mathbf{s}})^3-(\hat{\mathbf{p}}\cdot\hat{\mathbf{s}}), \tag{49}$$

and

$$(\hat{\mathbf{s}}\cdot\hat{\mathbf{p}}_2)^2(\hat{\mathbf{p}}\cdot\hat{\mathbf{s}}_2)^2+(\hat{\mathbf{s}}\cdot\hat{\mathbf{p}}_3)^2(\hat{\mathbf{p}}\cdot\hat{\mathbf{s}}_2)^2+(\hat{\mathbf{s}}\cdot\hat{\mathbf{p}}_2)^2(\hat{\mathbf{p}}\cdot\hat{\mathbf{s}}_3)^2+(\hat{\mathbf{s}}\cdot\hat{\mathbf{p}}_3)^2(\hat{\mathbf{p}}\cdot\hat{\mathbf{s}}_3)^2=1-2(\hat{\mathbf{p}}\cdot\hat{\mathbf{s}})^2+(\hat{\mathbf{p}}\cdot\hat{\mathbf{s}})^4. \tag{50}$$

These identities allow α_{TT} to be written in terms of $\hat{\mathbf{p}}\cdot\hat{\mathbf{s}}$ only as

$$\alpha_{TT}=\frac{\pi\omega^4}{8c_T^4}\int d^2\hat{s}\{((\hat{\mathbf{p}}\cdot\hat{\mathbf{s}})^2+1)\tilde{R}_{\rho\rho}^{TT}+(1-3(\hat{\mathbf{p}}\cdot\hat{\mathbf{s}})^2 +4(\hat{\mathbf{p}}\cdot\hat{\mathbf{s}})^4)\tilde{R}_{\mu\mu}^{TT}+4(\hat{\mathbf{p}}\cdot\hat{\mathbf{s}})^3\tilde{R}_{\rho\mu}^{TT}\}. \quad (51)$$

Integration in azimuth leaves

$$\alpha_{TT}=\frac{\pi^2\omega^4}{4c_T^4}\int_{-1}^{+1}\{(\chi^2+1)\tilde{R}_{\rho\rho}^{TT}(\chi)+(1-3\chi^2+4\chi^4) \times\tilde{R}_{\mu\mu}^{TT}(\chi)+4\chi^3\tilde{R}_{\rho\mu}^{TT}(\chi)\}d\chi. \quad (52)$$

The TT attenuation, α_{TT}, depends solely on the shear wave speed as expected.

The attenuations given by Eqs. (37), (41), (45), and (52) are the main results of this section. The results here have been reduced to integrations on the unit circle in terms of the correlation power spectra. The forward-weighted attenuations, α', which quantify the amount of forward scattering, are identical to the expressions for the attenuations given above with an additional factor of $\chi=\cos\Theta$ in the integrand.[14]

In the low-frequency limit, the correlation functions are expected to be constant. Also in this limit, the scattering is nonpreferential (an equal amount of energy is scattered forward as backward). In this case, the α's are zero. However, as seen in the equations for the attenuations, Eqs. (37), (41), (45), and (52), this is not necessarily the case. The angular dependence associated with the autocorrelations are all even functions. Those associated with the cross correlations are all odd. The implications of these results are still unclear. Must the cross correlations approach zero in the low-frequency limit? In the next section, assumptions about the form of the fluctuations are made and sample calculations presented.

III. MEAN FREE PATHS AND ELASTIC DIFFUSIVITY

The above expressions for the attenuations are now used for determination of mean free paths and elastic diffusivity. The above derivation resulted in expressions for the displacement attenuations. The energy density is proportional to the square of the displacements. Thus, the energy attenuations are twice the displacement attenuations. In addition, the scattering mean free paths, l_L and l_T, are the inverse of the energy attenuation. Therefore, we have

$$l_L=\frac{1}{2(\alpha_{LL}+\alpha_{LT})}, \quad (53)$$

$$l_T=\frac{1}{2(\alpha_{TL}+\alpha_{TT})}. \quad (54)$$

In the low-frequency Rayleigh limit, the correlation functions, $\tilde{R}$, are constant. Thus, the expressions for the attenuations, Eqs. (37), (41), (45), and (52) reduce considerably. In this limit, the integrands become simple polynomials in χ. After integration, we have

$$\alpha_{LL}=\frac{\pi^2\omega^4}{c_L^4}\left(\frac{1}{3}\tilde{R}_{\rho\rho}^{LL}+(1-2B^2)^2\tilde{R}_{\lambda\lambda}^{LL}+\frac{4}{5}B^4\tilde{R}_{\mu\mu}^{LL} +\frac{4}{3}B^2(1-2B^2)\tilde{R}_{\lambda\mu}^{LL}\right), \quad (55)$$

$$\alpha_{LT}=\frac{2\pi^2\omega^4}{3c_T^3c_L}\left(\tilde{R}_{\rho\rho}^{LT}+\frac{4}{5}B^2\tilde{R}_{\mu\mu}^{LT}\right), \quad (56)$$

$$\alpha_{TL}=\frac{\pi^2\omega^4}{3c_Tc_L^3}\left(\tilde{R}_{\rho\rho}^{LT}+\frac{4}{5}B^2\tilde{R}_{\mu\mu}^{LT}\right), \quad (57)$$

$$\alpha_{TT}=\frac{2\pi^2\omega^4}{c_T^4}\left(\frac{1}{3}\tilde{R}_{\rho\rho}^{TT}+\frac{1}{5}\tilde{R}_{\mu\mu}^{TT}\right), \quad (58)$$

where $B=c_T/c_L$ is the wave speed ratio. Also, in the low-frequency limit, the transforms of the correlation functions are independent of the wave types. In other words, $\tilde{R}_{\rho\rho}\equiv\tilde{R}_{\rho\rho}^{LL}=\tilde{R}_{\rho\rho}^{LT}=\tilde{R}_{\rho\rho}^{TT}$. Similar relations hold for the λ and μ terms as well. The scattering mean free paths then reduce to

$$l_L=\frac{15c_L^4B^3/(2\pi^2\omega^4)}{5(2+B^3)\tilde{R}_{\rho\rho}+15B^3(1-2B^2)^2\tilde{R}_{\lambda\lambda}+4B^2(3B^5+2)\tilde{R}_{\mu\mu}+20B^5(1-2B^2)\tilde{R}_{\lambda\mu}}, \quad (59)$$

$$l_T=\frac{15c_T^4/(2\pi^2\omega^4)}{5(2+B^3)\tilde{R}_{\rho\rho}+2(3+2B^5)\tilde{R}_{\mu\mu}}. \quad (60)$$

These results differ slightly from those given by Ryzhik *et al.*[10]

The elastic diffusivity, D, can be written[14]

$$D=\frac{c_Tl_T^*}{3}\frac{2+B^2L}{2+B^3}, \quad (61)$$

where $L=l_L^*/l_T^*$ is the ratio of transport mean free paths. The elastic diffusivity is a weighted average of the diffusivities of individual compressional and shear components. The weighting is determined by the diffusive equipartitioning law, which states that the energy in transverse form, E_T, and the longitudinal energy, E_L, are related by $E_T=2E_L/B^3$.[15] In the low-frequency limit, the transport mean free paths reduce to the scattering mean free paths

$$l_L^*=l_L=\frac{1}{2\alpha_L}, \quad (62)$$

$$l_T^*=l_T=\frac{1}{2\alpha_T}. \quad (63)$$

Outside the long wavelength limit, the transport mean free paths are given by[14]

$$l_L^* = \frac{1}{2}\frac{\alpha_T - \alpha'_{TT} + \alpha'_{LT}}{(\alpha_L - \alpha'_{LL})(\alpha_T - \alpha'_{TT}) - \alpha'_{LT}\alpha'_{TL}}, \tag{64}$$

$$l_T^* = \frac{1}{2}\frac{\alpha_L - \alpha'_{LL} + \alpha'_{TL}}{(\alpha_L - \alpha'_{LL})(\alpha_T - \alpha'_{TT}) - \alpha'_{LT}\alpha'_{TL}}, \tag{65}$$

where the primed attenuations are the forward-weighted displacement attenuations discussed above for the respective scattering processes. In the limit of nonpreferential scattering, the primed attenuations vanish and the mean free paths are seen to reduce to the inverse of the energy scattering attenuation as in Eqs. (62)–(63). Example calculations are now presented using the above derived quantities.

IV. EXAMPLE RESULTS

Example results are now presented using additional material assumptions. The results are put in dimensionless form for the most widespread applicability. It is first assumed that the material properties are uncorrelated with one another. This implies that the cross correlations are zero ($\tilde{R}_{\mu\lambda} = \tilde{R}_{\rho\mu} = \tilde{R}_{\rho\lambda} = 0$).

In this case, the attenuations become

$$\alpha_{LL} = \frac{\pi^2\omega^4}{2c_L^4}\int_{-1}^{+1}\{\chi^2\tilde{R}^{LL}_{\rho\rho}(\chi) + (1-2B^2)^2\tilde{R}^{LL}_{\lambda\lambda}(\chi) + 4B^4\chi^4\tilde{R}^{LL}_{\mu\mu}(\chi)\}d\chi, \tag{66}$$

$$\alpha_{LT} = \frac{\pi^2\omega^4}{2c_T^3c_L}\int_{-1}^{+1}(1-\chi^2)\{\tilde{R}^{LT}_{\rho\rho}(\chi) + 4B^2\chi^2\tilde{R}^{LT}_{\mu\mu}(\chi)\}d\chi, \tag{67}$$

and

$$\alpha_{TL} = \tfrac{1}{2}B^2\alpha_{LT}, \tag{68}$$

$$\alpha_{TT} = \frac{\pi^2\omega^4}{4c_T^4}\int_{-1}^{+1}\{(1+\chi^2)\tilde{R}^{TT}_{\rho\rho}(\chi) + (1-3\chi^2+4\chi^4)\times\tilde{R}^{TT}_{\mu\mu}(\chi)\}d\chi. \tag{69}$$

The forward-weighted attenuations[3,14] are defined as these with an extra factor of $\chi = \cos\Theta$ within the integrand. These attenuations are denoted with a prime.

Next, it is assumed that all autocorrelation functions have the same spatial dependence. They are assumed to have the form

$$R_{\gamma\gamma} = A_\gamma^2 R(r), \tag{70}$$

where the subscript γ refers to the material parameters of ρ, λ, and μ. The magnitude of the fluctuations for each material parameter is given by A_γ.

Finally, a form for the function $R(r)$ is assumed. As discussed by Stanke,[16] an exponential function describes the correlation of continuous and discrete materials reasonably well. Thus, it is assumed that

$$R(r) = e^{-r/H}, \tag{71}$$

where H is the correlation length. Such a model, with a single length scale, is perhaps oversimplified for materials with polydispersed scatterer sizes. However, for many materials such a model is expected to describe the statistics of the material properties well. In transform space

$$\tilde{R}(p) = \frac{H^3}{\pi^2(1+H^2p^2)^2}. \tag{72}$$

With the length scale of the spatial correlation introduced, dimensionless longitudinal and transverse frequencies are defined as $x_L = \omega H/c_L$ and $x_T = \omega H/c_T$. The transform of the difference between two wave vectors can then be written

$$\tilde{R}^{ij}(\chi) = \frac{H^3}{\pi^2(1+x_i^2+x_j^2-2x_ix_j\chi)^2}, \tag{73}$$

where the superscripts, ij, correspond to the possible wave types, L or T. The functions needed for determining the attenuations are

$$\tilde{R}^{LL}(\chi) = \frac{H^3}{\pi^2(1+2x_L^2(1-\chi))^2}, \tag{74}$$

$$\tilde{R}^{LT}(\chi) = \frac{H^3}{\pi^2(1+x_L^2+x_T^2-2x_Lx_T\chi)^2}, \tag{75}$$

$$\tilde{R}^{TT}(\chi) = \frac{H^3}{\pi^2(1+2x_T^2(1-\chi))^2}. \tag{76}$$

The attenuations may then be written in dimensionless form as

$$\alpha_{LL}H = \frac{x_L^4}{2}\int_{-1}^{+1}\frac{A_\rho^2\chi^2 + A_\lambda^2(1-2B^2)^2 + 4B^4A_\mu^2\chi^4}{(1+2x_L^2(1-\chi))^2}d\chi, \tag{77}$$

$$\alpha_{LT}H = \frac{x_L^4}{2B^3}\int_{-1}^{+1}\frac{(1-\chi^2)(A_\rho^2+4A_\mu^2B^2\chi^2)}{(1+x_L^2+x_T^2-2x_Lx_T\chi)^2}d\chi, \tag{78}$$

$$\alpha_{TT}H = \frac{x_T^4}{4}\int_{-1}^{+1}\frac{A_\rho^2(1+\chi^2) + A_\mu^2(1-3\chi^2+4\chi^4)}{(1+2x_T^2(1-\chi))^2}d\chi. \tag{79}$$

The integrals for the final form of the attenuations, both nonprimed and primed, can be calculated in closed-form.[3] All integrations are of the form

$$\int_{-1}^{+1}\frac{\chi^n}{(a-b\chi)^2}d\chi, \tag{80}$$

with n ranging from 0 to 5. The integrals were evaluated using numerical integration. Recursive adaptive Lobatto quadrature is available through the Matlab function ''quadl.''[17]

Figure 1 is a plot of the dimensionless longitudinal and transverse attenuations, $\alpha_L H$ and $\alpha_T H$, respectively, as a function of dimensionless frequency, x_L, for density fluctuations only ($A_\lambda = A_\mu = 0$). A wave speed ratio of $c_T/c_L = 1/\sqrt{3}$ has been used for these results and those that follow. The attenuations are seen to increase with the fourth power of frequency in the low-frequency limit. After a transition region, the attenuations increase as the frequency squared. The high-frequency geometric optics limit, in which the attenuations are constant, is not predicted due to the frequency-limiting assumption used above. The transverse

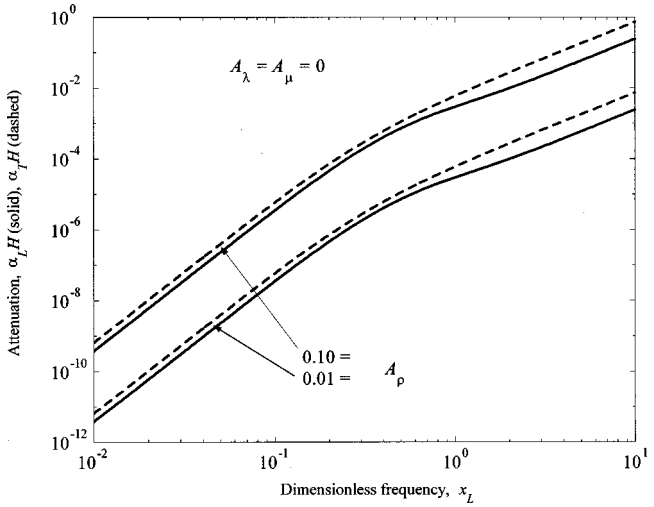


FIG. 1. Dimensionless longitudinal and transverse attenuations, $\alpha_L H$ and $\alpha_T H$, as a function of dimensionless frequency, x_L, for density fluctuations only.

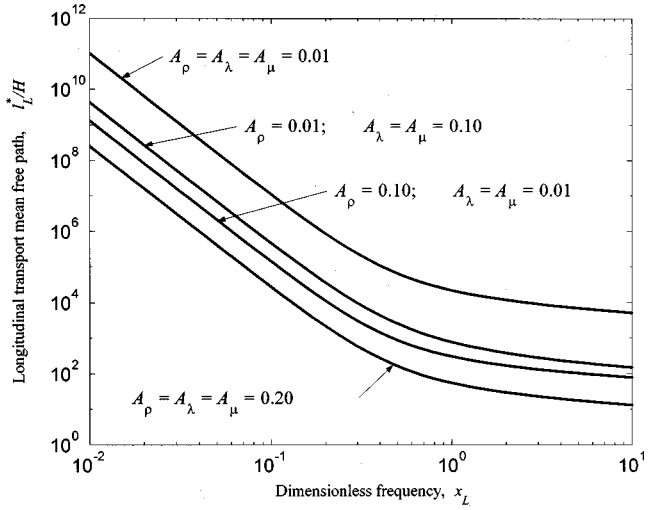


FIG. 3. Dimensionless longitudinal transport mean free path, l_L^*/H, as a function of dimensionless frequency, x_L, for different material fluctuation levels.

attenuation is larger than the longitudinal as expected. The difference between the two attenuations increases with frequency, but is a constant in each frequency regime. The actual ratio of the attenuations is a function of the fluctuations.

Figure 2 is a plot of the dimensionless longitudinal and transverse attenuations, $\alpha_L H$ and $\alpha_T H$, respectively, as a function of dimensionless frequency, x_L, for modulus fluctuations only ($A_\rho = 0$). The results have a similar form as those in Fig. 1. Combinations of both density and modulus fluctuations are simply the sum of the results from these two figures. The component of the attenuation attributed to density fluctuations (Fig. 1) is seen to be much larger than the component of attenuation attributable to modulus fluctuations for the same level of fluctuations. The importance of the density fluctuations has not been discussed previously.

The dimensionless transport mean free paths, l_L^*/H and l_T^*/H, are shown in Figs. 3 and 4. The complex forms of these terms, Eqs. (64)–(65), do not allow simple addition of factors from the different parameter fluctuations. Thus, a few different combinations of fluctuation levels have been chosen to highlight the range of these quantities. The ratio of the transport mean free paths is of interest since it appears in the definition of the diffusivity. It can be seen in Figs. 3 and 4 that the ratio, $L = l_L^*/l_T^*$, comes very near unity at the higher frequencies for many of the combinations of material fluctuations shown.

Finally, the dimensionless elastic diffusivity, D/Hc_T, is shown in Fig. 5 for various combinations of material fluctuations. Again, the low-frequency limit has the expected form, decreasing with the inverse fourth power of frequency. At higher frequencies, the diffusivity becomes nearly frequency independent as in the polycrystalline case.[3]

V. DISCUSSION

The propagation and scattering of elastic waves in heterogeneous media has been examined. Appropriate ensemble averaging of the elastic wave equation resulted in the Dyson

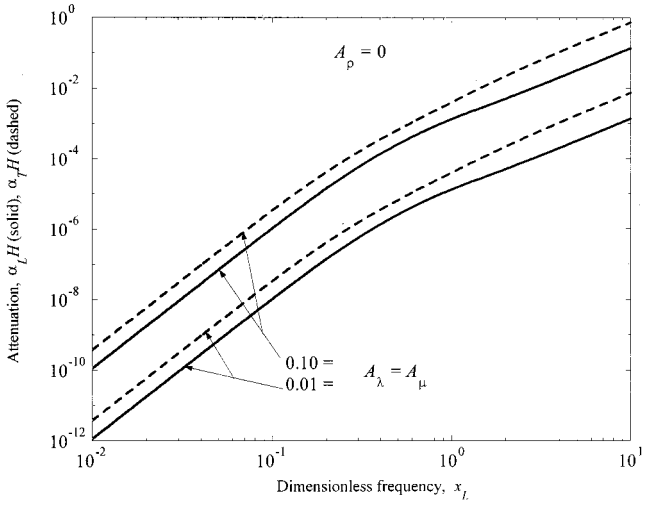


FIG. 2. Dimensionless longitudinal and transverse attenuations, $\alpha_L H$ and $\alpha_T H$, as a function of dimensionless frequency, x_L, for modulus fluctuations only.

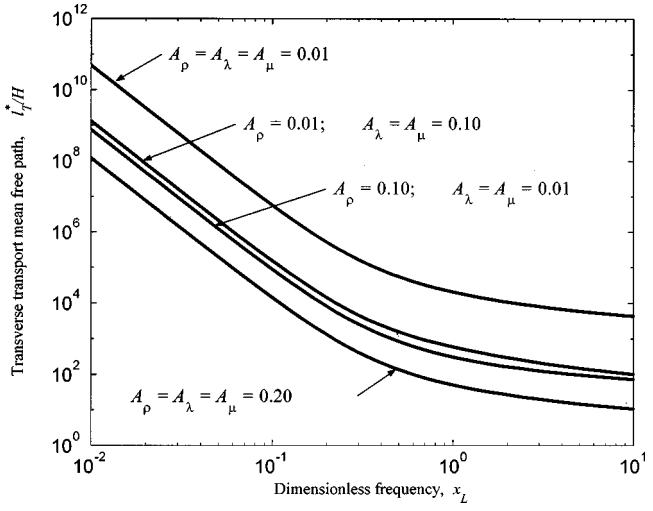


FIG. 4. Dimensionless transverse transport mean free path, l_T^*/H, as a function of dimensionless frequency, x_L, for different material fluctuation levels.

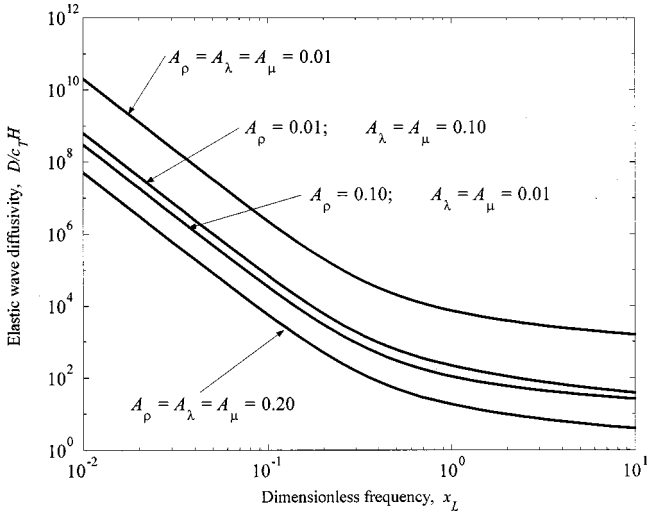


FIG. 5. Dimensionless elastic wave diffusivity, $D/c_T H$, as a function of dimensionless frequency, x_L, for different material fluctuation levels.

equation, governing the mean response. The problem was further specified for the case of both density and Lamé constants which varied spatially. The final forms of the attenuations are given by simple expressions involving integrations over the unit circle. The integrands are dependent upon inner products on the covariance of fluctuations of the material properties. The expressions derived here are in basic agreement with those results found using asymptotic methods.[10,11] Therefore, it is expected that the asymptotic approach would have the same restrictions as the current method. The fluctuations must be small and the frequency must not be so high that the phase velocity is appreciably altered.

The results presented here are also directly applicable to diffuse field methods such as backscatter techniques.[18] Recently, the above model has been further modified for two-phase materials. This model has been used for comparison with experiments of ultrasound diffusion in concrete.[19] The comparison between the theory and experiments is quite good despite the simplicity of many of the assumptions.

ACKNOWLEDGMENTS

The support of the Nebraska Research Initiative, the University of Nebraska Foundation Layman Fund, and the National Science Foundation (Grant No. CMS-9978707) is gratefully acknowledged. The authors also thank L. Ryzhik for helpful communication regarding this work.

[1] S. Hirsekorn, ''The scattering of ultrasonic waves by polycrystals,'' J. Acoust. Soc. Am. **72**, 1021–1031 (1982).
[2] F. E. Stanke and G. S. Kino, ''A unified theory for elastic wave propagation in polycrystalline materials,'' J. Acoust. Soc. Am. **75**, 665–681 (1984).
[3] R. L. Weaver, ''Diffusion of ultrasound in polycrystals,'' J. Mech. Phys. Solids **38**, 55–86 (1990).
[4] J. A. Turner and R. L. Weaver, ''Radiative transfer and multiple scattering of diffuse ultrasound in polycrystalline media,'' J. Acoust. Soc. Am. **96**, 3675–3683 (1994).
[5] S. Hirsekorn, ''Directional dependence of ultrasonic propagation in textured polycrystals,'' J. Acoust. Soc. Am. **79**, 1269–1279 (1986).
[6] S. Ahmed and R. B. Thompson, ''Propagation of elastic waves in equiaxed stainless-steel polycrystals with aligned [001] axes,'' J. Acoust. Soc. Am. **99**, 2086–2096 (1996).
[7] J. A. Turner, ''Elastic wave propagation and scattering in heterogeneous, anisotropic media: Textured polycrystalline materials,'' J. Acoust. Soc. Am. **106**, 541–552 (1999).
[8] J. A. Turner and R. L. Weaver, ''Radiative transfer of ultrasound,'' J. Acoust. Soc. Am. **96**, 3654–3674 (1994).
[9] J. A. Turner and R. L. Weaver, ''Time dependence of multiply scattered diffuse ultrasound in polycrystalline media,'' J. Acoust. Soc. Am. **97**, 2639–2644 (1995).
[10] L. V. Ryzhik, G. C. Papanicolaou, and J. B. Keller, ''Transport equations for elastic and other waves in random media,'' Wave Motion **24**, 327–370 (1996).
[11] G. C. Papanicolaou, L. V. Ryzhik, and J. B. Keller, ''Stability of the P-to-S energy ratio in the diffusive regime,'' Bull. Seismol. Soc. Am. **86**, 1107–1115 (1996).
[12] U. Frisch, ''Wave propagation in random media,'' in *Probalistic Methods in Applied Mathematics*, edited by A. T. Barucha-Reid (Academic, New York, 1968), Vol. I, pp. 75–198.
[13] F. C. Karal and J. B. Keller, ''Elastic, electromagnetic, and other waves in a random medium,'' J. Math. Phys. **5**, 537–547 (1964).
[14] J. A. Turner, ''Scattering and diffusion of seismic waves,'' Bull. Seismol. Soc. Am. **88**, 276–283 (1998).
[15] R. L. Weaver, ''On diffuse waves in solid media,'' J. Acoust. Soc. Am. **71**, 1608–1609 (1982).
[16] F. E. Stanke, ''Spatial autocorrelation functions for calculations of effective propagation constants in polycrystalline materials,'' J. Acoust. Soc. Am. **80**, 1479–1485 (1986).
[17] *Matlab, The Language of Technical Computing* (The Mathworks, Inc., Natick, MA).
[18] R. B. Thompson, ''A generalized model of the effects of microstructure on ultrasonic backscattering and flaw detection,'' in *Review of Progress in Quantitative NDE*, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 1996), Vol. 15, pp. 1471–1477.
[19] P. Anugonda, J. S. Wiehn, and J. A. Turner, ''Diffusion of ultrasound in concrete,'' Ultrasonics (under review).

Tomography of joint *P*-wave traveltime and polarization data: A simple approach for media with low to moderate velocity gradients

Lin-Ping Song[a)]
Department of Physics and Bernard Price Institute for Geophysical Research, University of the Witwatersrand, P.O. WITS 2050, South Africa

A. G. Every[b)]
Department of Physics, University of the Witwatersrand, P.O. WITS 2050, South Africa

C. Wright[c)]
Bernard Price Institute for Geophysical Research, University of the Witwatersrand, P.O. WITS 2050, South Africa

(Received 28 May 2000; revised 5 February 2001; accepted 5 February 2001)

An elastic wave tomography method utilizing joint traveltime and polarization data is proposed that is computationally simpler than the existing methods [Hu and Menke, Geophys. J. Int. **110**, 63 (1992); Farra and Begat, Geophys. J. Int. **121**, 371 (1995)]. In the linearization problem for the use of polarization data, we start with ray perturbation theory and assume that the medium is weakly inhomogeneous. Then the problem formulation for polarization data is approximately expressed as a linear integral of the gradient of the medium slowness perturbation along a reference ray. We call this a quasi-linear approximation which ignores the effect of the perturbation of the ray position on the first-order perturbation of the ray slowness vector. To efficiently obtain the solution for multi-data sets, a quadratic objective functional is constructed by including the data misfit terms and a model constraint term. Then a new conjugate gradient type of iterative reconstruction algorithm is developed to solve this minimization problem. This algorithm is also an extension of the conjugate gradient approach for standard least-squares problems. The feasibility and capability of the proposed tomography method is illustrated by conducting both noise-free and noisy synthetic experiments in a cross-hole geometry. The numerical results demonstrate that the additional use of polarization data not only improves the image quality, but also has a stabilizing effect on the iterative tomography solution. However, the limitation of the method is that it becomes inaccurate if the velocity variations in the medium change rapidly with position.
[DOI: 10.1121/1.1360238]

PACS numbers: 43.20.Dk, 43.35.Wa, 43.40.Ph [ANN]

I. INTRODUCTION

Elastic waves have long been used to remotely investigate physical properties of media in diverse scientific and industrial fields. Among these remote sensing approaches with acoustic or elastic waves, acoustical and seismic tomography have in recent decades received considerable attention for their demonstrated and potential ability to provide quantitative, high-resolution images of objects such as soft tissues, rock masses, marine sediments, and ocean flow.[1–9] As a broad generalization, both acoustic and seismic tomography may be loosely divided into ray-based and wave-based methods.[1,10] The former is concerned with the geometrical effect of wave propagation and is valid when the dominant wavelength is small compared to the scale of the inhomogeneities being probed. The latter (also known as diffraction tomography) attempts to account for the weak scattering/diffracting effect on wave propagation and is better suited to situations where the dominant acoustic wavelength is comparable to the size of the scattering body. Of the two methods, ray tomography is much simpler and easier to implement computationally. It sidesteps the complex issue of diffraction and its effects[3,10,11] and is able to directly yield useful velocity information.[5–9]

In this study, we confine attention to ray tomography, in which the validity of the geometrical ray approximation is assumed. In many cases, the observed data used in ray tomography are the first-arrival times of compressional waves, which are more clearly discernible than later arrivals. Besides this fundamental set of data, other distinct information contained in the first-arrival signal, such as the polarization properties of the acoustical waves, may be exploited to improve image quality. In this regard, Hu and Menke[12] first formulated the isotropic tomography problem of polarization data using a variational technique, and they applied the method to mapping the *P*-wave velocity structure of the crust and upper mantle.[13] Later Farra and Begat[14,15] extended the use of polarization data to transverse isotropic media by ray perturbation theory and reported synthetic experiments in vertical seismic profiling (VSP) geometry. Their studies showed that polarization data can offer useful and indepen-

[a)]Electronic mail: lpsong@hrz.uni-kassel.de
[b)]Electronic mail: every@physnet.phys.wits.ac.za
[c)]Electronic mail: 006CWRIG@cosmos.wits.ac.za

dent information in tomographic inversion. However, use of polarization data is only possible if elastic wave energy can be recorded using sensors such as geophones or seismometers that record two or three components of motion. At present, such sensors are largely restricted to seismic imaging, although new technologies may in the future permit more widespread use of polarization data.

Both traveltime and polarization data are nonlinearly dependent upon the physical property to be determined for the medium, e.g., velocity or its inverse, slowness. A tractable solution of the ray tomography problem requires a linear relation between data residuals and the perturbation parameters in a reference medium. Fortunately, using traveltime data, the nonlinear traveltime integral can be conveniently linearized along a reference ray by applying Fermat's principle. However, the problem of polarization data is not so simple because the polarization vectors do not obey a stationary principle along a ray.

Variational and perturbational formulas[12,14] in linearized procedures have been used to calculate the partial derivatives of the polarization vector with respect to the model parameters. However, the complicated calculations in both types of formulas are not easy to carry out. To some extent, these complications restrict their more extensive application.

The computational complexity of the existing methods and the valuable information in the polarization data demonstrates the need for a simpler but robust tomographic inversion scheme. The objective of this paper is to outline a simpler and feasible way of incorporating polarization data into the tomographic reconstruction scheme. In Sec. II, a perturbation technique is used to develop perturbed ray equations, which serve as the basis for formulating the relevant tomography problem. Then an approximately linearized formula is derived for using polarization data, and the problem of combining the traveltime and polarization data is expressed in a matrix form in Sec. III. Section IV presents a conjugate gradient-based tomographic iterative reconstruction scheme applicable to multi-data sets (traveltime and polarization data). In Sec. V, the feasibility and efficiency of the proposed approach is evaluated numerically in a synthetic cross-hole imaging experiment.

II. THE FIRST-ORDER PERTURBATION EQUATIONS OF TRAVELTIME AND SLOWNESS VECTOR

In recent years, ray perturbation theory has been well developed in Hamiltonian[16–18] and Lagrangian forms[19–21] to describe the effect of small perturbations in either the medium slowness or ray endpoints or reference curve (not being the true ray) on the ray position and traveltime. It can be used to model waveforms, trace a two-point ray, study higher-order traveltime perturbations associated with the change of ray positions, and determine Frechet derivatives of polarization data with respect to model parameters. The focus of this paper is on the last of these applications, but we will treat it here in a different way from Farra and Begat.[14] Following the ray perturbation analysis method, this section first gives the two fundamental perturbation equations on traveltime and slowness vector that are used in the subsequent formulation of the tomographic problem.

On the assumption of infinite frequency, the propagation of a kinematic ray is governed by the following two first-order ordinary differential equations:[22]

$$u(\mathbf{r})\frac{d\mathbf{x}}{ds}=\mathbf{p}, \tag{1a}$$

$$\frac{d\mathbf{p}}{ds}=\nabla u(\mathbf{r}), \tag{1b}$$

where $u(\mathbf{r})$ is the slowness of the medium, $\mathbf{r}$ the position vector of a ray, $\mathbf{p}$ the slowness vector of a ray, and s the arc length along a ray. The derivative of traveltime t along a ray is expressed as

$$\frac{dt}{ds}=u(\mathbf{r}). \tag{2}$$

Equations (1) and (2) show that the propagation of a ray depends on the slowness distribution of the medium which is the unknown in tomographic inverse problems.

Now suppose that the slowness $u(\mathbf{r})$ is expanded into a sum of reference slowness $u_0(\mathbf{r})$ and a small slowness perturbation $u_1(\mathbf{r})$,

$$u(\mathbf{r})=u_0(\mathbf{r})+\epsilon u_1(\mathbf{r}), \tag{3}$$

where the parameter ϵ is used to keep track of the order of approximation, a term containing the factor ϵ^n being the nth order of smallness. In this paper, the perturbation $u_1(\mathbf{r})$ is the object to be inverted in tomography based on the reference medium $u_0(\mathbf{r})$. A slight perturbation in the reference medium causes a perturbation of the reference ray. Using the ray arc length s_0 along the unperturbed ray as an independent parameter, the quantities describing a ray in a perturbed medium have the regular power series forms

$$\mathbf{r}(s_0)=\mathbf{r}_0(s_0)+\epsilon\mathbf{r}_1(s_0)+\cdots, \tag{4a}$$

$$\mathbf{p}(s_0)=\mathbf{p}_0(s_0)+\epsilon\mathbf{p}_1(s_0)+\cdots, \tag{4b}$$

$$t(s_0)=t_0(s_0)+\epsilon t_1(s_0)+\cdots. \tag{4c}$$

In the use of ray perturbation theory as indicated by Sneider and Sambridge,[20] one may explicitly specify the mapping from the unperturbed to the perturbed ray by an appropriate choice of a stretch factor $(\dot{\mathbf{r}}_0\cdot\dot{\mathbf{r}}_1)$, where a dot over a vector denotes the derivative with respect to the arc length s_0 along the reference ray. With this factor, the first-order relation between the arc length of corresponding points on the unperturbed and the perturbed rays is given by[20]

$$\frac{\partial s}{\partial s_0}=1+\epsilon(\dot{\mathbf{r}}_0\cdot\dot{\mathbf{r}}_1). \tag{5}$$

The stretch factor is optional[20] and could make the related problem simpler.

Following Moore,[23] we choose the stretch factor as

$$(\dot{\mathbf{r}}_0\cdot\dot{\mathbf{r}}_1)=0. \tag{6}$$

This choice maps points along the reference ray onto points along the perturbed with equal arc length or means $\partial s/\partial s_0=1$ to first order in small quantities. Similarly, in Hamiltonian ray perturbation theory,[14,18] the mapping from

the reference ray to the perturbed ray is specified by use of the same independent sampling τ_0-parameter (defined by $u\,d\tau = ds$) along the reference ray. Equivalently, for this case, $\partial\tau/\partial\tau_0$ equals unity to first order in small quantities.

With Eq. (4a) one expands the perturbed slowness around $\mathbf{r}_0$ as a Taylor series to first order

$$u(\mathbf{r}) = u_0(\mathbf{r}_0) + \epsilon[\nabla u_0 \cdot \mathbf{r}_1 + u_1(\mathbf{r}_0)] + \cdots. \tag{7}$$

Then, the first-order expansion of the slowness gradient is

$$\nabla u(\mathbf{r}) = \nabla u_0(\mathbf{r}_0) + \epsilon[\nabla\nabla u_0 \cdot \mathbf{r}_1 + \nabla u_1(\mathbf{r}_0)] + \cdots. \tag{8}$$

Substituting the perturbed expressions (4)–(8) into Eqs. (1) and (2) and equating the coefficients of powers of ϵ on the left- and right-hand sides of the relevant equations, we obtain the perturbed differential equations to first order:

$$\frac{dt_1}{ds_0} = u_1(\mathbf{r}_0) + \nabla u_0 \cdot \mathbf{r}_1, \tag{9a}$$

$$\frac{d\mathbf{p}_1}{ds_0} = \nabla u_1(\mathbf{r}_0) + \nabla\nabla u_0 \cdot \mathbf{r}_1. \tag{9b}$$

We see that Eqs. (9a)–(9b) have similar structures and involve the first-order perturbation in the ray position. The differential expression of the first-order perturbation $\mathbf{p}_1$ of the slowness vector contains higher-order spatial derivatives of the reference slowness and slowness perturbation than that of the first-order traveltime perturbation t_1. Both equations will serve as the basis to further formulate the tomographic problem in the next section.

III. PROBLEM FORMULATION

To start with, we consider integration of Eqs. (9a)–(9b) along a reference ray. Using Eq. (1) and integration by parts, the integral of the second term in the right-hand side of (9a) can be expressed as

$$\int_0^{S_0} (\nabla u_0 \cdot \mathbf{r}_1) ds_0 = [u_0(\dot{\mathbf{r}}_0 \cdot \mathbf{r}_1)]_0^{\epsilon_0}, \tag{10}$$

where S_0 is the total arc length of the reference ray. The right-hand term of Eq. (10) illustrates the changes of the traveltime resulting from the change in the endpoints of the reference ray.

In tomography problems the source and receiver points are fixed, i.e., $\mathbf{r}_1(0) = \mathbf{r}_1(S_0) = 0$. This immediately makes the terms $[u_0(\dot{\mathbf{r}}_0 \cdot \mathbf{r}_1)]_0^{S_0}$ vanish. Then the integral of Eq. (9a) is given by

$$t_1(S_0) = t_1(0) + \int_0^{S_0} u_1\, ds_0. \tag{11a}$$

Equation (11a) has been derived from perturbation theory by Snider and Sambridge[19,20] and shows that the first-order traveltime perturbation t_1 is the integral of the slowness perturbation u_1 along the reference ray. This is the well-known result from Fermat's principle, which states that the traveltime is stationary to first order for a slightly perturbed ray.[22]

However, the line integral of the second term on the right-hand side of Eq. (9b) does not have a compact form like Eq. (10), and is complicated since there are terms expressing the ray displacements. Note that the perturbation $\mathbf{r}_1$ is modulated by $\nabla\nabla u_0$ in Eq. (9b). When a reference medium is homogeneous or has a linear distribution of slowness, the term explicitly containing $\mathbf{r}_1$ could be removed and the related complication is avoided because $\nabla\nabla u_0 = 0$. Therefore in this case, the integral of Eq. (9b) is given by

$$\mathbf{p}_1(S_0) = \mathbf{p}_1(0) + \int_0^{S_0} \nabla u_1\, ds_0. \tag{11b}$$

Equation (11b) then shows that the first-order slowness vector perturbation $\mathbf{p}_1$ is now the integral of the gradient ∇u_1 of the medium slowness perturbation along the reference ray.

For a fixed two-point ray tracing problem, it is in principle necessary to give the initial perturbation slowness vector $\mathbf{p}_1(0)$ in order to ensure that the perturbed ray arrives at the desired receiver. In ray bending, Farra[18] numerically solved a linear system containing the endpoint boundary conditions for $\mathbf{p}_1(0)$ to get the required solution iteratively within a certain accuracy. The updating process may diverge due to the ill-conditioning of the linear system, caused by certain difficulties such as caustic surfaces or shadow zones. Farra and Begat[14] applied the algorithm for determining $\mathbf{p}_1(0)$ to their problem formulation of polarization data. Moore[24] presented a theoretical expression for $\mathbf{p}_1(0)$ which may be used to find the perturbed ray that reaches the required endpoint at the same value of the sampling parameter as on the unperturbed ray. Both involve the quite complex calculation of the ray propagator, the inverse of some related sub-matrix, and the first-order ray corrections.

To simplify the problem, we may choose $\mathbf{p}_1(0)$ differently by relaxing the two-point boundary condition to some extent in terms of a ray-shooting method/paraxial ray method and produce a perturbed ray in the vicinity of the receiver. At the source $\mathbf{r}_1(0) = 0$ the perturbation of initial slowness satisfies the perturbed Eikonal equation

$$\mathbf{p}_0 \cdot \mathbf{p}_1(0) = u_0 u_1, \tag{12}$$

which can be derived by expanding the Eikonal equation $u^2 = \mathbf{p}^2$ to first order.[14] From Eq. (12), we wish to trace a perturbed ray with

$$\mathbf{p}_1(0) = \frac{u_1}{u_0}\mathbf{p}_0. \tag{13}$$

When $|u_1/u_0| \ll 1$, we expect that the perturbed ray in the use of Eq. (13) deviates very slightly from the reference ray and may be thought of as a paraxial ray of the true perturbed ray joining a fixed source–receiver. Following Hu and Menke,[12] we assume that for a small $\mathbf{p}_1(0)$ defined in Eq. (13) the corresponding endpoint should be on the same wave front as the receiver. Therefore, perturbation of the slowness vector at the receiver is now approximated as

$$\mathbf{p}_1(S_0) = \frac{u_1}{u_0}\mathbf{p}_0 + \int_0^{S_0} \nabla u_1\, ds_0. \tag{14}$$

This expression is also equivalent to directly linearizing the integral of Eq. (1b) along a reference ray assuming the negligible effects of ray perturbation on the perturbation of the

slowness vector to first order for weakly inhomogeneous media (Farra, personal communication, 1999). In this case, we call Eq. (14) a quasi-linear formula for the use of polarization data defined by $\mathbf{p}_1(S_0)$. Its application is restricted to media with small velocity gradient. Currently, the applicability of this formula is well supported by the results of the synthetic tomographic experiment shown later.

For implementation of Eq. (11) as part of a two-dimensional tomographic scheme, the medium is digitized into a regular set of triangular cells and the slowness values are specified at the grid points of the cells. Inside each triangular cell, the slowness distribution varies linearly with the spatial coordinates x and z. The slowness field is thus continuous within the whole medium. With such digitization, the reference slowness field in the medium can be conveniently represented as

$$u_0(x,z)=\sum_{j=1}^{N} f_j(x,z)u_{0j}, \tag{15}$$

where $f_j(x,z)$ are N dimensionless basis functions and u_{0j} is the jth slowness coefficient. At the same time, the slowness perturbation has an expansion in the form

$$\Delta u(x,z)=\sum_{j=1}^{N} f_j(x,z)\Delta u_j. \tag{16}$$

Henceforth, the symbol ''Δ'' is now used before a letter to denote a perturbed quantity.

Inserting Eq. (16) into Eq. (11a), we obtain

$$\Delta t_i=\sum_{j=1}^{N} a_{ij}\Delta u_j, \quad i=1,\cdots,M, \tag{17}$$

where Δt_i is the traveltime perturbation or residual of the ith ray between the observed and theoretical traveltimes, M is the total number of rays, and

$$a_{ij}=\int_{L_{0i}} f_j(x,z)ds_0, \tag{18}$$

i.e., the integral of the jth basis function $f_j(x,z)$ along the ith ray path. We see that the a_{ij} are the elements of the traveltime Frechet derivative matrix.

Similarly, a substitution of Eq. (16) into Eq. (11b) leads to the equation

$$\Delta\mathbf{p}_i=\sum_{j=1}^{N} \mathbf{b}_{ij}\Delta u_j, \quad i=1,\cdots,M, \tag{19}$$

where $\Delta\mathbf{p}_i$ is the perturbation of the slowness vector of the ith ray at the receiver and calculated between the observed and the theoretical values, and

$$b_{ij}=f_j(x_{is},z_{is})\mathbf{n}_{is}+\int_{L_{0i}} \nabla f_j(x,z)ds_0, \tag{20}$$

where x_{is} and z_{is} are the source coordinates of the ith ray and $\mathbf{n}_{is}$ is the unit tangential vector of the ith ray at the source point in the reference medium. Now, the second term in Eq. (20) is the integral of the gradient $\nabla f_j(x,z)$ of the jth basis function along the ith ray path. $\mathbf{b}_{ij}$ are the elements of the approximate Frechet derivative matrix of the slowness vectors. The above line integrals in Eqs. (18) and (20) can be calculated by numerical integration for a number of ray segments of the ith ray in the discretized model.

In matrix notation, Eqs. (17) and (19) can be compactly written as

$$\begin{pmatrix}\Delta\mathbf{t}\\ \Delta\mathbf{p}\end{pmatrix}=\begin{pmatrix}A\\ B\end{pmatrix}\Delta\mathbf{u}, \tag{21}$$

where $\Delta\mathbf{t}$ is the M-dimensional residual traveltime column vector, A is the $M\times N$ matrix whose elements are given by Eq. (18), $\Delta\mathbf{p}$ is the $2M$-dimensional column vector of polarization, B the $2M\times N$ matrix with elements defined by Eq. (20), and $\Delta\mathbf{u}$ is the N-dimensional column vector of slowness perturbation coefficients. Equation (21) is the system of tomographic equations in matrix form that combine the traveltime and polarization data.

The basis functions we use are defined as follows. Denoting the position vectors of the three vertices of a triangular cell as $\mathbf{l}_k$, $\mathbf{l}_p$, $\mathbf{l}_q$, then the basis function for the vertex k is obtained as

$$f_k=\frac{1}{\Sigma}[|\mathbf{l}_p\times\mathbf{l}_q|+|\mathbf{l}\times(\mathbf{l}_p-\mathbf{l}_q)|], \tag{22}$$

where $\mathbf{l}$ is the position vector contained within the triangle and $\Sigma=|(\mathbf{l}_q-\mathbf{l}_k)\times(\mathbf{l}_q-\mathbf{l}_p)|$ is the area of a parallelogram whose two sides are $(\mathbf{l}_q-\mathbf{l}_k)$ and $(\mathbf{l}_q-\mathbf{l}_p)$. The basis functions corresponding to vertices p and q are obtained by permuting the indices in clockwise order of k, p, and q. These basis functions depend linearly on the position (x,z). Their sum is equal to one. Note that the basis functions defined by Eq. (22) are also valid for an irregular triangular parametrization model.

IV. TOMOGRAPHIC RECONSTRUCTION SCHEME

The implementation of nonlinear ray tomography for our situation comprises two fundamental steps: (1) constructing the matrices A and B and the theoretical traveltime and polarization data based on the previously derived slowness model; (2) solving the linear system of Eqs. (21) for the correction vector $\Delta\mathbf{u}$. Updating of the current model $\mathbf{u}_0$ by $\Delta\mathbf{u}$ is iterated until an acceptable solution is achieved or until a pre-defined number of iterations is completed. The first step is accomplished during the ray-tracing process. It is vital that a fast and efficient ray-tracing method be used in order to guarantee high performance of the tomographic procedure.

A. Ray tracing

Usual ray-tracing methods,[17,18,21,25,26] which are categorized into ray-shooting and ray-bending, cannot be applied easily in complicated inhomogeneous media. For example, the ray-shooting method has the problem of being slow to converge toward a specified point in the medium, or even fails when the point lies in a shadow zone. On the other hand, the ray-bending method can have difficulty in obtaining the global minimum traveltime paths that are required for first arrivals. To overcome these difficulties in the classical approaches, a number of grid traveltime methods have been

developed in the last decade. The basic idea of these methods is to construct traveltime fields on a regular grid of nodes, and then any two-point ray is readily found from this traveltime field information. Of this type of approach, the network ray-tracing method[27] is a suitable representative example. This method is extremely robust and can always locate the global minimum traveltime ray path from its origin to its destination for either direct or refracted arrivals. Its computer implementation is quite efficient and flexible. We adapt it here to carry out the two-point ray tracing required by our scheme.

The principle of network ray tracing is briefly the following. The fundamental task is to construct an appropriate network for a physical problem. In this paper, we consider the cell organization of networks[27] because it is particularly suitable for application to ray tomography. In a similar way to the rectangular cell network model, the network for the triangular cell model is set up by positioning nodes along the sides of each triangular cell, connecting nodes inside each triangular cell, and giving each connection a weight equal to the traveltime for that ray segment. The shortest path algorithm can then be used to quickly scan the minimum traveltime nodes and recursively complete the calculation of the entire first-arrival time field. This facilitates the determination of all information required by the tomographic system for a given model.

B. A conjugate gradient iterative algorithm

Now let us turn to the problem of solving Eq. (21). It is widely recognized that a tomographic solution is not unique. Also, the data do not lie in the range space of a matrix due to the presence of observational errors and the model approximations. A single ray from a source to a receiver being a line will transverse only a small fraction of the whole model, so that each row of a Frechet matrix will contain many zero elements. As a result, a tomographic inversion system is ill-posed, inconsistent, and sparse. To accommodate these attributes, the inverse problem is often cast as an optimization problem,[28–30] as we do below.

First of all, splitting the matrix B in Eq. (21) into the two submatrices B_x and B_z relating to the column vectors $\Delta\mathbf{p}_x$ and $\Delta\mathbf{p}_z$ of the horizontal and vertical polarization, we have

$$\begin{pmatrix} \Delta\mathbf{t} \\ \Delta\mathbf{p}_x \\ \Delta\mathbf{p}_z \end{pmatrix} = \begin{pmatrix} A \\ B_x \\ B_z \end{pmatrix} \Delta\mathbf{u}. \tag{23}$$

An optimum solution to Eq. (23) is obtained by minimizing the following objective function of $\Delta\mathbf{u}$:

$$\begin{aligned} F(\Delta\mathbf{u}) = {} & \lambda_t(\Delta\mathbf{t}-A\Delta\mathbf{u})^T(\Delta\mathbf{t}-A\Delta\mathbf{u}) \\ & + \lambda_x(\Delta\mathbf{p}_x-B_x\Delta\mathbf{u})^T(\Delta\mathbf{p}_x-B_x\Delta\mathbf{u}) \\ & + \lambda_z(\Delta\mathbf{p}_z-B_z\Delta\mathbf{u})^T(\Delta\mathbf{p}_z-B_z\Delta\mathbf{u}) + \lambda\Delta\mathbf{u}^T\Delta\mathbf{u}, \end{aligned} \tag{24}$$

where the first three terms are the data misfits which measure the discrepancies between the observed and theoretical quantities for each set of data (traveltime, horizontal, and vertical polarization data), and the last term pertains to the model perturbation constraints. The scalars λ_t, λ_x, and λ_z are introduced to control the relative importance of three sets of data and may be called the data balance factors. Similarly, λ is the model balance factor or simply damping factor and has the same role as the data balance factors. Differentiating Eq. (23) and setting the result to zero yields

$$(G^TG+\lambda I)\Delta\mathbf{u} = G^T\Delta\mathbf{d}, \tag{25}$$

where

$$G = \begin{pmatrix} \sqrt{\lambda_t}A \\ \sqrt{\lambda_x}B_x \\ \sqrt{\lambda_x}B_x \end{pmatrix}, \quad \Delta\mathbf{d} = \begin{pmatrix} \sqrt{\lambda_t}\Delta\mathbf{t} \\ \sqrt{\lambda_x}\Delta\mathbf{p}_x \\ \sqrt{\lambda_x}\Delta\mathbf{p}_x \end{pmatrix}.$$

Equation (25) is the damped least-squares normal equation for $G\Delta\mathbf{u}=\Delta\mathbf{d}$.

At this point, the problem of minimizing the objective functional Eq. (24) is reduced to solving Eq. (25). There are several basic reconstruction methods available to do this, including the algebraic reconstruction technique (ART), the simultaneous iterative reconstruction technique (SIRT), the conjugate gradient (CG) method, and the singular value decomposition (SVD) method.[28,31,32] The iterative solvers (ART, SIRT, and CG) require smaller computer memory and run faster than the SVD method. In addition, the CG type of algorithm is superior to ART and SIRT as analyzed in Ref. 32. For these reasons, in this section we choose to develop a new CG type of algorithm for obtaining the solution to Eq. (25).

The conjugate gradient method, originally propounded by Hestenes and Stiefel,[33] has several variants in which CGLS and LSQR[34] are the most popular in solving traveltime tomography problems.[35–37] Both CGLS and LSQR have similar qualitative properties.[34,36,37] Their fundamental implementation is to avoid direct calculation of G^TG, which can lead to not only the loss of information due to round-off but also poor solutions on ill-conditioned systems since the condition number of G^TG is the square of that of G.[33,34] In principle, either standard CGLS or LSQR could be chosen as a matter of taste to solve a damped and weighted least-squares problem (nonstandard least-squares problem) by having to form an augmented linear system. Based on the implementation idea of CGLS and the CG type algorithm of Song and Zhang[38] for the solution of such nonstandard problems, we propose a new conjugate gradient algorithm which preserves the operational features of CGLS without the need to construct augmented matrices during the computations. The proposed algorithm corresponding to Eq. (25) is as follows:

(1) Initialize: $\Delta\mathbf{u}_0=0$, $\mathbf{h}_0=\mathbf{f}_0=(\mathbf{d}-G\mathbf{u}_0)$, and $\mathbf{p}_0=\mathbf{r}_0=G^T\mathbf{h}_0=G^T\mathbf{f}_0$;

(2) Iterate for $k=0,1,2,\cdots$

$$\begin{aligned} & \mathbf{q}_k = G\mathbf{p}_k, \\ & \alpha_k = (\mathbf{r_k},\mathbf{r}_k)(\mathbf{q}_k^T\mathbf{q}_k+\lambda\mathbf{p}_k^T\mathbf{p}_k)^{-1}, \\ & \Delta\mathbf{u}_{k+1} = \Delta\mathbf{u}_k+\alpha_{k+1}\mathbf{p}_k, \\ & \mathbf{h}_{k+1} = \mathbf{h}_k-\alpha_k(\mathbf{q}_k+\lambda\mathbf{f}_k), \\ & \mathbf{r}_{k+1} = G^T\mathbf{h}_{k+1}, \end{aligned} \tag{26}$$

$$\beta_k=(\mathbf{r}_{k+1},\mathbf{r}_{k+1})(\mathbf{r}_k,\mathbf{r}_k)^{-1},$$

$$\mathbf{f}_{k+1}=\mathbf{h}_{k+1}+\beta_k\mathbf{f}_k,$$

$$\mathbf{p}_{k+1}=\mathbf{r}_{k+1}+\beta_k\mathbf{p}_k=G^T\mathbf{f}_{k+1}.$$

This is the conjugate gradient algorithm for damped least-squares (CGDLS), while the damping factor and the data balance factors (using the expansion of G) are now explicitly incorporated. As a result, the sparseness and the structure of the matrix is still fully retained. For the use of multiple data sets in our case, the matrix-vector product in this algorithm is most efficiently implemented via the three sub-matrices, A, B_x, B_z, without the need to assemble the whole matrix G. These facilitate the solution of the damped least-squares problem via the CG method.

In Eq. (26), $\mathbf{r}_k$ is the residual vector for the normal equations, i.e.,

$$\mathbf{r}_k=G^T\Delta\mathbf{d}-(G^TG+\lambda I)\Delta\mathbf{u}_k. \tag{27}$$

By the repeated use of $\mathbf{h}_{k+1}=\mathbf{h}_k-\alpha_k(\mathbf{q}_k+\lambda\mathbf{f}_k)$ in Eq. (25), we may express the introduced vector $\mathbf{h}_k$ as

$$\mathbf{h}_k=\mathbf{d}-G\mathbf{u}_k-\lambda\sum_{i=0}^{k-1}\alpha_k\mathbf{f}_k. \tag{28}$$

In structure, the algorithm Eq. (26) is similar to the standard CGLS algorithm except for the explicit use of the intermediate vector $\mathbf{f}_k$ and the damping factor λ. Thus the CGDLS algorithm is easily implemented with minor modifications to available CGLS code. In fact when $\lambda=0$, the algorithm Eq. (26) recovers the CGLS algorithm[33,34] where $\mathbf{h}_k=\mathbf{d}-G\mathbf{u}_k$ is the residual vector for the original equations $G\mathbf{u}=\mathbf{d}$, and the vector $\mathbf{f}_k$ is no longer required.

V. NUMERICAL RESULTS AND DISCUSSIONS

In this section, we numerically test our proposed tomographic scheme for jointly inverting traveltime and polarization data. The true model used has the dimensions 13×13 m^2. It is first divided into 1×1 m^2 square cells and then subdivided into triangular cells. Thus there are a total of 196 slowness or velocity unknowns to be inverted over the triangular grid of points in this tomographically parameterized model. We design an observing system in which both sources and receivers are positioned on the left- and right-hand sides of the model at intervals of 1 m. It follows that a total of 169 rays will be generated from left to right in this imaging system [Fig. 1(a)]. The choice of such a two-sided observing geometry is its relevance to practical applications in acoustical/seismic engineering investigations. Also, from a numerical point of view this corresponds to a class of typically ill-posed inverse problems on which it is ideal to illustrate a tomography method.

In ray-theoretic tomography, the measured first arrival data (e.g., traveltimes) generally contain insufficient information to resolve the features of a low-velocity object. When carrying out a reconstruction from a starting model, it is therefore more likely that an inaccurate image will be produced for a low-velocity inclusion than for a high-velocity body.[29–31,34] Our numerical simulations also show this point. Therefore, we will only discuss the reconstruction performance where the target model is assumed to have a low-velocity object centrally embedded in the background medium where the velocity is 5000 m/s. This model could represent a flaw in nondestructive evaluation or an anomaly in rock engineering. We specify the object as having a velocity of 4500 m/s or 4000 m/s at the 4×4 grid points of the cells. The model parameterization is chosen to ensure that the velocity transition between the background and the object is reasonably broad and smooth. The fractional velocity contrasts 10% and 20% between the background and the object can be regarded as conforming to weak and mild inhomogeneity, respectively. In the synthetic tests we also simulate noisy data by adding random errors to the exact traveltime and polarization data calculated from the true model. In the tomographic inversion, a global iteration may be defined as the process of updating a reference or current model. Within a global iteration, there is the CG iteration described by Eq. (26) to solve the linear system Eq. (25). Below, all iterative, nonlinear reconstructions in the different cases are conducted with a starting homogeneous model of 5000 m/s and a damping factor of 0.0001. This small damping factor is used to weaken the effects of the model constrained term in Eq. (24), since the main issue in this paper is concerned with the feasibility and capability of obtaining reconstructions from traveltime and polarization data in our proposed scheme. So the data balance factors λ_t, λ_x, and λ_z in Eq. (24) are simply chosen as either the combinations of 0 and 1 or both 1. Such choices of the data balance factors then are equivalent to the use of three sets of data: (1) traveltime data; (2) polarization data; (3) joint traveltime and polarization data. The reconstructed images are shown at global iteration 10.

We first consider noise-free synthetic data. Figure 1(a) shows the target model of 10% velocity contrast and the corresponding whole ray paths. Figures 1(b)–(d) are the results from the three sets of data with ten CG iterations. In Fig. 1(b) with only traveltime data and Fig. 1(d) with both traveltime and polarization data, the low-velocity object is well recovered. Both figures are superior to Fig. 1(c) which is derived from polarization data alone. Indeed the low-velocity object in Fig. 1(c) is not easily discerned as compared with those in Figs. 1(b) and (d). Figure 1(c), however, shows that polarization data are helpful in constraining the boundaries of the anomaly. The object shape in Fig. 1(d) appears better resolved than that in Fig. 1(b).

The CG iteration number has an effect on the traveltime-reconstructed image.[36,38] Too many CG iterations may lead to the solution divergence. For example, using a CG iteration of 20, the traveltime reconstruction becomes sensitive to the increased iteration number and the image in Fig. 1(e) has deteriorated. Nevertheless, reconstruction of joint traveltime and polarization data is stable with the same CG iterations as shown in Fig. 1(f), which is quite similar to Fig. 1(d).

Note that the horizontal resolution is low in these images, as exhibited in Fig. 1(a), due to the lack of vertical rays from top to bottom. Thus angular aperture limitation results in the artificial ''X'' shape in Figs. 1(b)–1(f).

For the case of 20% velocity contrast, the traveltime

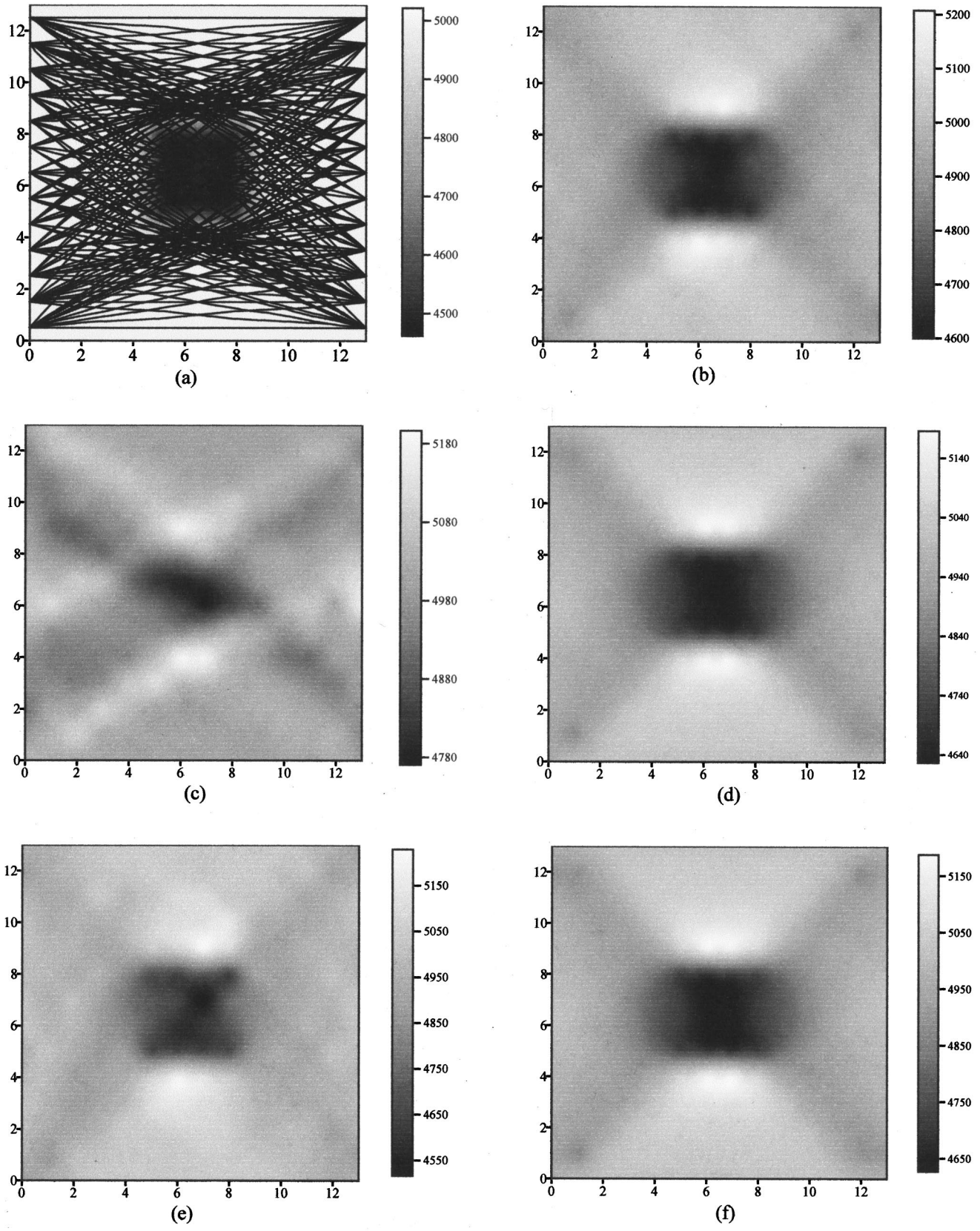


FIG. 1. (a) Target model of 10% velocity contrast and the whole ray paths. For the cases of noise-free synthetic data and no magnitude constraint, the reconstructions are conducted on different sets of data: (b) traveltime, (c) polarization, (d) traveltime+polarization, (e) traveltime, (f) traveltime+polarization. (b)–(d) and (e)–(f) are at CG iterations 10 and 20, respectively.

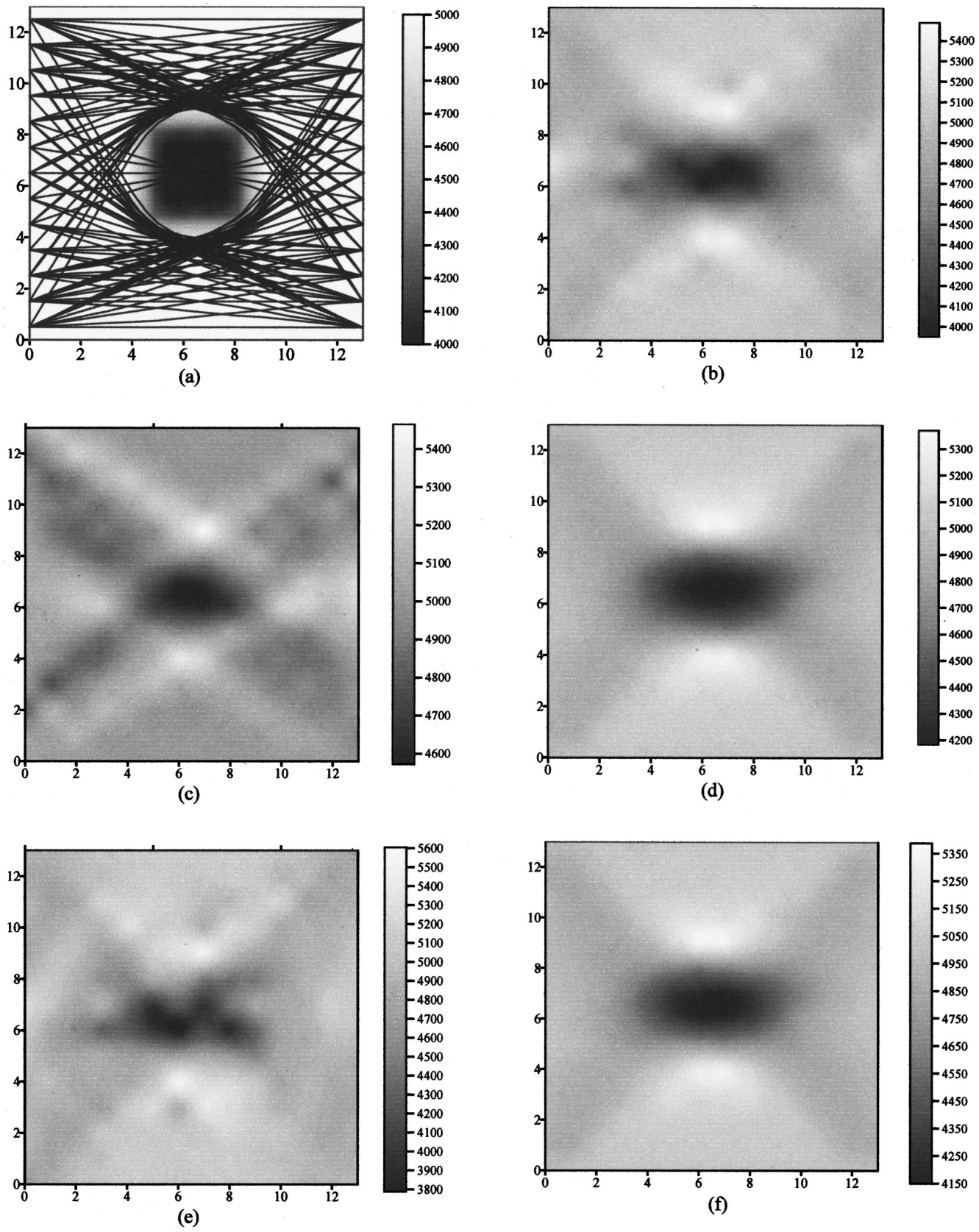


FIG. 2. As in Fig. 1, but for 20% velocity contrast.

reconstruction is not as good as that of 10% velocity contrast because now the degree of nonlinearity is much greater. The shape of the low-velocity object recovered in Fig. 2(b) is significantly elongated along the horizontal direction, due to the strong effect of ray bending and the limited two-sided observing system, as illustrated in Fig. 2(a). In this case, the low-velocity body is the poorly constrained part of the model because of limited ray sampling. In comparison, the image of the body is relatively well constrained in the case of small velocity contrast [Fig. 1(a)]. Using polarization data alone, the reconstructed image patterns for the mild contrast is basically similar to that of small contrast. But the low-velocity object in Fig. 2(c) seems somewhat clearer than in Fig. 1(c). This may result from the response of the polarization data to slowness gradient variations along the ray paths. Some useful information on the low-velocity object is evidently contained in the polarization data. When performing joint inversion of traveltime and polarization data, the inverted image in Fig. 2(d) is better and smoother. In particular, the image of the low-velocity object has considerably improved over that in Fig. 2(b). Also, on increasing the CG iteration to 20, we see that the traveltime-reconstruction image in Fig. 2(e) has become very poor, while the image in Fig. 2(f) from joint traveltime and polarization data remains the same as that in Fig. 2(d).

Next, we carry out the reconstruction of our object on noisy synthetic data. For 10% velocity contrast, the noisy data are simulated by adding Gaussian errors with standard deviations of 0.01 ms and 0.01 ms/m to the traveltime and polarization data generated from the corresponding target model. For the set of noisy data, the inverted images with ten CG iterations are shown in Figs. 3(a)–(b). The resulting traveltime image of Fig. 3(a) is quite blurred and the low-velocity object can barely be distinguished from the background. In contrast, the use of both traveltime and polarization data significantly improves the reconstruction as seen in Fig. 3(b). Indeed, by incorporating the polarization information into the reconstruction, the artifacts in Fig. 3(a) are now removed to a large extent, and the low-velocity object is revealed in Fig. 3(b). Applying 20 CG iterations to reconstruction, the images derived from traveltime data alone and the joint traveltime and polarization data are shown in Figs. 3(c)–(d), respectively. Figure 3(c) is so seriously distorted that the low-velocity body is almost unrecognizable, while Fig. 3(d) maintains approximately the same quality as Fig. 3(b). Naturally, the presence of noise in the data does degrade the image.

Compared to the true models, the above reconstructed images vary with different velocity magnitudes although the low-velocity object could be better distinguished from the background when adding polarization data. Such nonuniqueness of reconstructing velocity distribution is inherent, as discussed by Coen,[39] even with complete and precise first-arrival time tomography data, because part of the imaging region may not be covered by any first-arrival rays [e.g., Fig. 2(a)]. With the addition of noise in the data, poor angular coverage, and the limitations of the reconstruction algorithm itself, it is very difficult to uniquely determine the true slowness distribution of a medium.

In spite of these obstacles, the extent of nonuniqueness in the inverse problem can be minimized by adding some *a priori* knowledge. For example, in order to robustly perform nonlinear traveltime tomography, Berryman[29,30] constrained the solution to be in the feasible, convex region of the model space, which is governed by Fermat's principle and the positivity of slowness. Recently, Delsanto *et al.*[40] proposed a special ultrasonic traveltime tomography method to accurately predict the geometry of an inclusion in a homogeneous specimen by combining a genetic algorithm and the known physical properties of both the specimen and the inclusion. Within the notion of convex projection reconstruction,[41,42] the solution being sought must lie in the intersection of all the convex constraint sets defined by the allowable data and certain types of prior information. Therefore, using more constraints, the area of the intersection becomes smaller and the set of feasible solutions may tend to approach the true solution. This idea can be observed in the comparison of the results for traveltime data alone and joint traveltime and polarization data.

In the following, we shall further impose a magnitude constraint set[41,42] on each image pixel element to test the reconstruction's performance. Such a convex constraint set is used to make the reconstructed slowness values at each global iteration fall within ranges whose lower and upper limits can be specified as available prior information. Now suppose we have an ideal estimate of these limits in N-dimensional slowness model space in our synthetic examples, and simply postulate that $4500\ \text{m/s} \leqslant v_j \leqslant 5000\ \text{m/s}$ $(j=1,\cdots,N)$ for 10% velocity contrast and $4000\ \text{m/s} \leqslant v_j \leqslant 5000\ \text{m/s}$ $(j=1,\cdots,N)$ for 20% velocity contrast.

We continue to apply the previous noisy data to the reconstruction for the 10% velocity contrast. Figure 3(e) is the result for the traveltime data and the magnitude constraint. In comparison with Fig. 3(a), Fig. 3(e) is significantly improved. The low-velocity object is clearly defined despite there still being noise in the background. The best image quality, however, is derived from the traveltime and polarization data and the magnitude constraint as displayed in Fig. 3(f), in which both the low-velocity object and the background are close to the target solution.

To conclude this section, we consider another synthetic set of noisy theoretical data calculated for the target model of 20% velocity contrast with random errors of 0.04 ms and 0.04 ms/m standard deviations. For this large velocity contrast and the increased level of noise in the data, the reconstruction is bound to be poor. Figure 4(a), which results from the traveltime data and the magnitude constraint, is now very noisy and the low-velocity object is poorly determined. This worse situation can, however, be alleviated to a certain degree by incorporating polarization data into the reconstruction. The result is shown in Fig. 4(d), in which the low-velocity object is now well located.

As these examples demonstrate, with noise-free and noisy data, the additional use of polarization data in tomographic reconstruction not only improves the image quality but also has a stabilizing effect on the iterative solution with increasing CG iteration number.

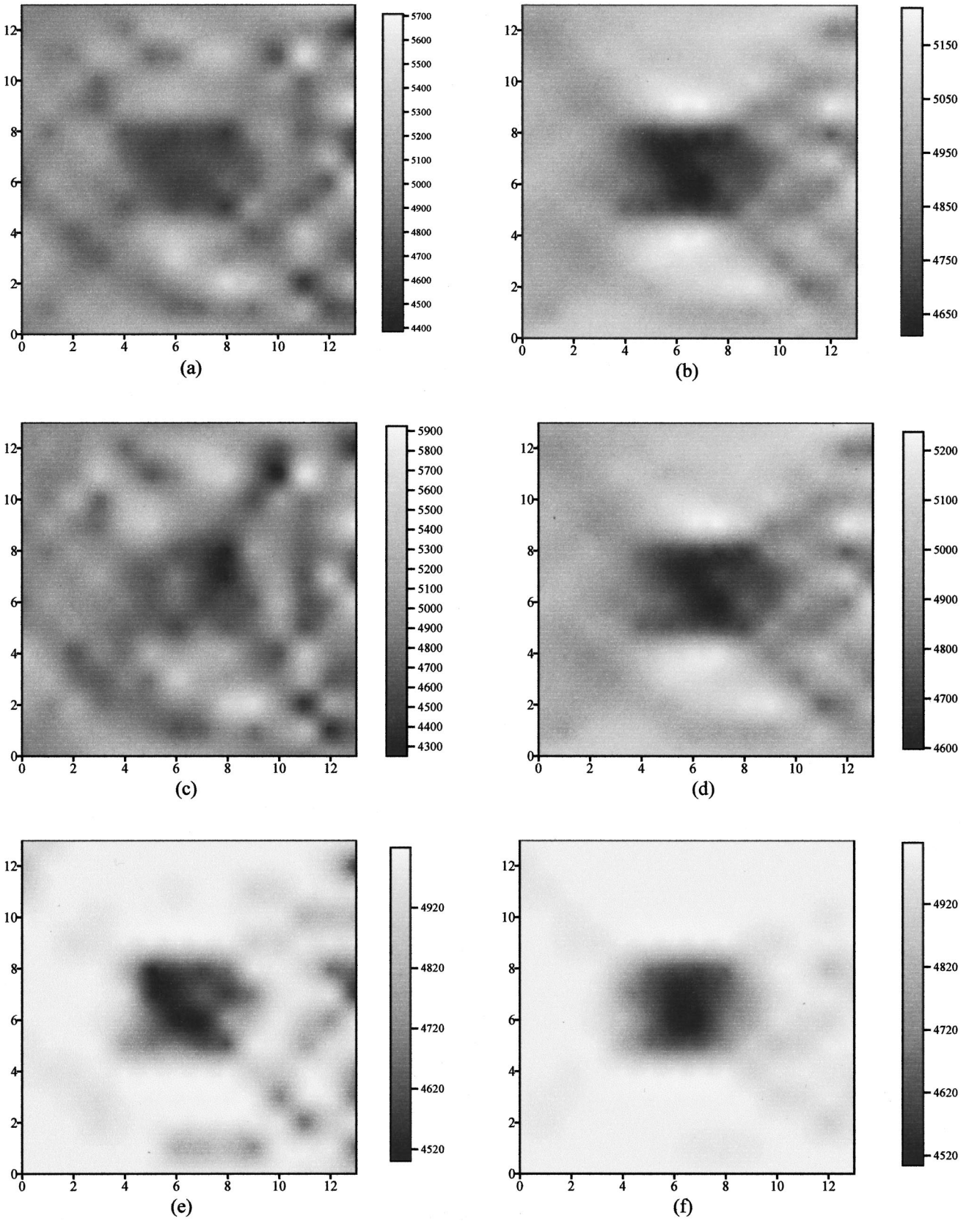


FIG. 3. The results of 10% velocity contrast are from noisy synthetic data of 0.01 ms and 0.01 ms/m standard deviation without [(a)–(d)] and with [(e)–(f)] magnitude constraint. (a) Traveltime, (b) traveltime+polarization, (c) traveltime, (d) traveltime+polarization, (e) traveltime, (f) traveltime+polarization. (a)–(b) and (e)–(f) are at CG iterations 10. (c)–(d) are at CG iterations 20.

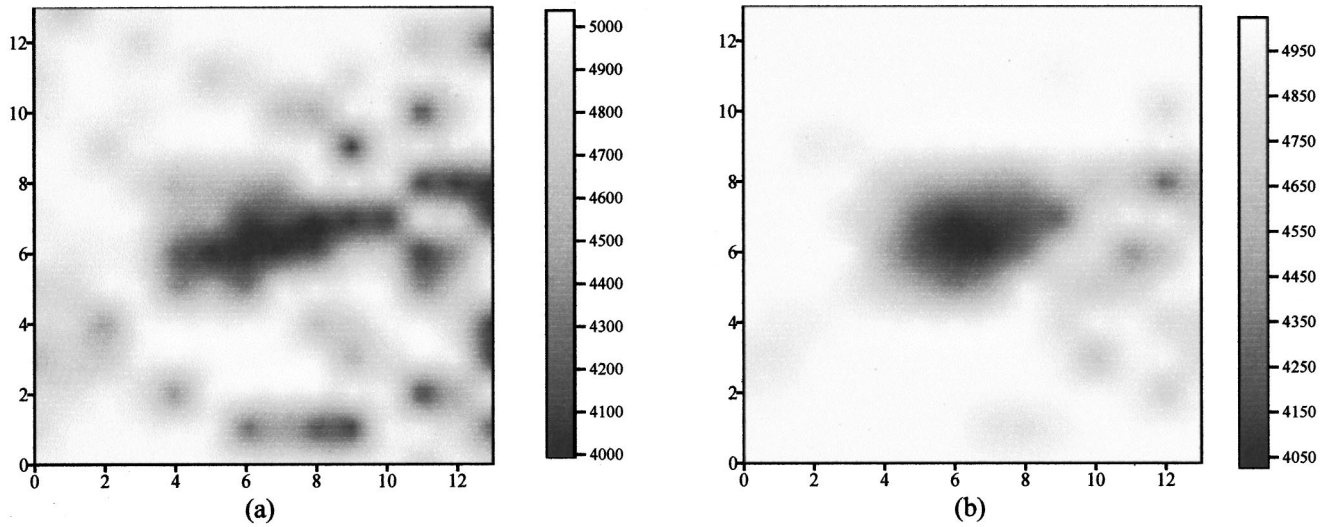


FIG. 4. The results of 20% velocity contrast are from the noisy synthetic data of 0.04 ms and 0.04 ms/m standard deviation with magnitude constraint. (a) Traveltime, (b) traveltime+polarization.

VI. CONCLUSION

We have developed a simplified tomography method for inverting joint *P*-wave traveltime and polarization data in isotropic media. Particular emphasis has been given to the treatment of polarization data. We have taken a quasi-linear approximation in which the effect of ray position perturbation on the perturbed slowness vector to first order is ignored in the integral for weakly inhomogeneous media. Then a linear integral relation is approximately formulated between the first-order perturbation of the ray slowness vector and the gradient of the medium slowness perturbation along a reference ray. The numerical tomography experiments justify this quasi-linear approximation. To solve the tomography problem of multiple data sets, a new CG-type iterative algorithm, which is an extension of the CGLS algorithm,[33,34] has been proposed that explicitly incorporates a damping factor and data value balance factors. In this algorithm, the submatrices related to each data point are used separately and their sparseness is retained; no construction of augmented matrices is required, while the numerical properties of CGLS in the solution of a damped least-squares problem are retained. These features make the proposed algorithm effective and flexible. To test the proposed tomography scheme, we have conducted a series of numerical experiments to reconstruct a low-velocity object in a two-sided (cross-hole) observing geometry. For 10% and 20% velocity contrasts, the behavior of the combined traveltime and polarization tomography is superior to that of simple traveltime tomography, regardless of whether the data are noise-free or noisy, with and without adding a magnitude convex constraint set. The numerical results demonstrate that the additional use of polarization data not only improves the image quality but also has a stabilizing effect on the iterative tomography solution, even when increasing the CG iteration number.

Our proposed tomography method is easy to implement, and has been shown to be feasible for weakly and mildly inhomogeneous media.

ACKNOWLEDGMENTS

L.P.S. was supported by a University of the Witwatersrand postdoctoral research fellowship, and funding was also provided for this research by the Bernard Price Institute for Geophysical Research, University of the Witwatersrand. L.P.S. is particularly grateful to Dr. V. Farra of the Institut de Physique du Globe de Paris for her valuable comments on an early version of this paper. Special thanks are given to Professor M. Koch of the University of Kassel for kindly supporting the revision work on this paper. The authors also thank the two anonymous reviewers for their useful comments and suggestions.

[1] R. K. Mueller, M. Kaven, and G. Wade, ''Reconstructive tomography and applications to ultrasonics,'' Proc. IEEE **67**, 567–587 (1979).
[2] N. Sponheim, L. J. Gelius, I. Johanson, and J. J. Stamnes, ''Ultrasonic tomography of biological tissue,'' Ultrason. Imaging **16**, 19–32 (1994).
[3] T. D. Mast, ''Wideband quantitative ultrasonic imaging by time-domain diffraction tomography,'' J. Acoust. Soc. Am. **106**, 3061–3071 (1999).
[4] C. Gelbke, F. Miranda, and G. Sattel, ''Results of a seismic transmission tomography survey at the Grimsel Rock Laboratory,'' The Log Analyst **30**, 243–260 (1989).
[5] A. J. Witten and W. C. King, ''Acoustic imaging of subsurface features,'' ASCE J. Env. Eng. Div. **116**, 166–181 (1990).
[6] T. Watanabe and K. Sassa, ''Seismic attenuation tomography and its application to rock mass evaluation,'' Int. J. Rock Min. Sci. **33**, 467–477 (1996).
[7] T. Yamamoto, ''Velocity variabilities and other physical properties of marine sediments measured by crosswell acoustic tomography,'' J. Acoust. Soc. Am. **98**, 2235–2248 (1995).
[8] B. Rapids, T. Nye, and T. Yamamoto, ''Pilot experiment for the acquisition of marine sediment properties via small scale tomography system,'' J. Acoust. Soc. Am. **103**, 212–224 (1998).
[9] W. Munk, P. Worchester, and C. Wunsch, *Ocean Acoustic Tomography* (Cambridge University Press, New York, 1995), and references therein.
[10] M. Slaney, A. C. Kak, and L. Larsen, ''Limitations of imaging with first-order diffraction tomography'' IEEE Trans. Microwave Theory Tech. **32**, 860–873 (1984).
[11] A. Witten and J. Tuggle, ''A practical approach to ultrasonic imaging using diffraction tomography,'' J. Acoust. Soc. Am. **83**, 1645–1652 (1988).
[12] G. Hu and W. Menke, ''Formal inversion of laterally heterogeneous velocity structure from *P*-wave polarization data,'' Geophys. J. Int. **110**, 63–69 (1992).

[13] G. Hu, W. Menke, and C. Powell, ''Polarization tomography for *P*-wave velocity structure in southern California,'' J. Geophys. Res. **99**, 15245–15256 (1994).
[14] V. Farra and S. L. Begat, ''Sensitivity of *qP*-wave traveltimes and polarization vectors to heterogeneity, anisotropy, and interfaces,'' Geophys. J. Int. **121**, 371–384 (1995).
[15] S. L. Begat and V. Farra, ''*P*-wave traveltime and polarization tomography of VSP data,'' Geophys. J. Int. **131**, 100–114 (1997).
[16] V. Farra and R. Madariaga, ''Seismic waveform modeling in heterogeneous media by ray perturbation theory,'' J. Geophys. Res. **92**, 2697–2712 (1987).
[17] J. Virieux, ''Fast and accurate ray tracing by Hamiltonian perturbation,'' J. Geophys. Res. **96**, 579–594 (1991).
[18] V. Farra, ''Bending method revisited: A Hamiltonian approach,'' Geophys. J. Int. **109**, 138–150 (1992).
[19] R. Snieder and M. Sambridge, ''Ray perturbation theory for travel times and ray paths in 3-D heterogeneous media,'' Geophys. J. Int. **109**, 294–322 (1992).
[20] R. Snieder and M. Sambridge, ''The ambiguity in ray perturbation theory,'' J. Geophys. Res. **98**, 22021–22034 (1993).
[21] J. Pulliam and R. Snieder, ''Fast, efficient calculation of rays and traveltimes with ray perturbation theory,'' J. Acoust. Soc. Am. **99**, 383–391 (1996).
[22] M. Born and E. Wolf, *Principle of Optics* (Pergamon, Oxford, 1980).
[23] B. J. Moore, ''Seismic rays in media with slight variation in velocity,'' Geophys. J. Int. **105**, 213–227 (1991).
[24] B. J. Moore, ''Analytical solutions for the ray perturbations in depth-varying media,'' Geophys. J. Int. **114**, 281–288 (1993).
[25] A. H. Andersen and A. C. Kak, ''Digital ray tracing in two-dimensional refractive field,'' J. Acoust. Soc. Am. **72**, 1593–1606 (1982).
[26] C. Cerveny, ''Ray tracing algorithms in three-dimensional laterally varying layered structures,'' in *Seismic Tomography*, edited by G. Nolet (Reidel, New York, 1987), pp. 99–133.
[27] T. J. Moser, ''Shortest path calculation of seismic rays,'' Geophysics **56**, 59–67 (1991).
[28] G. T. Herman, *Image Reconstruction from Projections* (Academic, New York, 1980).
[29] J. G. Berryman, ''Fermat's principle and nonlinear traveltime tomography,'' Phys. Rev. Lett. **62**, 2953–2956 (1989).
[30] J. G. Berryman, ''Stable iterative reconstruction algorithm for nonlinear traveltime tomography,'' Inverse Probl. **6**, 21–42 (1990).
[31] R. S. Schechter, R. B. Mignogna, and P. P. Delsanto, ''Ultrasonic tomography using curved ray paths obtained by wave propagation simulations on a massively parallel computer,'' J. Acoust. Soc. Am. **100**, 2103–2111 (1996).
[32] A. Van der Sluis and H. A. Van der Vorst, ''Numerical solution of large sparse linear algebraic systems arising from tomographic problems,'' in *Seismic Tomography*, edited by G. Nolet (Reidel, New York, 1987), pp. 49–84.
[33] M. Hestenes and E. Stiefel, ''Methods of conjugate gradients for solving linear systems,'' J. Res. Natl. Bur. Stand. **49**, 409–436 (1952).
[34] C. C. Paige and M. A. Saunders, ''LSQR: an algorithm for sparse linear equations and sparse least squares,'' ACM Trans. Math. Softw. **8**, 43–71 (1982).
[35] G. Nolet, ''Seismic wave propagation and seismic tomography,'' in *Seismic Tomography*, edited by G. Nolet (Reidel, New York, 1987), p. 1.23.
[36] J. A. Scales, ''Tomographic inversion via the conjugate gradient method,'' Geophysics **52**, 179–185 (1987).
[37] J. A. Scales, ''Fast l_p solution of large, sparse, linear systems: application to seismic travel time tomography,'' J. Comput. Phys. **75**, 314–333 (1988).
[38] L.-P. Song and S.-Y. Zhang, ''Stabilizing the iterative solution to ultrasonic transmission tomography,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**, 1117–1122 (1998).
[39] S. Coen, ''White holes in the Fresnel zone causing ambiguous first arrival time tomography interpretation,'' Phys. Rev. Lett. **73**, 3219–3222 (1994).
[40] P. P. Delsanto, A. Romano, and M. Scalerandi, ''Application of genetic algorithms to ultrasonic tomography,'' J. Acoust. Soc. Am. **104**, 1374–1381 (1998).
[41] D. C. Youla and H. Webb, ''Image restoration by the method of convex projections, Part I—theory,'' IEEE Trans. Med. Imaging **1**, 81–94 (1982).
[42] H. Stark and P. Oskoui, ''High-resolution image recovery from image-plane arrays using convex projections,'' J. Opt. Soc. Am. **6**, 1715–1726 (1989).

Visualization of multiple incoherent sources by the backward prediction of near-field acoustic holography

Kyoung-Uk Nam and Yang-Hann Kim[a)]
Department of Mechanical Engineering, Center for Noise and Vibration Control (NOVIC), Korea Advanced Institute of Science and Technology (KAIST), Science Town, Taejon-Shi, 305-701, Korea

(Received 25 January 2000; revised 1 February 2001; accepted 2 February 2001)

When there are multiple noise sources which share the same frequency bands, a sound field measured or estimated by near-field acoustic holography (NAH) is obviously the combination of the sound fields generated by the multiple sources. The objective of this paper is to estimate the sound field generated by each source when the coherence functions between sources are zeros. This objective can be achieved by obtaining signals coherent to sources. This paper proposes a method to obtain the coherent signals by using the spatial information of NAH. The proposed method obtains the coherent signals from pressure signals at source positions estimated by the backward prediction of NAH. Thus it does not require any prior information on source positions, unlike the conventional method, which placed sensors close to sources in order to obtain the coherent signals. The proposed method was verified by a numerical simulation using two incoherent monopole sources. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1358888]

PACS numbers: 43.20.−f, 43.50.Rq, 43.60.Cg, 43.60.Sx [ANN]

I. INTRODUCTION

When an acoustic engineer tries to reduce noise at a point, he/she will first measure pressure at the point. If the noise is stationary random, then he/she will estimate the spectrum in order to know the frequency characteristics. This is because a noise control strategy significantly depends on target frequency; i.e., the frequency that he/she wants to control. When there are multiple noise sources with the same frequency bands, the engineer will also want to know the contributions of the sources to the spectrum. This is because a noise control strategy also significantly depends on a target source; i.e., the source that he/she prefers to control. However, a spectrum does not give any information on the contributions.

Near-field acoustic holography[1] (NAH) has the same problem. NAH enables one to estimate the spectra of pressure and velocity at the region of interest even when there are multiple noise sources.[2] However, it does not provide any information on the contributions of sources. Figure 1 is an example. It shows the auto-spectra of pressure at a selected frequency on a hologram plane when there are two arbitrary sources. It is difficult to estimate the contributions from Fig. 1. This is because the sound field by a source overlaps that by the other.

Recently, contribution analysis has been one of the interests of NAH. References 3–7 are some of them. They assumed that sources are incoherent to one another, and tried to analyze contributions by placing sensors close to the sources.

The incoherent assumption means that the coherence functions between sources are zeros. It can be applied to independent sources which do not have any causal relations, such as the fan and the disk drive noise of a computer,[3] and the leading and the trailing edge noise of a tire.[6] We also use this assumption.

Placing sensors close to sources is required in order to apply a noise rejection method[8–11] to contribution analysis. This strategy regards the effect of a target source on outputs as a true signal, and that of other sources as noise. And, by comparing the outputs with the sensor signal close to the target source, it extracts the true signal or the contribution of the target source. This is because the target source dominates the closest sensor signal. However, it requires that one know where the sources are before applying NAH. It contradicts one of the important goals of NAH, the identification of source positions.

Our objective is to estimate the contributions of multiple incoherent sources in NAH without any prior information on source positions. In order to achieve this objective, we propose a method to use the spatial information of a hologram. We explain the method by the vector notation introduced in Ref. 12. It is reviewed and reinterpreted in Sec. II A for the completeness of this paper. We formulate the method for NAH based on a step-by-step measurement[2,13,14] using multiple references. For the sake of simplicity, we deal with a two-input problem.

II. PROBLEM DEFINITION

A. The review of vector notation

Let us consider two stationary random signals, $u(t)$ and $v(t)$. They have N samples, $u^{(1)}(t),\ldots,u^{(N)}(t)$, and $v^{(1)}(t),\ldots,v^{(N)}(t)$, measured simultaneously from 0 to T second. Their finite Fourier transforms are $u^{(1)}(f),\ldots,u^{(N)}(f)$, and $v^{(1)}(f),\ldots,v^{(N)}(f)$. If T and N are large enough, one can estimate their spectra as[8]

$$s_{uu}(f)=\frac{1}{TN}\sum_{i=1}^{N} u^{(i)*}(f)u^{(i)}(f), \tag{1a}$$

$$s_{vv}(f)=\frac{1}{TN}\sum_{i=1}^{N} v^{(i)*}(f)v^{(i)}(f), \tag{1b}$$

a)Electronic mail: yhkim@mail.kaist.ac.kr

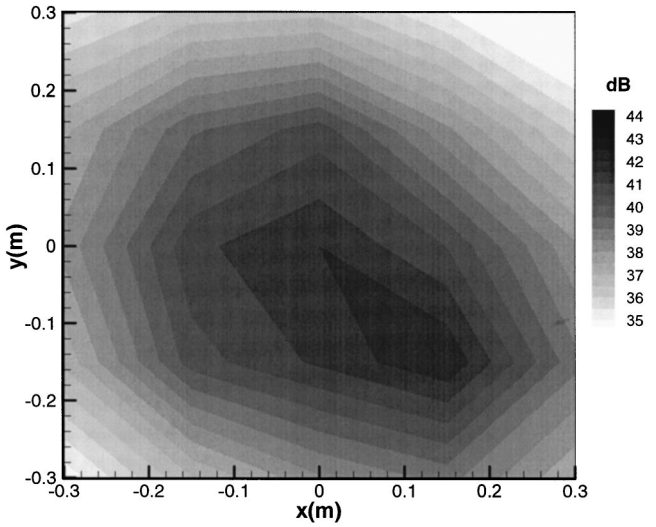


FIG. 1. The example of the auto-spectra of pressure at a selected frequency on a hologram plane when there are two arbitrary sources.

$$s_{uv}(f)=\frac{1}{TN}\sum_{i=1}^{N} u^{(i)*}(f)v^{(i)}(f), \tag{1c}$$

where the superscript ''*'' denotes the complex conjugate or the matrix conjugate transpose. We will omit the frequency variable ''f'' in later expressions because we are interested in only frequency domain.

We define spectral vectors[12] as

$$\mathbf{u}=\frac{1}{TN}[u^{(1)}\cdots u^{(N)}]^T, \tag{2a}$$

$$\mathbf{v}=\frac{1}{TN}[v^{(1)}\cdots v^{(N)}]^T, \tag{2b}$$

where the superscript ''T'' denotes the matrix transpose. The vectors enable one to write the spectra of Eqs. (1a)–(1c) as

$$s_{uu}=\mathbf{u}^*\mathbf{u}, \tag{3a}$$

$$s_{vv}=\mathbf{v}^*\mathbf{v}, \tag{3b}$$

$$s_{uv}=\mathbf{u}^*\mathbf{v}, \tag{3c}$$

Equations (3a)–(3c) express the spectra as the inner products of the two spectral vectors.

Using Eqs. (3a)–(3c), one can express the coherence function as

$$\gamma_{uv}^2=\frac{|s_{uv}|^2}{s_{uu}s_{vv}}=\frac{|\mathbf{u}^*\mathbf{v}|^2}{(\mathbf{u}^*\mathbf{u})(\mathbf{v}^*\mathbf{v})}. \tag{4}$$

In order to investigate the graphical meaning of Eq. (4), let us assume that **u** and **v** are defined in Euclidean space with real elements, although they are defined in Unitary space with complex ones. If the angle between them is θ, the coherence function is $\cos^2\theta$ (Fig. 2). That is, if the projection vector of **u** on **v** is $\mathbf{u}_{:v}$, the coherence function is the square of the magnitude ratio of $\mathbf{u}_{:v}$ to **u**. Therefore, if it is one (coherent case), one vector is linearly dependent on the other. If it is zero (incoherent case), one vector is orthogonal to the other. That is, a coherence function represents the orthogonal relation between two spectral vectors.

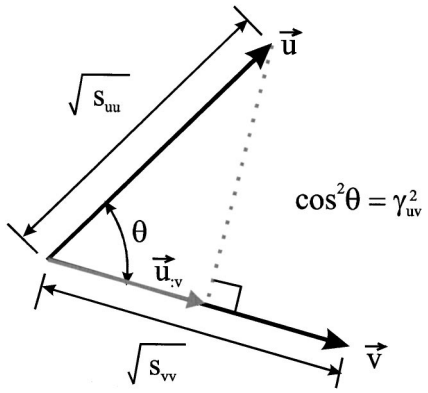


FIG. 2. The visualization of a coherence function (**u** and **v**: the spectral vectors of signal u and v, $\mathbf{u}_{:v}$: the projection vector of **u** on **v**, s_{uu} and s_{vv} : the auto-spectra of u and v, γ_{uv}^2: the coherence function between u and v, and θ: the angle between **u** and **v**).

As an example using spectral vectors, let us consider a single-input and single-output system, where u is the input, v is the output, and h_u^v is the transfer function. If the record length T is large enough, one can write all the samples as

$$v^{(i)}=h_u^v u^{(i)}, \quad i=1,\cdots,N. \tag{5}$$

Therefore one can write a vector equation as

$$\mathbf{v}=h_u^v\mathbf{u}. \tag{6}$$

This equation means that the input vector is linearly dependent on the output vector. Inserting Eq. (6) to Eq. (4) shows the well-known fact that the coherence function between the input and the output is one in a single-input and single-output system.[8]

B. Problem definition

Figure 3 illustrates a planar acoustic holography when there are two sources. q_1 and q_2 are input signals at a selected frequency. r_1 and r_2 are reference signals. $p_H(x,y)$ is pressure at (x,y) point on a hologram plane, and $p_S(x',y')$ is pressure at (x',y') point on a source plane.

If $h_{qi}^H(x,y)$ is the transfer function between q_i and $p_H(x,y)$, one can express $p_H(x,y)$ and $p_S(x',y')$ as

$$p_H(x,y)=h_{q1}^H(x,y)q_1+h_{q2}^H(x,y)q_2, \tag{7a}$$

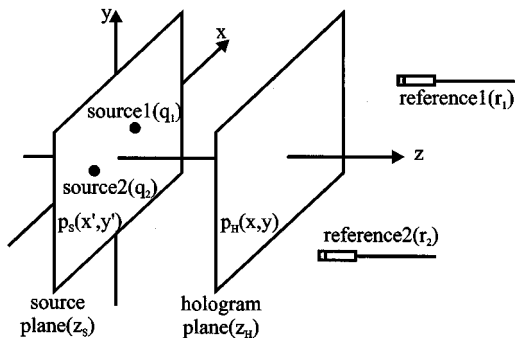


FIG. 3. The illustration of a planar acoustic holography [q_1 and q_2: two input signals, r_1 and r_2: two reference signals, $p_H(x,y)$ and $p_S(x',y')$: pressure on a hologram and a source plane].

$$p_S(x',y') = L_H^S[p_H(x,y)]$$
$$= h_{q1}^S(x',y')q_1 + h_{q2}^S(x',y')q_2, \quad (7b)$$

where

$$h_{qi}^S(x',y') = L_H^S[h_{q1}^H(x,y)]. \quad (7c)$$

L_H^S is the linear operator that propagates pressure from the hologram to the source plane.[1]

We restrict our analysis only to a stationary random sound field. Then we can define spectral vectors $\mathbf{q}_1$, $\mathbf{q}_2$, $\mathbf{p}_H(x,y)$, and $\mathbf{p}_S(x',y')$, as defined in the Sec. II A. Thus Eqs. (7a) and (7b), like Eq. (6), can be rewritten as

$$\mathbf{p}_H(x,y) = h_{q1}^H(x,y)\mathbf{q}_1 + h_{q2}^H(x,y)\mathbf{q}_2 = QH_q^H(x,y), \quad (8a)$$

$$\mathbf{p}_S(x',y') = h_{q1}^S(x',y')\mathbf{q}_1 + h_{q2}^S(x',y')\mathbf{q}_2$$
$$= QH_q^S(x',y'), \quad (8b)$$

where

$$Q = [\mathbf{q}_1 \quad \mathbf{q}_2], \quad (8c)$$

$$H_q^H(x,y) = [h_{q1}^H(x,y) \quad h_{q2}^H(x,y)]^T, \quad (8d)$$

$$H_q^S(x',y') = [h_{q1}^S(x',y') \quad h_{q2}^S(x',y')]^T. \quad (8e)$$

It is noteworthy that $\mathbf{p}_H(x,y)$ and $\mathbf{p}_S(x',y')$ lie in the two-dimensional subspace spanned by $\mathbf{q}_1$ and $\mathbf{q}_2$, although the number of the vector elements is N. That is, input vectors limit the degree of freedom of output vectors.

Our incoherent assumption means that $\mathbf{q}_1$ is orthogonal to $\mathbf{q}_2$. Therefore the inner products of Eqs. (8a) and (8b) produce the auto-spectra of the pressure on the hologram and the source plane, $s_{HH}(x,y)$ and $s_{SS}(x',y')$, as

$$s_{HH}(x,y) = |h_{q1}^H(x,y)\sqrt{s_{q1q1}}|^2 + |h_{q2}^H(x,y)\sqrt{s_{q2q2}}|^2$$
$$= H_q^{H*}(x,y)S_{qq}H_q^H(x,y), \quad (9a)$$

$$s_{SS}(x',y') = |h_{q1}^S(x',y')\sqrt{s_{q1q1}}|^2 + |h_{q2}^S(x',y')\sqrt{s_{q2q2}}|^2$$
$$= H_q^{S*}(x',y')S_{qq}H_q^S(x',y'), \quad (9b)$$

where

$$s_{HH}(x,y) = \mathbf{p}_H^*(x,y)\mathbf{p}_H(x,y), \quad (9c)$$

$$s_{SS}(x',y') = \mathbf{p}_S^*(x',y')\mathbf{p}_S(x',y'), \quad (9d)$$

$$s_{q1q1} = \mathbf{q}_1^*\mathbf{q}_1, \quad (9e)$$

$$s_{q2q2} = \mathbf{q}_2^*\mathbf{q}_2, \quad (9f)$$

$$S_{qq} = Q^*Q = \begin{bmatrix} \mathbf{q}_1^*\mathbf{q}_1 & \mathbf{q}_1^*\mathbf{q}_2 \\ \mathbf{q}_2^*\mathbf{q}_1 & \mathbf{q}_2^*\mathbf{q}_2 \end{bmatrix} = \begin{bmatrix} s_{q1q1} & 0 \\ 0 & s_{q2q2} \end{bmatrix}. \quad (9g)$$

The first and the second terms of Eqs. (9a) and (9b) represent the contributions of the first and the second source to the spectra.

By measuring the inputs and the pressure on the hologram plane, one can estimate $h_{qi}^H(x,y)$ and s_{qiqi}.[8] Therefore one can estimate the contributions on the hologram plane. By propagating $h_{qi}^H(x,y)$ [Eq. (7c)], one can estimate $h_{qi}^S(x',y')$. Therefore one can also estimate the contributions on the source plane. This process is identical to estimating the magnitude squares of the projection vectors of $\mathbf{p}_H(x,y)$ and $\mathbf{p}_S(x',y')$ on $\mathbf{q}_1$ and $\mathbf{q}_2$, as shown in Fig. 4. This also means that one can estimate the contributions by obtaining signals coherent to the inputs, of which the vectors are linearly dependent on the input vectors. However, in many practical cases, it is difficult to directly obtain inputs or signals coherent to the inputs.

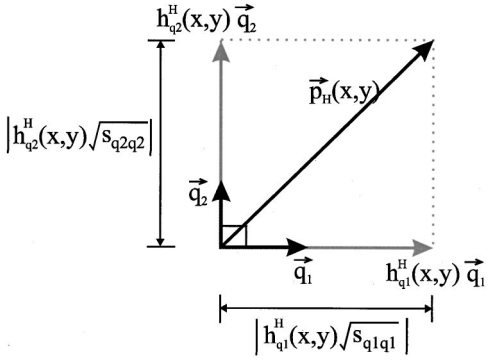


FIG. 4. The expression of a pressure vector by input vectors [$\mathbf{q}1$ and $\mathbf{q}_2$: two input vectors, $\mathbf{p}_H(x,y)$: pressure vector on the hologram plane, and $h_{qi}^H(x,y)$: the transfer function between q_i and $p_H(x,y)$].

Section III proposes a method to obtain such signals by using the backward prediction of NAH. We formulate the proposed method on the basis of a step-by-step measurement using reference signals. Before Sec. III, therefore, we describe NAH based on reference signals in terms of the vector notation.

C. NAH based on reference signals

Because the reference signals in Fig. 3 are also outputs, the reference vectors, $\mathbf{r}_1$ and $\mathbf{r}_2$, lie in the two-dimensional vector space spanned by the input vectors. Thus the reference vectors can be used as the basis to express other output vectors if they are linearly independent. In most of the cases, their linear independence can be achieved by measuring them at different points. Using the reference vectors, the pressure vectors can be written as

$$\mathbf{p}_H(x,y) = h_{r1}^H(x,y)\mathbf{r}_1 + h_{r2}^H(x,y)\mathbf{r}_2 = RH_r^H(x,y), \quad (10a)$$

$$\mathbf{p}_S(x',y') = h_{r1}^S(x',y')\mathbf{r}_1 + h_{r2}^S(x',y')\mathbf{r}_2 = RH_r^S(x',y'), \quad (10b)$$

where

$$h_{ri}^S(x',y') = L_H^S[h_{ri}^H(x,y)], \quad (10c)$$

$$R = [\mathbf{r}_1 \quad \mathbf{r}_2], \quad (10d)$$

$$H_r^H(x,y) = [h_{r1}^H(x,y) \quad h_{r2}^H(x,y)]^T, \quad (10e)$$

$$H_r^S(x',y') = [h_{r1}^S(x',y') \quad h_{r2}^S(x',y')]^T. \quad (10f)$$

$h_{ri}^H(x,y)$ is the coefficient of $\mathbf{r}_i$ when $\mathbf{p}_H(x,y)$ is expressed by the linear combination of $\mathbf{r}_1$ and $\mathbf{r}_2$. We call it the transfer function between r_i and $p_H(x,y)$. The Appendix explains a method to estimate it. In Eq. (10b), $h_{ri}^S(x',y')$ can be estimated by propagating $h_{ri}^H(x,y)$ [Eq. (10c)]. Figure 5 visualizes Eq. (10a).

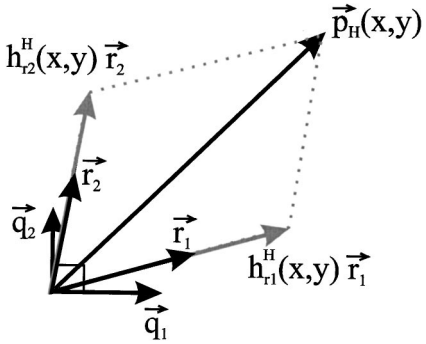


FIG. 5. The expression of a pressure vector by reference vectors [$\mathbf{q}_1$ and $\mathbf{q}_2$: two input vectors, $\mathbf{r}_1$ and $\mathbf{r}_2$: two reference vectors, $\mathbf{p}_H(x,y)$: pressure vector on the hologram plane, and $h^H_{r1}(x,y)$: the transfer function between r_i and $p_H(x,y)$].

The underlying assumption of Eqs. (10a) and (10b) is that the number of independent sources is two. In general, however, one has to estimate it from a prior experiment. There are large publications available on this topic. References 15–17 are some of them. However, because this topic is not our interest, we assume that one can estimate it.

By the inner products of Eqs. (10a) and (10b), one can estimate the auto-spectra on the hologram and the source plane as

$$s_{HH}(x,y)=H_r^{H*}(x,y)S_{rr}H_r^H(x,y), \tag{11a}$$

$$s_{SS}(x',y')=H_r^{S*}(x',y')S_{rr}H_r^S(x',y'), \tag{11b}$$

where

$$S_{rr}=R^*R=\begin{bmatrix}\mathbf{r}_1^*\mathbf{r}_1 & \mathbf{r}_1^*\mathbf{r}_2\\ \mathbf{r}_2^*\mathbf{r}_1 & \mathbf{r}_2^*\mathbf{r}_2\end{bmatrix}=\begin{bmatrix}s_{r1r1} & s_{r1r2}\\ s_{r2r1} & s_{r2r2}\end{bmatrix}. \tag{11c}$$

However, we cannot directly estimate the contributions of the sources from Eqs. (11a) and (11b). This is because the reference vectors are not linearly dependent on the input vectors in general, as shown in Fig. 5.

III. SOLUTION METHOD

A. The idea of the proposed method

Equation (11b) enables one to estimate the auto-spectra of pressure on the source plane. Thus one can estimate source positions from the local maximum points of the spectra and pressure signals at the source positions. For contribution analysis, our proposed method uses the pressure signals at the source positions estimated by NAH. It is conceptually identical to the conventional method addressed in Sec. I. However, it does not require any prior information on source positions before applying NAH because of using the estimated signals, not the measured signals.

Let us consider an ideal case, where the two source positions are (x_1',y_1') and (x_2',y_2') on the source plane. Figure 6 is the one-dimensional expression of the ideal case. The contributions have their maximum values at the source positions. The shapes are steep enough to neglect the overlap of the contributions, or the distance between the sources is large enough. Then the pressure at the source positions, $p_S(x_1',y_1')$ and $p_S(x_2',y_2')$, are coherent to the inputs because there is no overlap. Thus one can estimate the contributions by simply projecting $\mathbf{p}_H(x,y)$ and $\mathbf{p}_S(x',y')$ on $\mathbf{p}_S(x_1',y_1')$ and $\mathbf{p}_S(x_2',y_2')$. As mentioned above, one can estimate the source positions by Eq. (11b) and the pressure vectors at the source positions by Eq. (10b). $\mathbf{p}_S(x_1',y_1')$ and $\mathbf{p}_S(x_2',y_2')$ are expressed by the linear combination of the reference vectors, but are linearly dependent on the input vectors, as shown in Fig. 7.

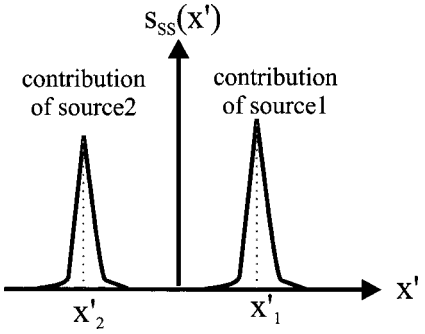


FIG. 6. The one-dimensional expression of an ideal case for the proposed method [x_1' and x_2': two source positions, and $s_{SS}(x')$: the auto-spectra of pressure on the source plane].

For the simplicity of notation, we rewrite $\mathbf{p}_S(x_1',y_1')$ and $\mathbf{p}_S(x_2',y_2')$ as $\mathbf{w}_1$ and $\mathbf{w}_2$. From Eq. (10b), $\mathbf{w}_1$ and $\mathbf{w}_2$ are expressed as

$$\mathbf{w}_1=\mathbf{p}_S(x_1',y_1')=h_{r1}^{w1}\mathbf{r}_1+h_{r2}^{w1}\mathbf{r}_2=RH_r^{w1}, \tag{12a}$$

$$\mathbf{w}_2=\mathbf{p}_S(x_2',y_2')=h_{r1}^{w2}\mathbf{r}_1+h_{r2}^{w2}\mathbf{r}_2=RH_r^{w2}, \tag{12b}$$

where

$$h_{ri}^{wj}=h_{ri}^S(x_j,y_j), \tag{12c}$$

$$H_r^{w1}=[h_{r1}^{w1}\quad h_{r2}^{w1}]^T, \tag{12d}$$

$$H_r^{w2}=[h_{r1}^{w2}\quad h_{r2}^{w2}]^T. \tag{12e}$$

Using Eqs. (10a), (11c), (12a), and (12b), one can express the projection vectors of $\mathbf{p}_H(x,y)$ on $\mathbf{w}_1$ and $\mathbf{w}_2$, $\mathbf{p}_{H:w1}(x,y)$ and $\mathbf{p}_{H:w2}(x,y)$, as[18,19]

$$\mathbf{p}_{H:w1}(x,y)=\frac{\mathbf{w}_1^*\mathbf{p}_H(x,y)}{\mathbf{w}_1^*\mathbf{w}_1}\mathbf{w}_1=\frac{H_r^{w1*}S_{rr}H_r^H(x,y)}{s_{w1w1}}\mathbf{w}_1, \tag{13a}$$

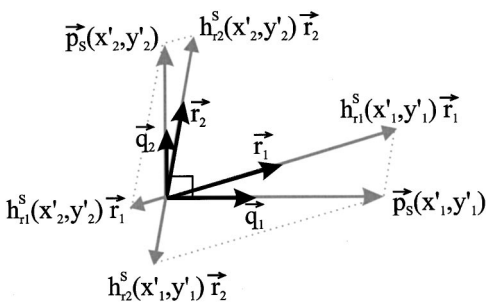


FIG. 7. The expression of pressure vectors at source positions by reference vectors in the ideal case [$\mathbf{q}_1$ and $\mathbf{q}_2$: two input vectors, $\mathbf{r}_1$ and $\mathbf{r}_2$: two reference vectors, $\mathbf{p}_S(x_1',y_1')$ and $\mathbf{p}_S(x_2',y_2')$: pressure vectors at the source positions, and $h_{ri}^S(x_j',y_j')$: the transfer function between r_i and $p_S(x_j',y_j')$].

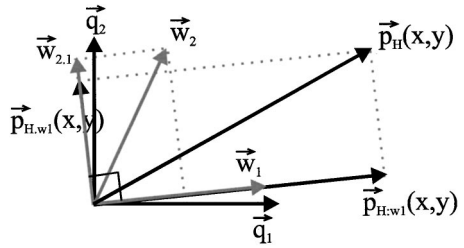


FIG. 8. The vector expression of the partial coherence procedure [$\mathbf{q}_1$ and $\mathbf{q}_2$: two input vectors, $\mathbf{w}_1$ and $\mathbf{w}_2$: pressure vectors at the source positions, $\mathbf{w}_{2\cdot1}$: the orthogonal component of $\mathbf{w}_2$ to $\mathbf{w}_1$, $\mathbf{p}_H(x,y)$: pressure vector on the hologram plane, $\mathbf{p}_{H:w1}(x,y)$ and $\mathbf{p}_{H\cdot w1}(x,y)$: the projection vectors of $\mathbf{p}_H(x,y)$ on $\mathbf{w}_1$ and $\mathbf{w}_{2\cdot1}$].

$$\mathbf{p}_{H:w2}(x,y)=\frac{\mathbf{w}_2^*\mathbf{p}_H(x,y)}{\mathbf{w}_2^*\mathbf{w}_2}\mathbf{w}_2=\frac{H_r^{w2*}S_{rr}H_r^H(x,y)}{s_{w2w2}}\mathbf{w}_2, \quad (13b)$$

where

$$s_{w1w1}=\mathbf{w}_1^*\mathbf{w}_1, \quad (13c)$$

$$s_{w2w2}=\mathbf{w}_2^*\mathbf{w}_2. \quad (13d)$$

s_{w1w1} and s_{w2w2} are the auto-spectra at the source positions. By the magnitude squares of Eqs. (13a) and (13b), one can estimate the contributions on the hologram plane. They are $|H_r^{w1*}S_{rr}H_r^H(x,y)|^2/s_{w1w1}$ and $|H_r^{w2*}S_{rr}H_r^H(x,y)|^2/s_{w2w2}$. Similarly, one can also estimate the contributions on the source plane.

As the spatial overlap becomes large, the proposed method will produce large error. This is because w_1 and w_2 are not perfectly coherent to q_1 and q_2. Therefore it can be applied to local sources with small overlap, but not to sources with large overlap. This is its limitation.

B. The proposed method combined with the partial coherence procedure

When there is the overlap, the partial coherence procedure[8,20] can improve the second contribution estimation in some cases. In terms of the vector notation, it is equivalent to Gram–Schmidt orthogonalization,[18,19] as shown in Ref. 12.

Figure 8 illustrates the proposed method combined with the partial coherence procedure. $\mathbf{w}_1$ and $\mathbf{w}_2$ are the pressure vectors at the source positions. The partial coherence procedure obtains new orthogonal basis $\mathbf{w}_1$ and $\mathbf{w}_{2\cdot1}$ by applying Gram–Schmidt orthogonalization to $\mathbf{w}_1$ and $\mathbf{w}_2$, where $\mathbf{w}_{2\cdot1}$ is the orthogonal component of $\mathbf{w}_2$ to $\mathbf{w}_1$. The partial coherence procedure estimates the contributions by projecting $\mathbf{p}_H(x,y)$ on $\mathbf{w}_1$ and $\mathbf{w}_{2\cdot1}$. The projection vector on $\mathbf{w}_1$ is $\mathbf{p}_{H:w1}(x,y)$ [Eq. (13a)]. Therefore it does not change the first contribution. However, it changes the second contribution because $\mathbf{w}_{2\cdot1}$ is different than $\mathbf{w}_2$. As shown in Fig. 8, the projection vector on $\mathbf{w}_{2\cdot1}$, $\mathbf{p}_{H\cdot w1}(x,y)$, can be obtained by subtracting $\mathbf{p}_{H:w1}(x,y)$ from $\mathbf{p}_H(x,y)$. Using Eqs. (10a), (12a), and (13a), the result is expressed as

$$\mathbf{p}_{H\cdot w1}(x,y)=\mathbf{p}_H(x,y)-\mathbf{p}_{H:w1}(x,y)=RH_{r\cdot w1}^H(x,y), \quad (14a)$$

where

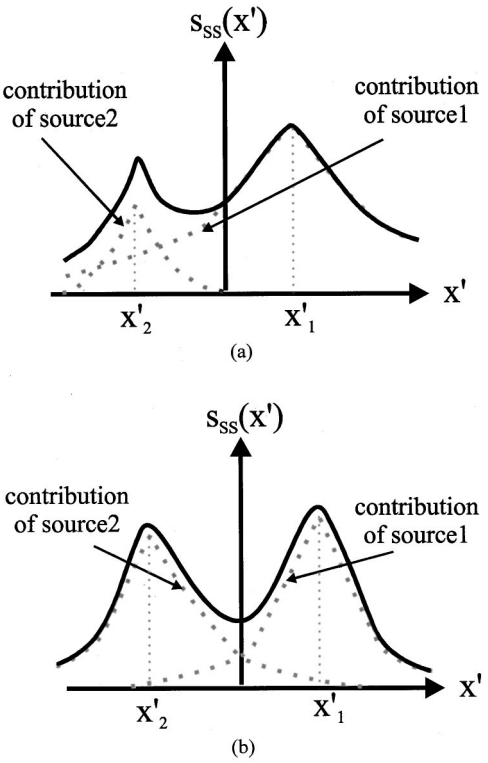


FIG. 9. The two examples of spatial overlaps: (a) the case that the strength of one source is much larger than that of the other; (b) the case that the strength of one source is similar to that of the other [x_1' and x_2': two source positions, and $s_{SS}(x')$: the auto-spectra of pressure on the source plane].

$$H_{r\cdot w1}^H(x,y)=H_r^H(x,y)-H_r^{w1}(H_r^{w1*}S_{rr}H_r^H(x,y))/s_{w1w1}. \quad (14b)$$

The partial coherence procedure estimates the second contribution by the magnitude square of Eq. (14a), $H_{r\cdot w1}^{H*}(x,y)S_{rr}H_{r\cdot w1}^H(x,y)$.

The partial coherence procedure makes the second contribution estimation more accurate in the case of Fig. 9(a). In Fig. 9(a), the first source significantly affects the pressure at the second source position, but the second source does not affect the pressure at the first source position. Thus $\mathbf{w}_1$ is very linearly dependent on $\mathbf{q}_1$, but $\mathbf{w}_2$ is not linearly dependent on $\mathbf{q}_2$, as shown in Fig. 8. However, $\mathbf{w}_{2\cdot1}$ is very linearly dependent on $\mathbf{q}_2$ because of the orthogonal property. Therefore the second contribution estimation becomes more accurate by the partial coherence procedure. However, it does not in the case of Fig. 9(b) that $\mathbf{w}_1$ and $\mathbf{w}_2$ are similarly distorted.

C. Numerical simulation

In order to verify the proposed method, we performed a numerical simulation using two incoherent monopole sources. The hologram plane was located at 15 cm (z_H), and the source plane was at 0 cm. Measurement spacing was 15 cm, and the number of measurement points was 32×32. The two monopole sources were located at (15 cm, 0, 0) and (−15 cm, 0, 0). Reference microphones were located at (30 cm, 0, 0) and (−30 cm, 0, 0), which were considered to be sufficiently far from the sources. Frequency was 300 Hz. The input signals were the strengths of the monopole sources.

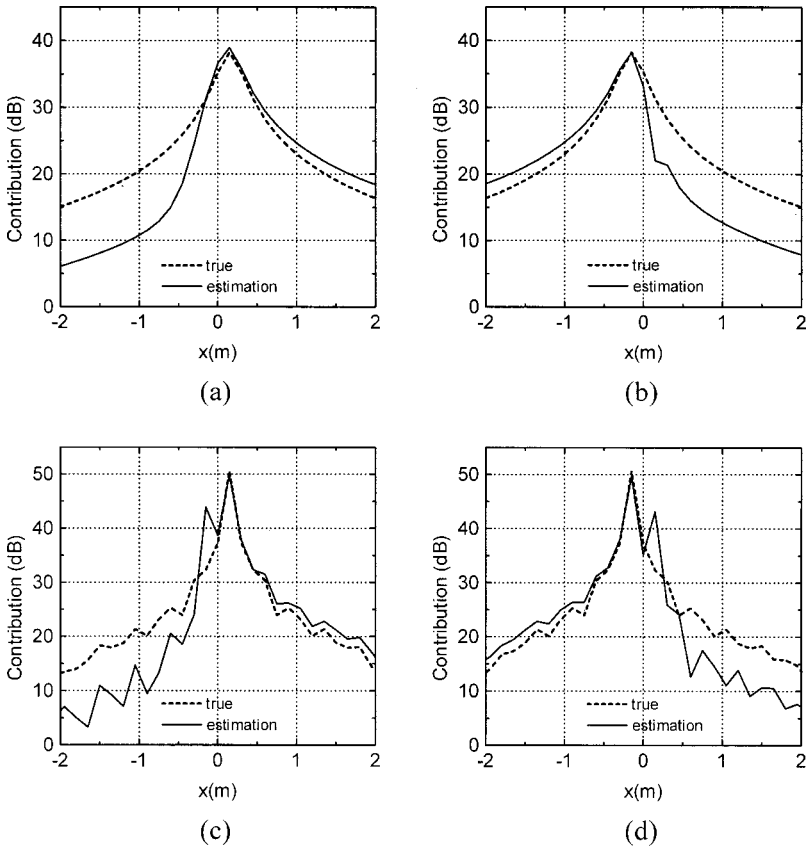


FIG. 10. The results of Estimation 1 at $y=0$ line: (a) the first contribution on the hologram plane; (b) the second contribution on the hologram plane; (c) the first contribution on the source plane; (d) the second contribution on the source plane.

The auto-spectra of the inputs were the same. The cross-spectrum between the inputs was zero. This assured our incoherent assumption. In order to estimate the pressure on the source plane, we added zeros so that total data number is 96×96, and used the window[21] that assures minimum error.

We estimated the contributions by two methods, Estimation 1 and 2. Estimation 1 estimated the contributions by simply applying the partial coherence procedure to the reference signals. It is the conventional method except for not placing the reference microphones close to sources. Estimation 2 estimated them by the proposed method combined with the partial coherence procedure.

Figures 10 and 11 show the results of Estimation 1 and 2 at $y=0$ line. The true contributions on the source plane were estimated by propagating the true ones on the hologram plane. Figure 10 shows large error. This is simply because the reference signals are not coherent to the inputs. Figure 11 shows relatively small error. However, the results are not perfect because of the overlap on the source plane. We can see that the second contribution at the first source position is $-\infty$ dB [Fig. 11(d)]. This is because the orthogonal component of $\mathbf{w}_1$ to $\mathbf{w}_1$ is zero. In other words, that is because the proposed method combined with the partial coherence procedure regards that only the first source affects the pressure at the first source position.

D. Optimal decomposition plane

The overlap produces error. This section introduces a method to reduce it.

The overlap on the source plane will be less than that on the hologram plane. The overlap on the plane between the hologram and the source plane will decrease, as the plane goes to the source plane. Let us consider a decomposition plane (z_D) for contribution analysis. That is, we use the local maximum pressure vectors on the decomposition plane instead of the source plane. As the decomposition plane goes to the source plane, the estimated contributions will go to the true ones because of the reduction of the overlap. This motivated us to send the decomposition plane to the back of the source plane. The pressure on the decomposition plane behind the source plane ($z_D<z_S$) cannot be accomplished in reality, but can obviously produce more coherent signals.

Figure 12 shows the first contributions on the hologram plane estimated by using three different decomposition planes. The result by the $z_D=-5$ cm plane is better than that by the $z_D=0$ cm plane, which is the source plane. It implies that a decomposition plane, which is located at the back of the source plane, can be better than the source plane. However, the result by the $z_D=-40$ cm are worse than that by the $z_D=-5$ cm. It implies that there is an optimal position of the decomposition plane.

In order to understand these facts, we examined the first true contributions in space and wave number domain on the three decomposition planes (Fig. 13). In the wave number domain [Fig. 13(a)], the true contribution on the $z_D=-5$ cm plane has the flattest shape because evanescent wave components properly increase. Therefore the true contribution on the $z_D=-5$ cm plane has the steepest shape in the space domain [Fig. 13(b)]. It means that the $z_D=-5$ cm plane has the least overlap and is closest to the ideal case.

The next step is to determine the optimal plane of many

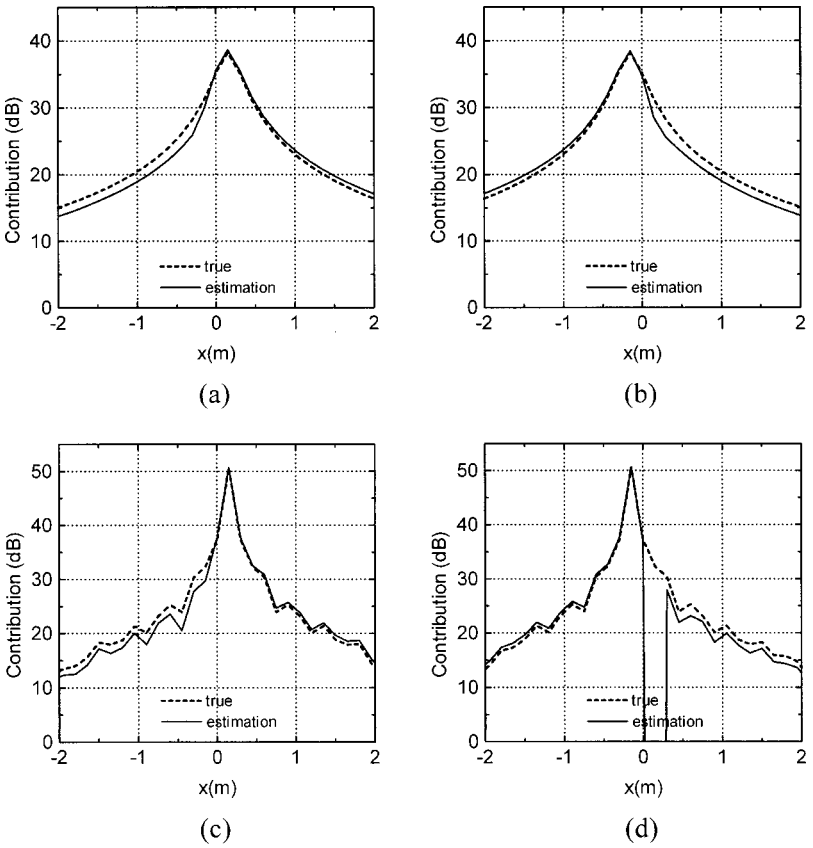


FIG. 11. The results of Estimation 2 at $y=0$ line: (a) the first contribution on the hologram plane; (b) the second contribution on the hologram plane; (c) the first contribution on the source plane; (d) the second contribution on the source plane.

possible planes. The optimal plane must have the least overlap. This means that the contribution on the optimal plane has to be steepest. Therefore, in order to find it, we have to devise a measure for the steepness.

If a decomposition plane is not far from the optimal plane, the auto-spectrum near a maximum point will be similar to the first contribution. Therefore, as a measure for the steepness, one can use the ratio of the maximum auto-

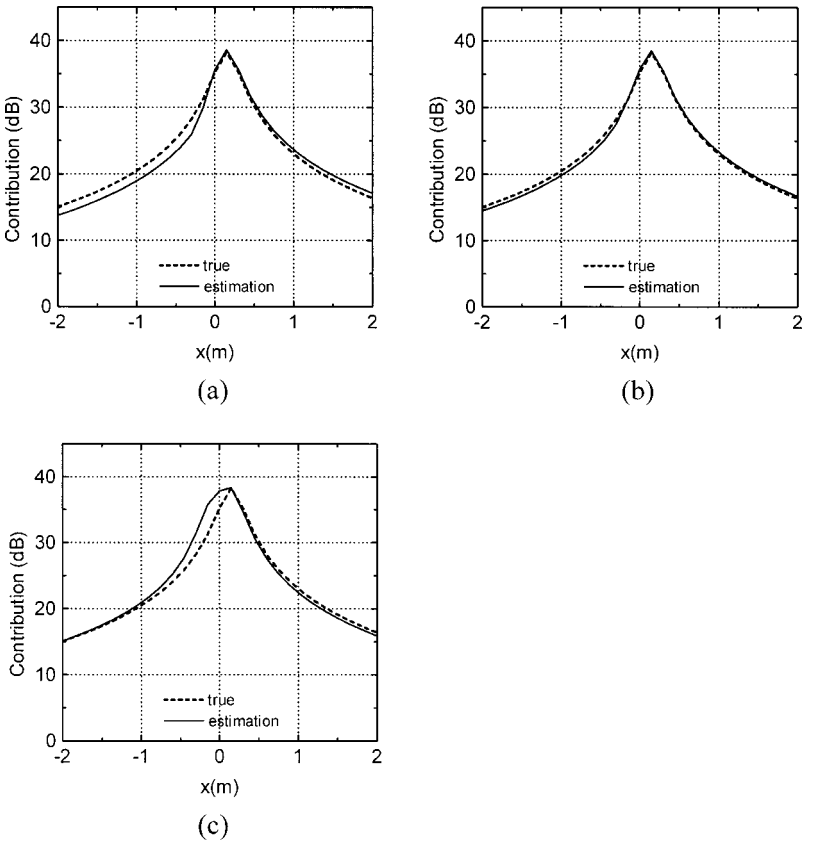


FIG. 12. The first contributions at $y=0$ line on the hologram plane estimated by three different decomposition planes, which are (a) $z_D=0$ cm, (b) $z_D=-5$ cm, and (c) $z_D=-40$ cm.

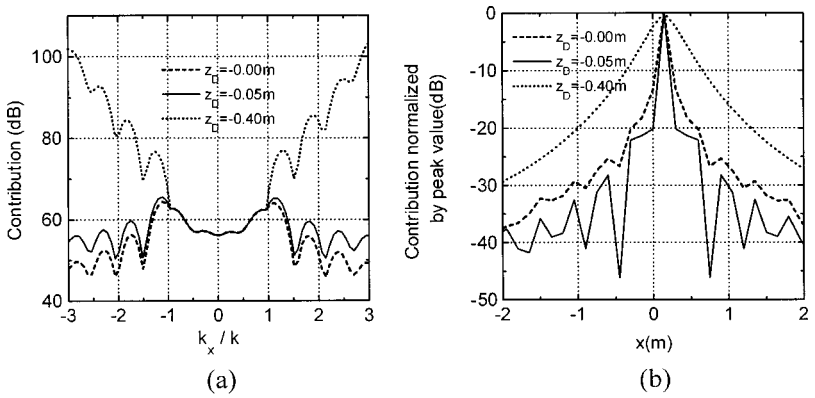


FIG. 13. The first true contributions on three decomposition planes: (a) at $k_y=0$ line (y-directional wave number) in a wave number domain (k: specific wave number, k_x : x-directional wave number); (b) at $y=0$ line in a space domain, where the contributions are normalized by their maximum values.

spectrum to the neighborhood. If the auto-spectrum on a decomposition plane is $s_{DD}(m,n)$ in a discrete form and the maximum point is $(m_{\max},n_{\max})$, then one can express the ratio, $\kappa(z_D)$, as

$$\kappa(z_D)=4s_{DD}(m_{\max},n_{\max})/\{s_{DD}(m_{\max}+1,n_{\max}) + s_{DD}(m_{\max}-1,n_{\max})+s_{DD}(m_{\max},n_{\max}+1) + s_{DD}(m_{\max},n_{\max}-1)\}. \quad (15)$$

Therefore one can determine the plane to maximize this ratio as the optimal plane. Using this ratio, we determined that the optimal plane is $z_D=-5$ cm of $z_D=0$, -1 cm, ..., -40 cm in the above simulation. Figure 14 shows the results by this optimal plane. We can see better results than those of Fig. 11.

IV. CONCLUSIONS

We proposed a method to estimate the contributions of multiple incoherent sources in near-field acoustic holography (NAH). The proposed method predicts source positions by NAH, and uses the estimated pressure at the source positions. It does not require any prior information on source positions before applying NAH unlike the conventional method. The detail procedure was explained by the projection of the spectral vectors. We verified the proposed method by the numerical simulation using two incoherent monopole sources. We also introduced an optimal decomposition plane in order to reduce the spatial overlap, which produces the error in the proposed method.

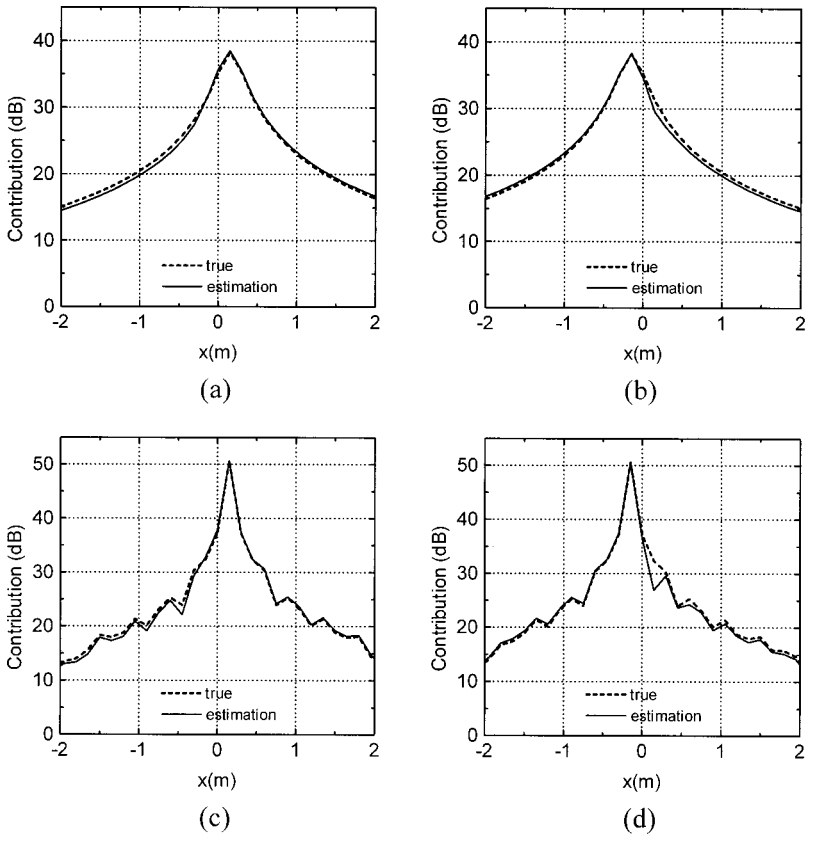


FIG. 14. The contributions at $y=0$ line by the optimal decomposition plane ($z_D=-5$ cm): (a) the first contribution on the hologram plane; (b) the second contribution on the hologram plane; (c) the first contribution on the source plane; (d) the second contribution on the source plane.

ACKNOWLEDGMENTS

This study was partly supported by the NRL (National Research Laboratory) project of KISTEP (Korea Institute of Science & Technology Evaluation and Planning) and the BK21 (Brain Korea 21) project initiated by the Ministry of Education & Human Resources & Development of Korea.

APPENDIX: THE ESTIMATION OF THE TRANSFER FUNCTIONS BETWEEN REFERENCES AND THE PRESSURE

Let us rewrite Eq. (10a) as

$$\mathbf{p}_H(x,y) = RH_r^H(x,y), \tag{A1a}$$

where

$$R = [\mathbf{r}_1 \quad \mathbf{r}_2], \tag{A1b}$$

$$H_r^H(x,y) = [h_{r1}^H(x,y) \quad h_{r2}^H(x,y)]^T. \tag{A1c}$$

In Eq. (A1a), the unknowns are $h_{r1}^H(x,y)$ and $h_{r2}^H(x,y)$. By the pseudo-inverse of matrix R, they can be estimated as[8,12]

$$\begin{aligned} H_r^H(x,y) &= [h_{r1}^H(x,y) \quad h_{r2}^H(x,y)]^T \\ &= (R^*R)^{-1}(R^*\mathbf{p}_H(x,y)) = S_{rr}^{-1}S_{rH}(x,y), \end{aligned} \tag{A2a}$$

where

$$S_{rr} = R^*R = \begin{bmatrix} \mathbf{r}_1^*\mathbf{r}_1 & \mathbf{r}_1^*\mathbf{r}_2 \\ \mathbf{r}_2^*\mathbf{r}_1 & \mathbf{r}_2^*\mathbf{r}_2 \end{bmatrix} = \begin{bmatrix} s_{r1r1} & s_{r1r2} \\ s_{r2r1} & s_{r2r2} \end{bmatrix}, \tag{A2b}$$

$$S_{rH}(x,y) = R^*\mathbf{p}_H(x,y) = \begin{bmatrix} \mathbf{r}_1^*\mathbf{p}_H(x,y) \\ \mathbf{r}_2^*\mathbf{p}_H(x,y) \end{bmatrix} = \begin{bmatrix} s_{r1H}(x,y) \\ s_{r2H}(x,y) \end{bmatrix}. \tag{A2c}$$

[1] E. G. Williams and J. D. Maynard, ''Holographic imaging without the wavelength resolution limit,'' Phys. Rev. Lett. **45**, 554–557 (1980).
[2] J. Hald, ''STSF—A unique technique for scan-based near-field acoustic holography without restrictions on coherence,'' B&K Technical Review No. 1 (1989).
[3] K. B. Ginn and J. Hald, ''STSF—Practical instrumentation and applications,'' B&K Technical Review No. 2 (1989).
[4] D. Hallman and J. S. Bolton, ''Multi-reference near-field acoustical holography,'' Proc. of Inter-Noise '92, 1165–1170 (1992).
[5] H.-S. Kwon and J. S. Bolton, ''Partial field decomposition in nearfield acoustical holography by the use of singular value decomposition and partial coherence procedures,'' Proc. of Noise-Con '98, 649–654 (1998).
[6] R. J. Ruhala and C. B. Burroughs, ''Separation of leading edge, trailing edge, and sidewall noise sources from rolling tires,'' Proc. of Noise-Con '98, 109–114 (1998).
[7] M. A. Tomlinson, ''Partial source discrimination in near field acoustic holography,'' Appl. Acoust. **57**, 243–261 (1999).
[8] J. S. Bendat and A. G. Piersol, *Random Data: Analysis and Measurement Procedures*, 2nd ed. (Wiley, New York, 1986), pp. 109–251.
[9] R. J. Alfredson, ''The partial coherence technique for source identification on a diesel engine,'' J. Sound Vib. **55**, 487–494 (1977).
[10] M. W. Trethewey and H. A. Evensen, ''Identification of noise sources of forge hammers during production: An application of residual spectrum techniques to transients,'' J. Sound Vib. **77**, 357–374 (1981).
[11] M. E. Wang and M. J. Crocker, ''On the application of coherence techniques for source identification in a multiple noise source environment,'' J. Acoust. Soc. Am. **74**, 861–872 (1983).
[12] J. M. Danthez and R. Aquilina, ''Separation of broadband sources processing concept of the labrador software,'' Mech. Syst. Signal Process. **11**, 91–106 (1997).
[13] D. Otte, P. V. Ponseele, and J. Leuridan, ''Operational deflection shapes in multisource environments,'' Proc. of the 8th International Modal Analysis Conference, 413–421 (1990).
[14] S. H. Yoon and P. A. Nelson, ''A method for the efficient construction of acoustic pressure cross-spectral matrices,'' J. Sound Vib. **233**, 897–920 (2000).
[15] R. O. Schmidt, ''Multiple emitter location and signal parameter estimation,'' IEEE Trans. Antennas Propag. **34**, 276–280 (1986).
[16] S. M. Price and R. J. Bernhard, ''Virtual coherence: A digital signal processing techniques for incoherent source identification,'' Proc. of the 4th International Modal Analysis Conference, 1256–1262 (1986).
[17] M. S. Kompella, P. Davis, R. J. Bernhard, and D. A. Ufford, ''A technique to determine the number of incoherent sources contributing to the response of a system,'' Mech. Syst. Signal Process. **8**, 363–380 (1994).
[18] G. Strang, *Linear Algebra and its Applications*, 3rd ed. (Harcourt Brace, New York, 1988), pp. 63–182.
[19] P. Lancaster and M. Tismenetsky, *The Theory of Matrices with Application*, 2nd ed. (Academic, New York, 1985), pp. 1–219.
[20] C. J. Dodds and J. D. Robson, ''Partial coherence in multivariate random process,'' J. Sound Vib. **42**, 243–249 (1975).
[21] H.-S. Kwon and Y.-H. Kim, ''Minimization of bias error due to windows in planar acoustic holography using a minimum error window,'' J. Acoust. Soc. Am. **98**, 2104–2111 (1995).

Adaptive time-reversal mirror

J. S. Kim,[a] H. C. Song, and W. A. Kuperman
Marine Physical Laboratory, Scripps Institute of Oceanography, University of California, San Diego, La Jolla, California 92093-0238

(Received 20 December 2000; revised 1 February 2001; accepted 2 February 2001)

The time-reversal mirror uses the received signal from a probe source to refocus the signal at the probe source location by backpropagating the time-reversed version of the received signal. In this study, an adaptive method is described to steer a null to an arbitrary position in a waveguide while maintaining a distortionless response at the probe source location. As an application, selective focusing in free space is demonstrated. © *2001 Acoustical Society of America.*
[DOI: 10.1121/1.1358299]

PACS numbers: 43.20.Fn, 43.30.Vh [ANN]

I. INTRODUCTION

The time-reversal mirror (TRM) has been demonstrated in ultrasonics[1,2] and underwater acoustic environments.[3] Further, the theory has been extended to the detection of scatterers and to focusing through an inhomogeneous medium. Recently, the development of the DORT method[4] allowed the selective focusing on a weak scatterer.

In this paper, the concept of an adaptive weighting on the TRM array before backpropagation is introduced to steer nulls. Accordingly, the expression for the weight vector to steer a null at the arbitrary location, assuming that the cross-spectral density matrix (CSDM) for the null location is known, is formulated based on optimization theory with constraints.[5–7] Then, it is demonstrated in simulation that the null can be steered in an ocean waveguide. Even when the CSDM is not available, a null in the vicinity of the probe source still can be steered in the range direction based on the theory on the invariants of the waveguide,[8–11] for which simulation results are also given.

As an application of null steering, selective focusing on the weaker target among two targets is demonstrated, and compared with the previous work known as the DORT method.[4,12] In order to focus on the strong target, the iterative time-reversal mirror[13–15] is used. It is shown that the focused field on the strong scatterer is equivalent to the field due to the eigenvector with the largest eigenvalue propagated to the control plane, where the scatterers are located. Since the scattered field from the strong scatterer contains the pressure vector from the scatterer to be nulled, the CSDM is constructed to compute the weighting vector to steer the null. Once the adaptive weighting on the array is computed, the time-reversal mirror can be focused onto the weak target with one iteration. The simulation shows that the focused field in this way gives the identical field to that resulting from the eigenvector with the second largest eigenvalue of the time-reversal operator (TRO) of the DORT method. It is noted that the DORT method requires measurement of the $(N\times N)$ interelement response matrix, while the selective focusing method based on nulling process needs only a few iterations depending on the relative strength of the targets. Typically the number of iteration requires less than the number of array elements. For two targets with target strengths differing by 6 dB, it has been found that ten iterations in total were sufficient to focus on the weaker target.

The potential application of null steering can be found in many areas, including selective focusing, silencing certain locations during transmission and communication associated with TRM, suppressing reverberation from discrete scatterers, and the practical problems that require manipulation of the field based on nulling.

In Sec. II, the formulation of TRM is reviewed and extended to accommodate adaptively weighted backpropagation. Simulation results are presented in Secs. III and IV for null steering and selective focusing, respectively.

II. THEORY

A. Review of TRM

As described in Fig. 1, the phase-conjugate field at the field location $\vec{r}$ in frequency domain is written as[10]

$$p(\vec{r})=\sum_{i=1}^{N} g^*(\vec{r}_i|\vec{r}_{ps})g(\vec{r}|\vec{r}_i)=\mathbf{g}^{\dagger}(\mathbf{r}_{\text{array}}|\vec{r}_{ps})\mathbf{g}(\vec{r}|\mathbf{r}_{\text{array}}), \quad (1)$$

where $g(\vec{r}_i|\vec{r}_{ps})$ represents the received acoustic pressure at the ith array element location $\vec{r}_i$ propagated from the probe source position $\vec{r}_{ps}$. Likewise, $g(\vec{r}|\vec{r}_i)$ represents the field propagated from the ith array element location $\vec{r}_i$ to the arbitrary receiver location $\vec{r}$ as shown in Fig. 1. N is the number of array elements. Superscripts $(\,)^*$ and $(\,)^{\dagger}$ denote complex conjugate and Hermitian transpose, respectively. In a vector notation, $\mathbf{g}$ and $\mathbf{r}_{\text{array}}$ are $(N\times 1)$ column vectors. Note that the position vectors are written in *italic* letters with arrows and the column vectors and matrices are written in boldface letters.

[a]The work was done during the author's sabbatical leave at MPL. Permanent affiliation: Korea Maritime University, Pusan, 606-701, Korea. Electronic mail: jskim@hanara.kmaritime.ac.kr

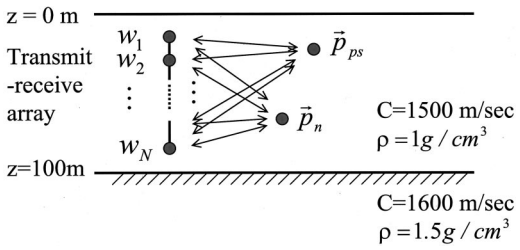


FIG. 1. Description of Pekeris waveguide.

B. Derivation of weight vector

Introducing a signal vector for backpropagation

$$\mathbf{w}=\begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_{N-1} \\ w_N \end{bmatrix}, \tag{2}$$

Eq. (1) can, more generally, be written as

$$p(\vec{r})=\sum_{i=1}^{N} w_i^* g(\vec{r}|\vec{r}_i)=\mathbf{w}^\dagger \mathbf{g}(\vec{r}|\mathbf{r}_{\text{array}}), \tag{3}$$

where $\mathbf{w}$ reduces to $\mathbf{g}$ in a conventional time-reversal mirror.

Now, let us find a signal vector $\mathbf{w}$ which gives the distortionless response at the focal point and minimizes the intensity coming from elsewhere when backpropagated. Denoting $\mathbf{K}$ as the CSDM of the observed signal vector, this problem reduces to the optimization problem of the following objective function:

$$\min_{\mathbf{w}} \mathbf{w}^\dagger \mathbf{K} \mathbf{w}, \tag{4}$$

with the distortionless response constraint at the focal location, which can be expressed as

$$\mathbf{w}^\dagger \mathbf{g}(\vec{r}_{ps}|\mathbf{r}_{\text{array}})=1. \tag{5}$$

The solution is well known and referred to as the minimum variance method in the adaptive array signal processing;[5–7] that is

$$\mathbf{w}=\frac{\mathbf{K}^{-1}\mathbf{g}(\vec{r}_{ps}|\mathbf{r}_{\text{array}})}{\mathbf{g}^\dagger(\vec{r}_{ps}|\mathbf{r}_{\text{array}})\mathbf{K}^{-1}\mathbf{g}(\vec{r}_{ps}|\mathbf{r}_{\text{array}})}. \tag{6}$$

In practice, the signal vector $\mathbf{w}$ is found from the minimum variance formulation with diagonal loading known as white noise constraint (WNC) in order to make it robust to the noise and the inversion of matrix $\mathbf{K}$ possible.

For the case of steering a null with a single probe source, the CSDM is defined as

$$\mathbf{K}=\mathbf{g}(\vec{r}_{ps}|\mathbf{r}_{\text{array}})\mathbf{g}^\dagger(\vec{r}_{ps}|\mathbf{r}_{\text{array}})+\mathbf{g}(\vec{r}_n|\mathbf{r}_{\text{array}})\mathbf{g}^\dagger(\vec{r}_n|\mathbf{r}_{\text{array}}), \tag{7}$$

where $\vec{r}_n$ is the location where the null is to be placed. The signal vector $\mathbf{w}$ in Eq. (6) places a null at the location $\vec{r}_n$ where the Green's function is denoted as $\mathbf{g}=\mathbf{g}(\vec{r}_n|\mathbf{r}_{\text{array}})$. The properties of the signal vector $\mathbf{w}$ give the distortionless response in the look location when backpropagated, which is expressed as a constraint given in Eq. (5). In addition, since the weighting minimizes the acoustic intensity coming from elsewhere, the weighting tends to null out the contributions from other CSDMs such as the second term in Eq. (7), which satisfy the following requirement:

$$\min_{\mathbf{w}} |\mathbf{w}^\dagger \mathbf{g}(\vec{r}_n|\mathbf{r}_{\text{array}})|^2. \tag{8}$$

If the placement of nulls is needed at more than one location, additional CSDMs which correspond to the desired null locations can be added incoherently to Eq. (7).

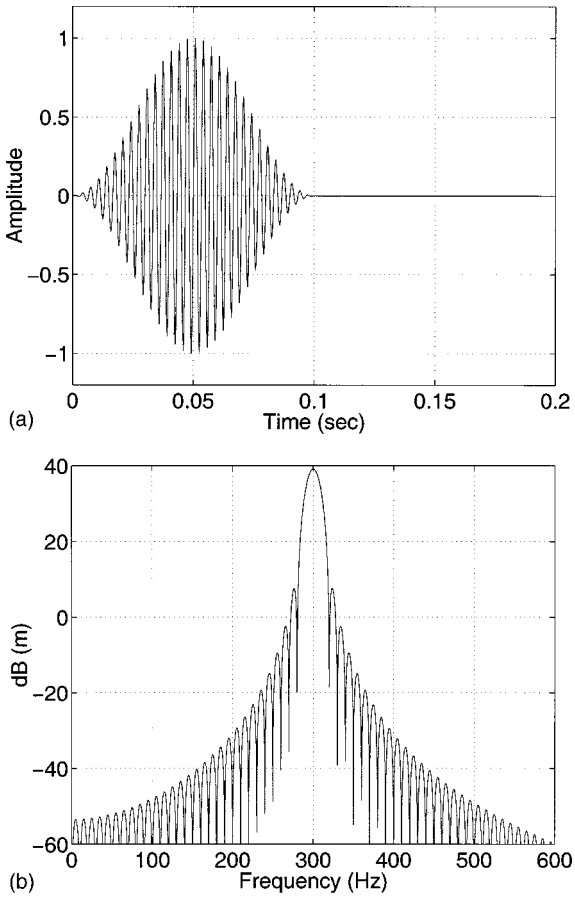


FIG. 2. Probe source (a) signal and (b) spectrum.

C. Theory on the waveguide invariants

As described above, the adaptive weight vector normally requires measurement of CSDM for the nulling location. The CSDM, however, can be predicted in an acoustic waveguide if the null is to be placed in the vicinity of the probe source in the range direction using the waveguide invariants theory.

In a dispersive and multimodal waveguide, the lines of constant sound intensity lead to a constant slope between the certain parameters of the waveguide.[8–11] The invariant, denoted as β, characterizes the relation between the range and frequency as

$$\beta=\frac{R}{\omega}\frac{\delta\omega}{\delta R}, \tag{9}$$

where R is a horizontal range and ω is an angular frequency.

For a Pekeris waveguide, β is one as shown in Fig. 9 so that

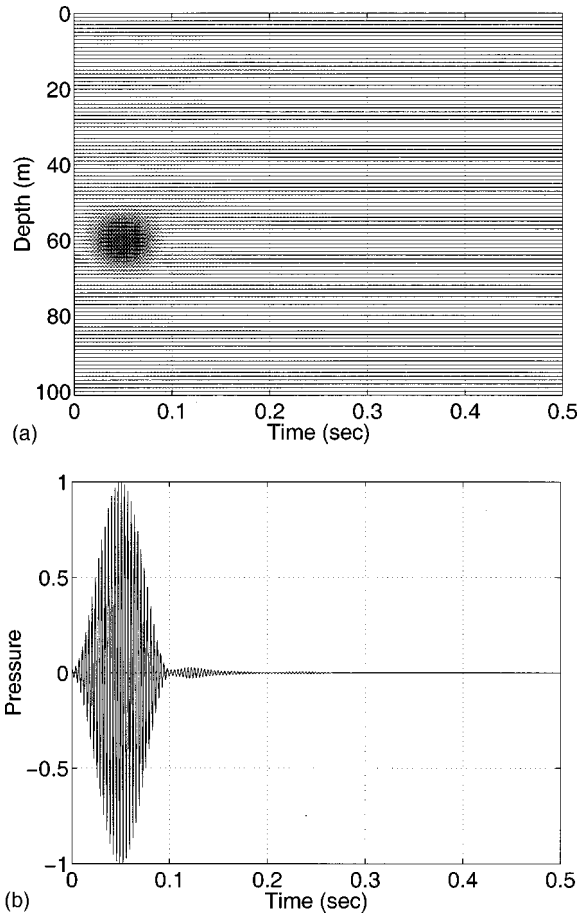


FIG. 3. Simulated time series focused on the probe source location at 60-m depth and 6000-m range.

$$R' = \frac{\omega'}{\omega} R. \tag{10}$$

Equation (10) states that the acoustic field at (R', ω) approximates the value at (R, ω'). Therefore, the signal vector received at the array can be used to calculate the CSDM in the vicinity of the probe source range at the same depth so as to place a null without measuring the Green's function from the null location.

III. SIMULATION IN WAVEGUIDE

In this section, the steering of null is demonstrated in a Pekeris waveguide as shown in Fig. 1. A time-reversal mirror is used to focus on the probe source location in Sec. III A. In Sec. III B, the null is steered to the location of which the transfer function to the array is known. In Sec. III C, the null is steered into the horizontal direction using the transfer function predicted from the theory on the waveguide invariants reviewed in Sec. II C.

For the purpose of simulation, the pulse with center frequency $f_c = 300$ Hz and ping duration $\tau = 0.1$ s is used (Fig. 2). The sampling frequency is 3600 Hz, FFT size is 8192, so that the time window is 2.3 s long. The probe source is located at $\vec{r}_{ps} = (r_{ps}, z_{ps}) = (6000\ \text{m}, 60\ \text{m})$.

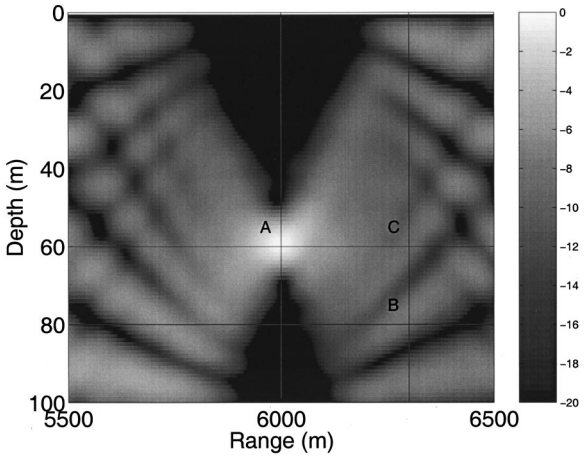


FIG. 4. Single frequency phase conjugation experiment at 300 Hz. The probe source is at 60-m depth and 6000-m range.

A. Time-reversal mirror

When the received signal at the array is time reversed and backpropagated, the signal focuses back to the probe source location as shown in Fig. 3. Figure 3(a) shows the time series received at the range 6000 m as function of depths. The time is referenced as the source signal for the focused signal, since the field response is 1 at the focal point. The focusing is noted at 60-m depth, the time series of which is shown in Fig. 3(b).

Figure 4 shows the focusing at the center frequency 300 Hz. The position A: $\vec{r}_{ps} = (r_{ps}, z_{ps}) = (6000\ \text{m}, 60\ \text{m})$ is the focal location, B: $\vec{r}_n = (r_n, z_n) = (6300\ \text{m}, 80\ \text{m})$ is the location to be nulled in Sec. III B, and C: $\vec{r}_n = (r_n, z_n) = (6300\ \text{m}, 60\ \text{m})$ is the location to be nulled in Sec. III C.

B. Placing a null at an arbitrary location with known field response

When the pressure vector or Green's function at the nulling position $\vec{r}_n = (r_n, z_n)$ from an array is known, the CSDMs are constructed using Eq. (7) and the weight vector is found from Eq. (6). The backpropagated field response with this weighting is shown in Fig. 5 for the center fre-

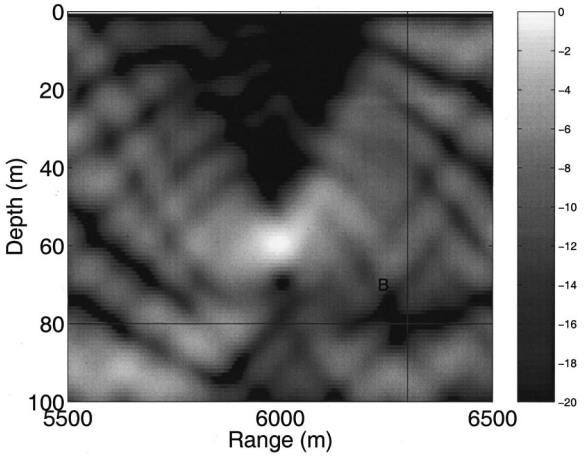


FIG. 5. Single frequency phase conjugation experiment at 300 Hz. The probe source is at 60-m depth and 6000-m range, and a null is placed at the range of 6300 m and at the depth of 80 m.

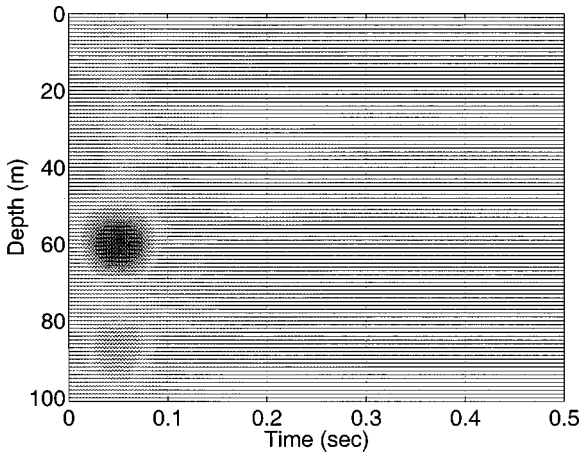


FIG. 6. Simulated backpropagated time series at range 6000 m with a null at 80-m depth and 6300-m range.

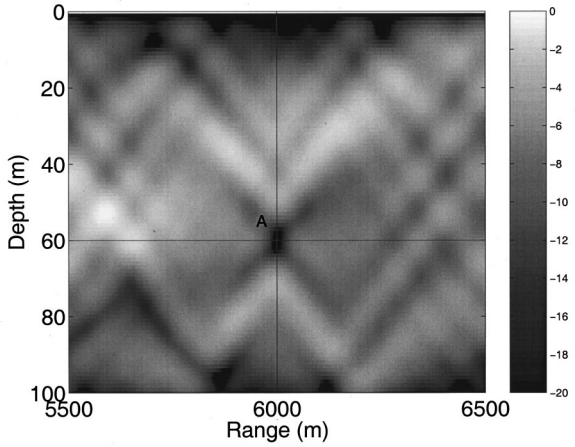


FIG. 8. Single frequency phase conjugation experiment at 300 Hz. The null is steered to the probe source location at the range of 6000 m and at the depth of 60 m.

quency of the probe source signal. The null at point B $\vec{r}_n=(r_n,z_n)$ is noted as a result of the nulling process.

The time series at the range of 6000 m is shown in Fig. 6. Although the distortionless response at the focal location $\vec{r}_{ps}=(r_{ps},z_{ps})$ is maintained, the weak sidelobes are observed at different depths. This is due to the shading effect of nulling, so that the field except the focal location at the range of 6000 m is modified to some extent. The degree of modification becomes insignificant when the pressure vectors from the probe source location and the nulling location are well resolved. In Fig. 7, the nulling at 80-m depth is shown. The sidelobe structure along the depth at the range of 6300 m is highly correlated with the sidelobe structure in Fig. 5 for the center frequency. Also, the time delay of 0.2 s due to the distance of 300 m from the focal location is noted.

In Fig. 8, a null is steered to the probe source location. A unit vector was used to construct the first term in Eq. (7), which is equivalent to steering the look location to the plane wave propagating normal to the array. The second term in Eq. (7) is now the signal vector from the probe source. This example demonstrates that null can be steered even without the specific look location as a probe source.

C. Placing a null and a focus at the same depth using information from probe source signal only

In Sec. III B, *a priori* knowledge of the pressure vector at the nulling location is required to steer the null. However, even if the pressure vector at the array from the nulling location is not known and the pressure vector from the probe source is the only information available, it is possible to steer the null at the same depth of the probe source within the range where the invariants of waveguide hold.

The transfer function at the null position is predicted based on the theory on the waveguide invariants as described in Sec. II C. Since the null is to be placed at the range of 6300 m, the frequency of the wave vector to be backpropagated in place of 300 Hz can be calculated from Eq. (10) as

$$f' = 300\ \text{Hz} \times \frac{6000\ \text{m}}{6300\ \text{m}} = 285.7\ \text{Hz}. \tag{11}$$

Similarly, the frequency bins are shifted according to Eq. (9), phase conjugated, and backpropagated to realize the null at 6300 m. The null position can also be predicted from Fig. 9.

In Fig. 10, the simulated field response to place a null at 60-m depth and 6300-m range is shown at the center frequency.

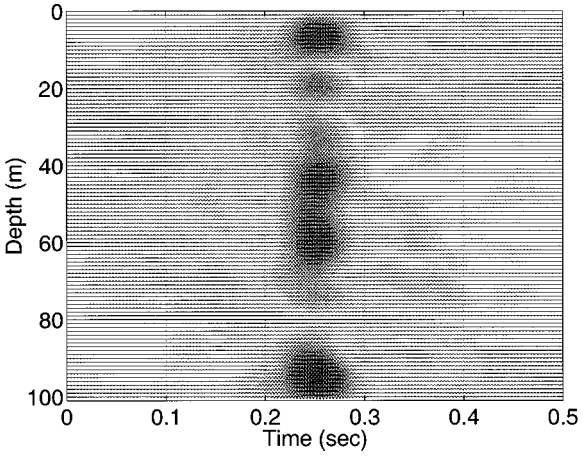


FIG. 7. Simulated backpropagated time series at range 6300 m. Note the null at 80-m depth. Also, 0.2 s of time delay is noted.

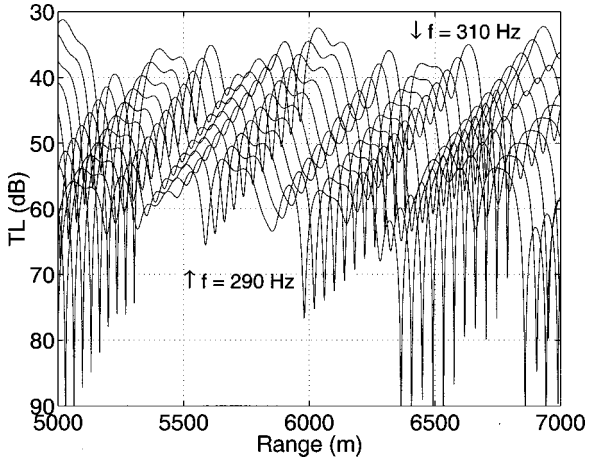


FIG. 9. Interference structure of the sound field in the Pekeris waveguide.

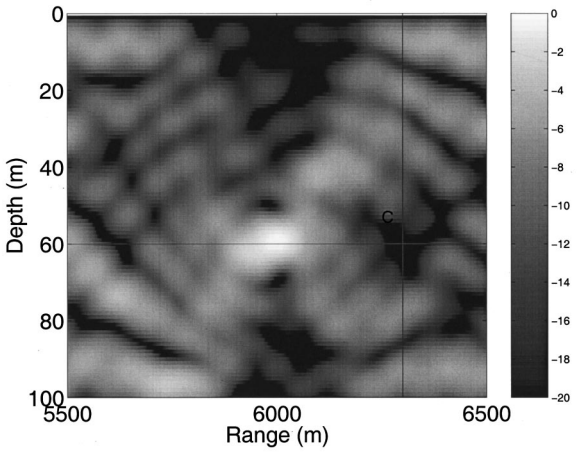


FIG. 10. Single frequency phase conjugation experiment at 300 Hz. The probe source is at 60-m depth and 6000-m range, and a null is placed at the range of 6300 m and at the depth of 60 m.

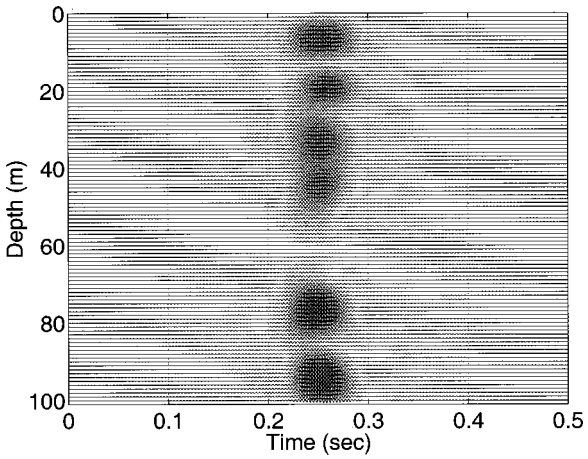


FIG. 12. Simulated backpropagated time series at range 6300 m. Note the null at 60-m depth.

Figure 11 shows the depth-stacked time series at 6000-m range. In Fig. 12, the depth-stacked time series at 6300-m range are displayed. The null at 60-m depth and time delay of 0.2 s is noted.

IV. SELECTIVE FOCUSING ON TWO TARGETS IN FREE SPACE

As an application of null steering, selective focusing associated with the iterative TRM is demonstrated in this section.

Selective focusing on the point-like scatterers has been demonstrated using the DORT method.[4,12] The DORT method utilizes the time-reversal operator (TRO) defined as **K*K**, where the matrix **K** is the transfer matrix, of which elements K_{lm} are defined as the pressure received at lth element with source located at mth element in a transmit–receive array. Note that this **K** is different from the one used in Eq. (4). When the eigenvalues of TRO are not degenerate, the eigenvectors of the TRO correspond to the pressure vectors received from the scatterers. The propagation of the eigenvectors results in selective focusing at each scatterer.

The focusing on the weak target based on nulling process requires the knowledge of CSDM at the location of the strong scatterer in order to calculate the nulling weight vector. This is achieved by an iterative time-reversal operation.[13–15] The received pressure vector from the strong scatterer provides the CSDM at the nulling location. Once the CSDM is known, the weight is calculated and applied on the array before backpropagation. This beam pattern has a null at the location of the strong target and, since the strong eigenvector is suppressed, the beam pattern focuses weakly on the second target. The scattering at this stage occurs from the second target, since the strong target has been nulled. Time reversing and backpropagating this signal yields a focus on the second target.

While the DORT method provides an elegant way for selective focusing, the construction of TRO requires measurement of the $(N\times N)$ interelement response matrix (or, half due to the symmetry of the **K** matrix) and the eigenvector decomposition. The nulling method requires only a few iterations to focus on the second target for the case of two scatterers.

The selective focusing process based on the nulling method is illustrated in the following subsections. Section IV A describes the setup for the numerical experiment and the simulation parameters. In Sec. IV B, the iterative time-

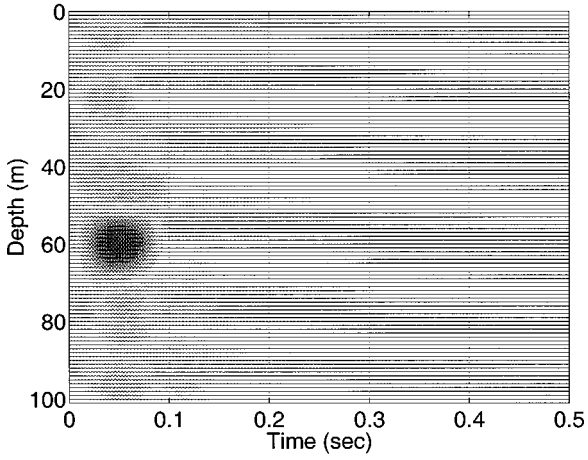


FIG. 11. Simulated backpropagated time series at range 6000 m.

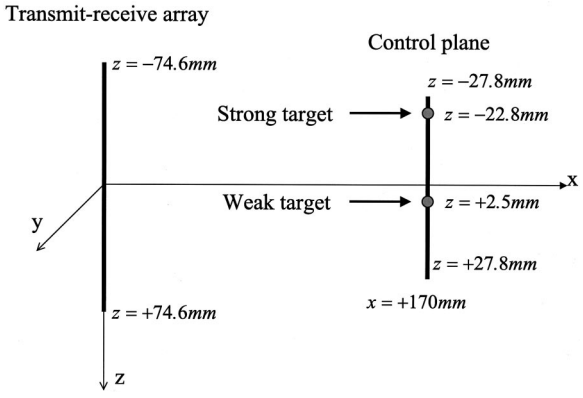


FIG. 13. Setup for numerical experiment of selective focusing.

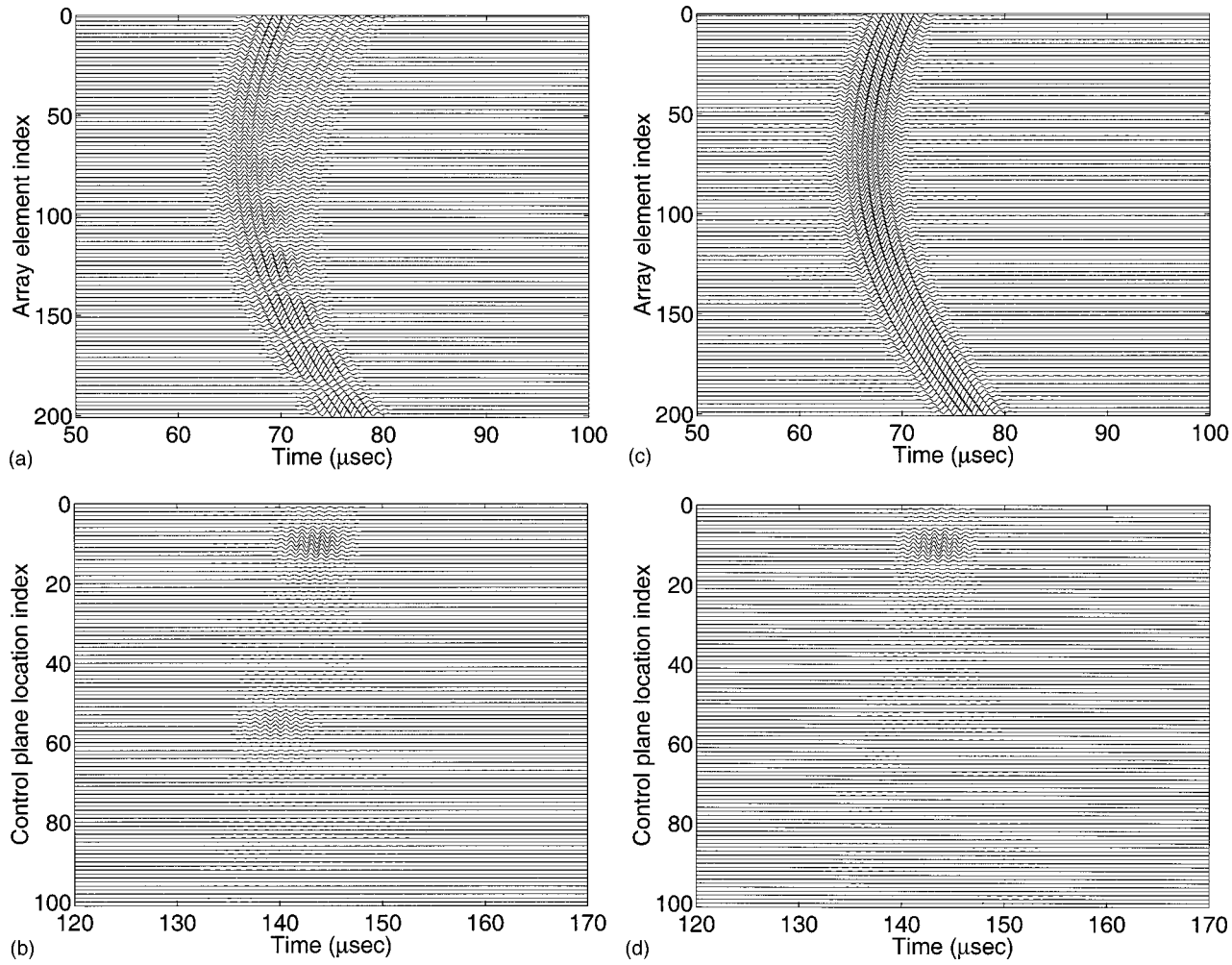


FIG. 14. The time series at the array: (a) and (c), and at the control plane: (b) and (d). (a) Shows the signals received at the array from the scatters insonified by a monopole source. (b) is for the received signals at the control plane when signal (a) is phase conjugated and backpropagated. (c) is the signal received at the array after ten iterations, and (d) is the received signal at the control plane when signal (c) is phase conjugated and backpropagated.

reversal mirror is demonstrated, with selective focusing on the weak scatterer in Sec. IV C.

A. Setup for numerical experiment

Figure 13 shows the schematic setup for the numerical experiment. This setup has been adopted from the experimental setup used in the demonstration of the DORT method[4] with minor modification. The center frequency is 1 MHz, and the sound speed is 1500 m/s. The transmit–receive array has 200 elements with element spacing of half a wavelength, so that z spans between $z=(-74.6,+74.6)$ mm. In order to make the mathematical manipulation more feasible, a control plane (more precisely a line) is defined at 170 mm apart from, and parallel with, the transmit–receive array. The pressure field along this control plane will be displayed to visualize the focusing on the scatterers, which are purposely placed on the control plane. This is not necessary for either the DORT method or the nulling method, since the scatterers can be spread arbitrarily in three-dimensional space. The control plane consists of 100 points. The spacing between the control points are 3/4 of that of the spacing in the transmit–receive array, so that z spans between $z=(-27.8, +27.8)$ mm. Again, this spacing is selected for the purpose of displaying the focusing with fine resolution. Two point scatterers are placed on the control plane. The strong scatterer is at $(x,y,z)=(170,0,-22.8)$, and the weak scatterer is at $(x,y,z)=(170,0,2.5)$. All units are in mm. The scatterers are separated in a way that the scatterers are well resolved. The difference in the target strength is 6 dB for the current simulation. The numerical experiment has been carried out with -30-dB Gaussian white noise. In the simulation, it is assumed that multiple scattering is negligible.

B. Iterative TRM on strong target

In order to achieve the focusing on the strong scatterer, the first step is to insonify the control plane using a point source. The reflected signals from two point scatterers are shown in Fig. 14(a). This signal shows the considerable reflection from the strong scatterer, while there is a weak signal from the second target. When this signal is time reversed and backpropagated, the signal received at the control plane is shown in Fig. 14(b). This process is iterated to achieve a focus on the strong target. After ten iterations, the signal received at the transmit–receive array is shown in Fig. 14(c), for which the time-reversed and backpropagated signal is shown in 14(d). It is clear that the signal is focused on the

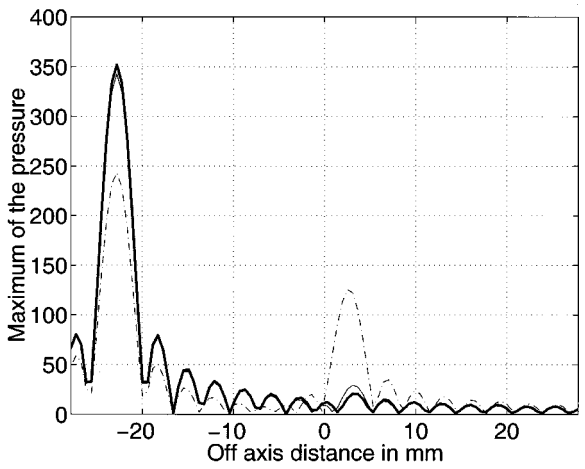


FIG. 15. The field received at the control plane at the center frequency of 1 MHz as the iteration increases. The broken line is the signal received from scatterers insonified by a point source. The amplitude at the strong scatterer is twice that of the weak scatterer. The solid thin line is after five iterations, and the thick line is for the tenth iterations. It shows that the frequency response converges to the strong scatterers.

location of the strong scatterer.

The field projected on the control plane for the center frequency is plotted in Fig. 15. The plot shows how the field is focused to the strong target as the iteration increases. The broken line shows the initial field at the control plane corresponding to the case shown in Fig. 14(b). The difference in the target strengths between scatterers is 6 dB. The thick solid line shows the focusing on the strong target after ten iterations.

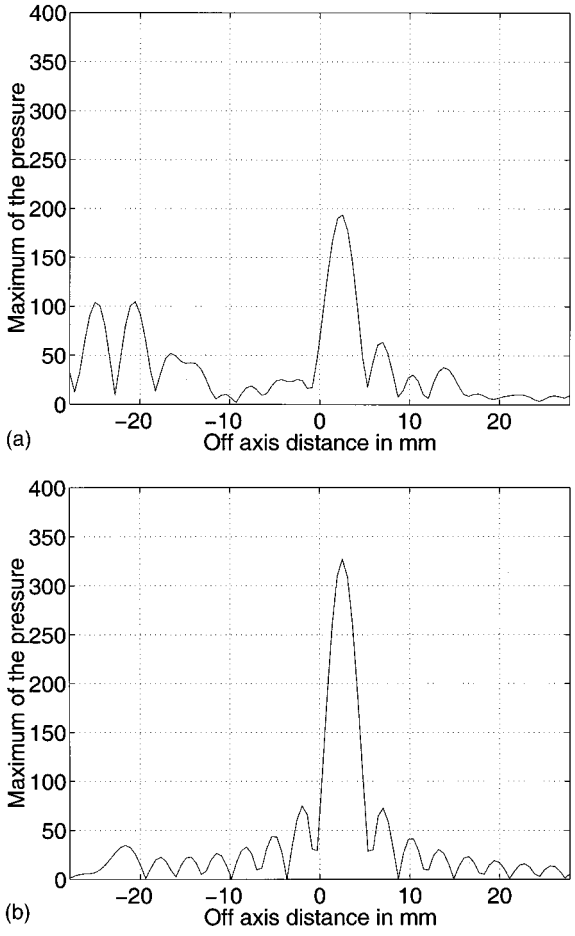


FIG. 16. The wave vectors received at the control plane with nulling. (a) The wave vectors received at the control plane after applying a nulling vector on the phase conjugated transmitting vector at the center frequency. (b) One iteration is applied to focus on the weak target.

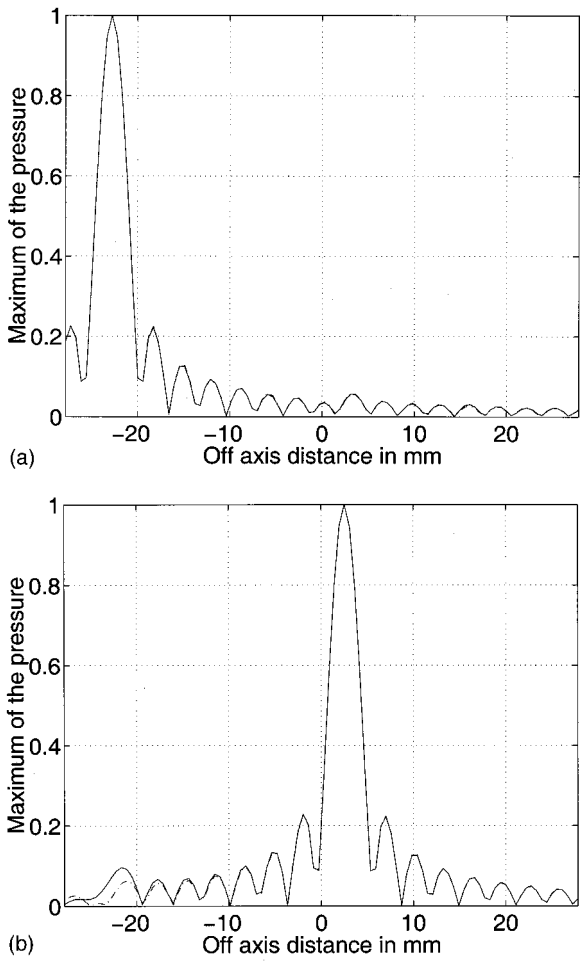


FIG. 17. The comparison with TRO (time-reversal operator). (a) shows the strong scatterer's eigenvectors (broken line: by DORT method; solid line: by iteration). (b) shows the weak scatterer's eigenvectors (broken line: by DORT method; solid line: by nulling method).

The rate of convergence is dependent on the relative target strength of the scatterers. The detailed convergence analysis can be found in Refs. 13–15.

C. Focusing on weak target

Once the focusing on the strong target is accomplished, the signal received at the receive array plays the role of a probe signal where we want to place a null, assuming the scatterer can be treated as a point scatterer. Now, the pressure vector received from the strong target is equivalent to $\mathbf{g}(\mathbf{r}_{\text{array}}|\vec{r}_n)$ in Eq. (7). Note that for this case $\vec{r}_n$ in the second term is the same as $\vec{r}_{ps}$ since the strong scatterer is to be nulled. However, there is the problem of setting the wave vector for which the weighted and backpropagated field has the distortionless response, which plays the same role as the first term in Eq. (7). Although this wave vector can be an arbitrary vector sufficiently far away from the nulling point

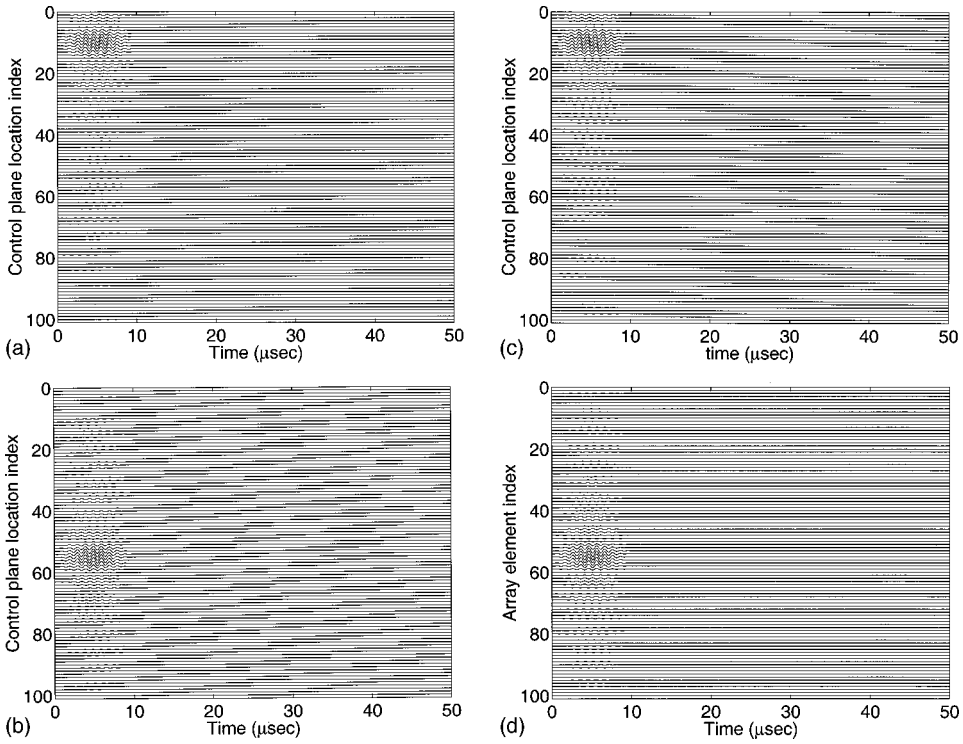


FIG. 18. (a) and (c) show the time series focused on the first target by nulling and DORT methods, respectively, and (b) and (d) show the time series focused on the second target by nulling and DORT method.

(or equivalently well resolved), a unit vector has been used in this study. The meaning of using a unit vector is interpreted as the wave vector received in the far field so that the distortionless response is achieved at the far field. It is noted that the random Gaussian number has been generated to be used as the wave vector for the distortionless response, and the result was unaffected, which is not surprising because the number of elements have enough degrees of freedom to null out the strong eigenvector.

Figure 16(a) shows the pressure vector received at the control plane with the nulling vector applied at the center frequency of 1 MHz. It is noted that considerable focusing is achieved on the weak target. It is considered that, when the strong target eigenvector component is suppressed, the second eigenvector is taking the place of the dominant wave vector. After one iteration, Fig. 16(b) shows that the focusing is achieved on the weak target. If the iteration continues, the strong target will pick up the energy and eventually dominate.

Figures 17(a) and (b) show the focused fields on the strong and the weak targets, respectively, using both the adaptive nulling method and the DORT method. Figure 18 corresponds to the time-domain solution of Fig. 17. Figures 18(a) and (b) show the signal received at the control plane and focused on the strong scatterer located at -22.8 mm, which corresponds to the control panel index of 10, and the weak target located at 2.5 mm, the control panel index of which corresponds to 55. Figures 18(c) and (d) are the same plots using the DORT method. The time delay for the nulling method is arbitrary since there are many iterations involved, while the time delay in the DORT method represents the distance from the transmit–receive array.

V. CONCLUSION

The concept of the weighted time-reversal mirror is introduced. This is used to steer nulls in the Pekeris ocean waveguide to the location where the pressure vector is known at the array while maintaining the distortionless response at the desired location such as a focal point. Using information from the probe source only, the theory on the waveguide invariants is utilized to demonstrate the steering of nulls in the range direction. As an application, selective focusing on the weak target among two point scatterers in free space is realized, giving the identical eigenvectors as obtained from the DORT method.

ACKNOWLEDGMENT

This work was supported by the Office of Naval Research.

[1] M. Fink, ''Time-reversal mirrors,'' J. Phys. D **26**, 1330–1350 (1993).
[2] M. Fink, ''Time-reversed acoustics,'' Phys. Today **50**, 34–40 (1997).
[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] C. Prada, S. Manneville, D. Spoliansky, and M. Fink, ''Decomposition of the time reversal operator: Detection and selective focusing on two scatterers,'' J. Acoust. Soc. Am. **99**, 2067–2076 (1996).
[5] Don H. Johnson and Dan E. Dudgeon, *Array Signal Processing—Concepts and Techniques*, Prentice-Hall Signal Processing Series, Alan V. Oppenheim, Series Editor (Prentice Hall, Englewood Cliffs, NJ, 1993).
[6] H. Cox, ''Robust adaptive beamforming,'' IEEE Trans. Acoust., Speech, Signal Process. **ASSP-35**, 1365–1376 (1987).
[7] J. Capon, ''High resolution frequency wavenumber spectrum analysis,'' Proc. IEEE **58**, 1408–1418 (1969).
[8] S. D. Chuprov, ''Interference structure of a sound field in a layered

ocean,'' Acoustics of the Ocean: Current Status (in Russian), edited by L. M. Brekhovskikh and I. B. Andreevoi (Nauka, Moscow, 1982), pp. 71–91.
[9] G. A. Grachev, ''Theory of acoustic field invariants in layered waveguide,'' Acoust. Phys. **39**, 33–35 (1993).
[10] H. C. Song, W. A. Kuperman, W. S. Hodgkiss, T. Akal, and C. Ferla., ''A time-reversal mirror with variable range focusing,'' J. Acoust. Soc. Am. **103**, 3234–3240 (1998).
[11] 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).
[12] N. Mordant, C. Prada, and M. Fink, ''Highly resolved detection and selective focusing in a waveguide using the D.O.R.T. method,'' J. Acoust. Soc. Am. **105**, 2634–2642 (1999).
[13] C. Prada, F. Wu, and M. Fink, ''The iterative time reversal mirror: A solution to self-focusing in the pulse echo mode,'' J. Acoust. Soc. Am. **90**, 1119–1129 (1991).
[14] C. Prada, J. L. Thomas, and M. Fink, ''The iterative time reversal process: Analysis of the convergence,'' J. Acoust. Soc. Am. **97**, 62–71 (1995).
[15] H. C. Song, W. A. Kuperman, and W. S. Hodgkiss, ''Iterative time reversal in the ocean,'' J. Acoust. Soc. Am. **105**, 3176–3184 (1999).

Roughness characterization of porous soil with acoustic backscatter

Michael L. Oelze,[a] James M. Sabatier, and Richard Raspet
National Center for Physical Acoustics and Department of Physics and Astronomy, University of Mississippi, University, Mississippi 38677

(Received 28 November 2000; revised 22 February 2001; accepted 23 February 2001)

The use of acoustics to determine the pore properties of soils, such as porosity, permeability, and tortuosity, is well established. A theoretical surface impedance and complex bulk wavenumber was developed by K. Attenborough for porous media that incorporated the soil pore properties as parameters [J. Acoust. Soc. Am. **73**, 785–799 (1983)]. Acoustic level difference measurements were used as a noninvasive means of finding the soil pore properties. Acoustic reflection measurements showed that the sound field over porous rough surfaces is modified by the surface impedance and by surface roughness. It is not possible to separate the signal modification due to impedance and the signal modification from roughness scattering in a forward scattering measurement. In order to accurately determine the soil pore properties, the roughness effects must be known independently from the surface impedance. A means of measuring roughness apart from impedance would allow the effects of roughness to be taken out of the level difference measurements. The underwater acoustics community has used acoustic backscatter for many years to examine surface roughness. The feasibility of adapting these acoustic backscatter techniques to outdoor porous soil surfaces is examined.
[DOI: 10.1121/1.1366320]

PACS numbers: 43.20.Fn, 43.30.Hw [ANN]

I. INTRODUCTION

Predictions of sound propagation over porous soil using three pore parameters have been fit to measured data to determine the pore properties. This work was based upon models constructed by Attenborough to characterize the properties of porous media.[1] The Attenborough models have been extensively compared to soil measurements and therefore are useful for modeling the range expected.[2,3] These models calculate the complex bulk density and bulk wavenumber from the porosity, tortuosity, and flow resistivity of soils. The low frequency limits of the density and wavenumber are given by

$$\rho_b(\omega)=\frac{\rho_0}{\Omega}\left(\frac{4}{3}T+i\frac{4\sigma_{\mathrm{eff}}}{\omega\rho_0}\right), \tag{1}$$

$$k_b(\omega)=k_0\sqrt{\gamma}\left(aT+i\frac{4\sigma_{\mathrm{eff}}}{\rho_0\omega}\right)^{1/2}. \tag{2}$$

In these equations γ is the ratio of specific heats in air, k_0 is the wavenumber in air, ρ_0 is the density of air, T is the tortuosity, Ω is the porosity, $\sigma_{\mathrm{eff}}=s_p^2\sigma\Omega$ is the effective flow resistivity with s_p^2 the pore shape factor and σ the flow resistivity, and a is defined by

$$a=\frac{4}{3}-\left(\frac{\gamma-1}{\gamma}\right)N_{\mathrm{Pr}},$$

where N_{Pr} is the Prandtl number. The bulk density and bulk wavenumber of the soil can be combined to calculate the normalized surface impedance

$$Z(\omega)=\left[\frac{1}{\rho_0 c_0}\right]\left[\frac{\omega\rho_b(\omega)}{k_b(\omega)}\right], \tag{3}$$

where c_0 is the speed of sound in air.

In a review of sound propagation over a smooth impedance surface, calculations of the sound pressure level emanating from a point source are described by Attenborough *et al.*[4] In that review paper, the sound pressure level is determined by the source-receiver geometry, the frequency of interest, and the impedance of the surface. The received sound level is the combination of the direct signal from the source to receiver and the signal reflected off of the surface. The reflected signal contains information about the surface since it is altered in amplitude and phase upon reflection. This information is embedded in a reflection coefficient, which is a function of the angle of incidence and the surface impedance. Information about the porous properties of the soil can be inferred from the value and frequency dependency of the surface impedance. Sabatier *et al.* used the acoustic level difference technique illustrated in Fig. 1 to measure the pore properties of soils.[2] The level difference technique is attractive because it is noninvasive method of measuring pore properties. Later work showed only the ratio of tortuosity/porosity and flow resistivity/porosity could be determined.[3]

In taking the level difference measurements, it was found that the experimental results matched fairly well with the developed models over smooth surfaces. However, when the theoretical model was compared with results from ex-

[a]Present address: Bioacoustics Research Laboratory, Department of Electrical and Computer Engineering, University of Illinois at Urbana–Champaign, 405 North Mathews Avenue, Urbana, IL 61801. Electronic mail: oelze@brl.uiuc.edu

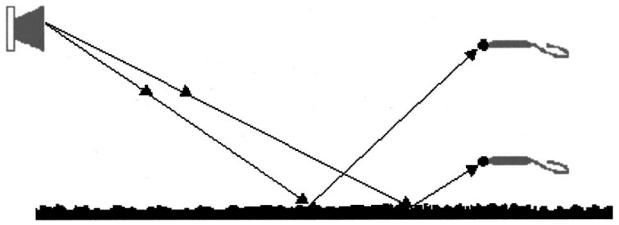


FIG. 1. Level difference measurement experimental setup showing sound source and two vertically separated microphones above the ground surface.

periments over rough surfaces, the agreement diverged. Hence, roughness was causing a significant effect upon the forward propagation of sound.

Attenborough proposed that the roughness of a surface might be able to be characterized by examining the sound reflected from the surface.[5] Efforts were then made to model the effects of roughness on the propagation of sound over porous surfaces. Attenborough *et al.* chose to treat roughness as an effective admittance, or inverse impedance.[6]

Attenborough used the analysis by Howe[7] for constructing an effective admittance for hemispherical roughness elements, valid for low grazing angles. In this model, the rough surface was treated as a smooth surface with an altered admittance. Combining Howe's results with results from Tolstoy,[8] Attenborough derived an effective admittance for roughness scatterers of arbitrary shape. The effective admittance is

$$\beta_3^* = \beta_s - i(k_0\sigma_v/2 + \beta_s k_b \sigma_v), \tag{4}$$

where β_s is normal surface admittance, $1/Z$, and σ_v is the roughness length scale defined as the projected volume of roughness per unit area of the surface. In laboratory experiments, the model predicted the effects of roughness well.[5]

Using the effective admittance from Attenborough, roughness experiments were conducted by Chambers on porous surfaces.[9] The experimental setup by Chambers was similar to Attenborough's level difference method except that the source and receiver were placed closer to the ground. In so doing, the patch ensonified by the source had a larger footprint. It was postulated that by allowing a larger area to react with the sound field, a better representation of the surface roughness could be obtained since local anomalies of rough or smooth sections would be averaged.

Chambers observed that the effects of roughness mimicked those of porosity and permeability. Roughness acts to decrease reflected sound levels for the grazing incidence geometry due to scattering. Increasing the permeability or porosity of a surface also decreases sound levels due to changes in viscous drag forces within the pore spaces of the ground. In the models themselves, the roughness and the other soil properties are not independent and, therefore, experimentally it is impossible to measure the effects independently. If one does not properly characterize the soil properties for the porosity and permeability with some other method, the observed acoustic response may be inaccurately apportioned between the effects of roughness and the other soil properties. The result is that measurements of the roughness scale will be inaccurate. The models may accurately describe the effects of roughness if the relevant assumptions are not violated and the soil properties are properly characterized. It may also be that the models for forward scattering need to be altered to better describe the field test data.

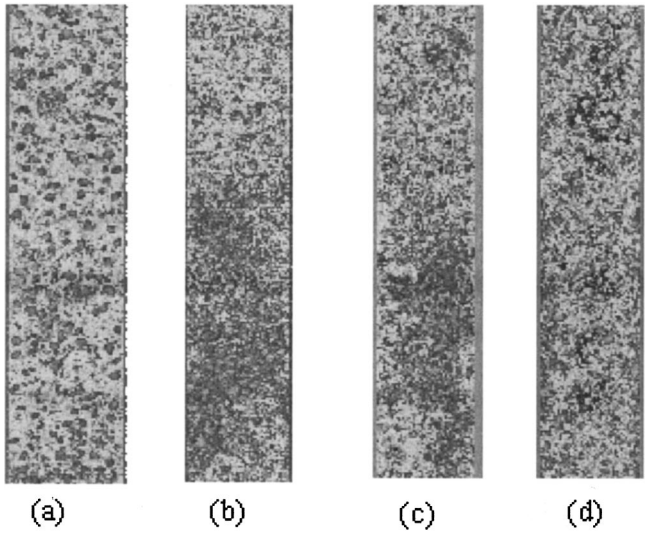


FIG. 2. Height profiles of four soil plots taken by laser microreliefmeter.

II. ROUGHNESS STATISTICS OF OUTDOOR SOIL SURFACES

One method that has been used to measure surface roughness without the effects of the acoustic surface impedance is the laser microprofiler.[10] The laser microprofiler measures the height profile of a surface with a resolution of 0.25 mm in the vertical direction and in square sections of $0.5 \times 0.5\,\mathrm{mm}^2$ in the horizontal. Römkens *et al.* used the laser profiler to measure the roughness of several soil surfaces.[11] Figure 2 shows profiles of four different soil tracts measured by the laser profiler.

From a surface profile, the roughness power spectrum can be derived. The power spectrum is the squared magnitude of the Fourier transform of the surface profile. The rms height and the correlation function can be derived from the 2-D power spectrum $W(k)$ by integrating over the roughness wavenumber, k, giving

$$h_{\mathrm{rms}}^2 = 2\pi \int_{k_c}^{\infty} W(k) k\, dk, \tag{5}$$

$$C(x) = \frac{2\pi}{h_{\mathrm{rms}}^2} \int_{k_c}^{\infty} W(k) e^{-ikx} k\, dk, \tag{6}$$

where k_c is the cutoff wavenumber. The correlation length is the length at which $C(x)$ decreases to $1/e$ of its initial value and describes how the roughness elements are packed together on the surface. The larger the correlation length for the roughness the more sparse the roughness elements are spread about on the surface.

Intuitively, not all scales of roughness will be included in any surface profile since the profile will have a finite length. The low wavenumber cutoff is chosen appropriate to the size of the roughness being examined. For example, if a low wavenumber cutoff were chosen which corresponds to a

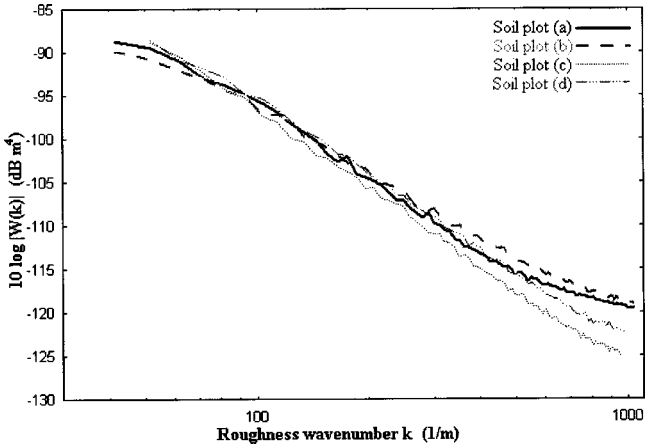


FIG. 3. Two-dimensional Power spectrums for the four soil profiles of Fig. 2 plotted in log–log space.

cutoff wavelength of several kilometers, the roughness of local hills and valleys would be incorporated into the statistics evaluated from the power spectrum. The cutoff wavelength determines the size of the plot from which the statistics of the surface are calculated. Appropriate cutoff wavenumbers for normal agricultural soil surfaces would correspond to wavelengths of a meter or less. Any values of rms height and correlation length should be given with reference to the cutoff wavenumber chosen.

Figure 3 shows the 2-D power spectrums for the four profiles of Fig. 2 obtained by the laser profiler. The soil plots were made by breaking up the soil into clods using various farming implements. As the soil is worked the larger clods continue to be broken down into smaller and smaller clods. Kolmogorov predicted that a process where particles are broken down into smaller and smaller particles would have a power law distribution.[12] The 2-D power spectrums in Fig. 3, which were plotted in log–log space, show that the power spectra for the soil profiles are also power laws over a range of wavenumbers. The power laws of the power spectrum for the random rough surfaces are of dimensionality such that the surfaces are fractal, which means the surfaces contain structure on all scales and the form of the structure is similar at each scale.

III. UNDERWATER BACKSCATTER

The underwater community has utilized acoustic backscatter techniques for many years to examine the roughness of ocean surfaces. For a review of acoustic backscatter in underwater sound, see Ref. 13. Backscatter is measured in terms of the scatter strength, which is defined as

$$S_s = 10 \log \sigma_s(\theta), \tag{7}$$

where σ_s is the backscatter cross-section.[14] The backscatter cross-section is determined by looking at the ratios of the scattered intensity to the intensity incident upon the surface. The cross-section is given by

$$\sigma(\theta) = \frac{r^2}{A}\frac{I_s}{I_0}, \tag{8}$$

where θ is the grazing angle, r is the distance from source to surface, A is the area of the ensonified surface, and I_s and I_0 are the scattered and incident intensities.

Scattering from rough surfaces can be modeled by use of small roughness perturbation theories. For rough surfaces, perturbation theory is valid under the restriction[13]

$$k_a h < 0.5, \tag{9}$$

where h is the rms height of the surface. First order perturbation theory under Dirichlet boundary conditions yields for the backscatter cross-section[15]

$$\sigma_p(\theta) = 4k_a^4 \sin^4 \theta\, W(2k_a \cos\theta), \tag{10}$$

where k_a is the acoustic wavenumber and $W(2k_a \cos\theta)$ is the 2-D power spectrum evaluated at the roughness wavenumber, $2k_a \cos\theta$. While scattering occurs at many points within the ensonified surface, only scatter points with roughness wavenumber separation on the surface equal to $2k_a \cos\theta$ are in phase for a given graze angle and wavelength of sound. Hence, the scatter contribution will predominately come from the wavenumber, $2k_a \cos\theta$, which is the Bragg wavelength.[16] In principle, the roughness power spectrum can be evaluated by inverting scatter strength measurements taken at different frequencies and graze angles. This would give the power spectrum in terms of a logarithmic function

$$10 \log W(k,\theta) = S_s(k,\theta) - 10 \log 4k^4 \sin^4 \theta. \tag{11}$$

It has been shown that profiles of underwater surfaces tend to have power roughness spectrums with power law behaviors similar to those for soil surfaces.[17,18] The fact that the power spectrums are power law means that in log-log space the power spectrum is linear. Theoretically this means that all that is needed to approximate the full power spectrum is two data points from which the slope and intercept of the spectrum line can then be calculated. The power law characteristic of the roughness power spectrum has been utilized to examine statistics of underwater surfaces.[18]

IV. BACKSCATTER FROM POROUS SOIL SURFACES

The success of underwater acoustic backscatter at examining sea floor roughness and the fact that agricultural soil surfaces display the power law behavior indicates that backscatter may be a viable means of measuring roughness of porous soil surfaces. The value of using acoustic backscatter on porous soil surfaces depends upon the ability of backscatter to give roughness statistics independent of the surface impedance effects. The effect of surface impedance on backscatter is investigated by incorporating impedance boundary conditions into the calculation of the perturbation cross-section.

Figure 4 shows the boundary of a rough surface of finite impedance with the incident, reflected, and transmitted velocity potentials given by

$$\Phi i = e^{-ik_1[\beta_{1x}x + \beta_{1y}y + \beta_{1z}z]}, \tag{12a}$$

$$\Phi r = \int\int e^{-ik_1[\beta_{rx}x + \beta_{ry}y + \beta_{rz}z]} T_r(\mathbf{k}_r, \mathbf{k}_i)\, d\mathbf{k}_r, \tag{12b}$$

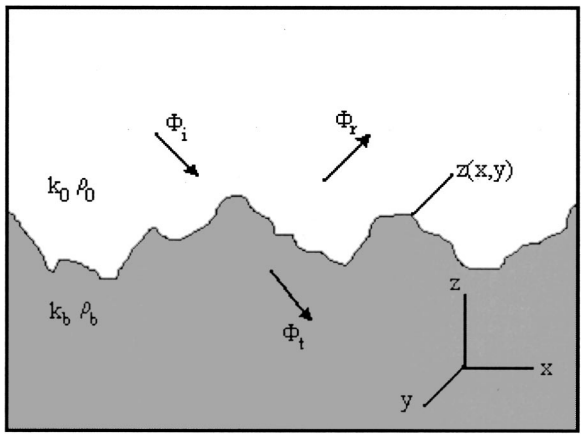


FIG. 4. Surface geometries for sound fields on a porous surface with roughness small compared to the wavelength.

$$\Phi t = \int\int e^{-ik_2[\beta_{tx}x+\beta_{ty}y+\beta_{tz}z]} T_t(\mathbf{k}_t,\mathbf{k}_i)\,d\mathbf{k}_t \tag{12c}$$

with the cosine angles represented by the β_{ij}'s, the $\mathbf{k}$'s representing the transverse acoustic wavenumbers, and T the transition matrix. The rms height of the roughness is assumed to be small compared to the wavelength of sound and the local slope is assumed to be small. To obtain the perturbation cross-section, the T-matrix is expanded in a power series in terms of h and S, the rms height and local slope of the surface, with

$$T_r = \sum_{m,n} h^m S^n T_{m,n}. \tag{13}$$

For a surface with height profile given by the function $z(x,y)$, the following relations hold:

$$\begin{aligned} &\langle z^2(x,y)\rangle = h^2, \quad z(x,y) = h\zeta(x,y), \\ &\langle \zeta^2(x,y)\rangle = 1, \quad S\zeta_i = \partial z/\partial x_i. \end{aligned} \tag{14}$$

The first-order correction to the T-matrix enables the first-order perturbation cross section to be calculated through the relation[19]

$$\sigma = (k_1\beta_{1z})^2 C(\mathbf{k}_r,\mathbf{k}_i), \tag{15}$$

$$\begin{aligned} C(\mathbf{k}_r,\mathbf{k}_i)\,\delta(\mathbf{k}_i-\mathbf{k}_i') = \langle (T_r(\mathbf{k}_r,\mathbf{k}_i) - \langle T_r(\mathbf{k}_r,\mathbf{k}_i)\rangle) \\ \times (T_r(\mathbf{k}_r,\mathbf{k}_i') - \langle T_r(\mathbf{k}_r,\mathbf{k}_i)\rangle)^*\rangle. \end{aligned} \tag{16}$$

To incorporate the soil surface pore properties, the boundary conditions for the continuous pressure and normal component of velocity at the surface are used:

$$\rho_1 \frac{\partial(\Phi_i+\Phi_r)}{\partial t} = \rho_2 \frac{\partial \Phi_t}{\partial t}, \tag{17}$$

$$\frac{\partial(\Phi_i+\Phi_r)}{\partial n} = \frac{\partial \Phi_t}{\partial n}. \tag{18}$$

Using the relation

$$\frac{\partial}{\partial n} = \frac{\partial}{\partial z} - S\left(\varsigma_i \frac{\partial}{\partial x_i}\right)_T \tag{19}$$

and solving the boundary condition equations for Φ_r in terms of Φ_i yields

$$\begin{aligned} &[-\rho\beta_{1z} + \kappa\beta_{tz} + S[\varsigma_j\beta_{1j}](\rho-1)]e^{-ik_1(\beta_{1j}x_j+\beta_{1z}z)} \\ &\quad = \int\int [-\rho\beta_{rz} - \kappa\beta_{tz} + S[\varsigma_j\beta_{rj}](1-\rho)] \\ &\qquad \times e^{-ik_1(\beta_{rj}x_j+\beta_{rz}z)} T_r\,d\mathbf{k}_r \end{aligned} \tag{20}$$

with

$$\rho = \frac{\rho_2}{\rho_1}, \quad \kappa = \frac{k_2}{k_1}.$$

Expanding the exponential

$$e^{-ik_1\beta_z z} = e^{-ik_1\beta_z h\varsigma(x,y)} \approx 1 - ik_1\beta_z h\varsigma(x,y)$$

and relating terms of order h^0S^0, h^1S^0, and h^0S^1 in Eq. (20) gives for h^0S^0

$$\begin{aligned} &(\rho\beta_{1z} - \kappa\beta_{tz})e^{-ik_1(\beta_{1j}x_j)} \\ &\quad = \int\int (\rho\beta_{rz} + \kappa\beta_{tz})e^{-ik_1(\beta_{rj}x_j)} T_{0,0}\,d\mathbf{k}_r; \end{aligned} \tag{21}$$

for h^1S^0

$$\begin{aligned} &(\rho\beta_{1z} - \kappa\beta_{tz})e^{-ik_1(\beta_{1j}x_j)}[-ik_1\beta_{1z}\varsigma(\mathbf{x})] \\ &\quad = \int\int (\rho\beta_{rz} + \kappa\beta_{tz})[(ik_1\beta_{1z}\varsigma(\mathbf{x}))T_{0,0} + T_{1,0}] \\ &\qquad \times e^{-ik_1(\beta_{rj}x_j)}\,d\mathbf{k}_r; \end{aligned} \tag{22}$$

for h^0S^1

$$\begin{aligned} &[\varsigma_j\beta_{1j}](\rho-1)e^{-ik_1(\beta_{1j}x_j)} \\ &\quad = \int\int \{[\varsigma_j\beta_{rj}](1-\rho)T_{0,0} \\ &\qquad - [\rho\beta_{rz} + \kappa\beta_{tz}]T_{0,1}\}e^{-ik_1(\beta_{rj}x_j)}\,d\mathbf{k}_r. \end{aligned} \tag{23}$$

The Fourier transform of Eq. (21) yields $T_{0,0}$; once $T_{0,0}$ is known, $T_{1,0}$ and $T_{0,1}$ can be obtained from the Fourier transforms of Eqs. (22) and (23). The results are

$$T_{0,0}(\mathbf{k}_r,\mathbf{k}_i) = \mathrm{R}\,\delta(\beta_{1x}-\beta_{rx})\,\delta(\beta_{1y}-\beta_{ry}), \tag{24}$$

$$T_{1,0}(\mathbf{k}_r,\mathbf{k}_i) = \frac{-1}{(2\pi)^2}\int\int_x 2ik_1\beta_{1z}\mathrm{R}e^{i(\mathbf{k}_r-\mathbf{k}_i)\cdot\mathbf{x}}\varsigma(\mathbf{x})\,d\mathbf{x}, \tag{25}$$

$$\begin{aligned} T_{0,1}(\mathbf{k}_r,\mathbf{k}_i) = \frac{1}{(2\pi)^2}\int\int_x \frac{(1-\rho)(1+\mathrm{R})}{\rho\beta_{rz}+\kappa\beta_{tz}} \\ \times(\varsigma_j\beta_{1j})e^{i(\mathbf{k}_r-\mathbf{k}_i)\cdot\mathbf{x}}\,d\mathbf{x}, \end{aligned} \tag{26}$$

where R is the Rayleigh reflection coefficient

$$\mathrm{R} = \frac{\rho\beta_{1z} - \kappa\beta_{tz}}{\rho\beta_{1z} + \kappa\beta_{tz}}. \tag{27}$$

The first order T's are seen to be related to the Fourier transform of the surface function by

$$hT_{1,0}(\mathbf{k}_r,\mathbf{k}_i) = -2ik_1\beta_{1z}\mathrm{R} \times\left\{\frac{1}{(2\pi)^2}\int\int_x z(\mathbf{x})e^{i(\mathbf{k}_r-\mathbf{k}_i)\cdot\mathbf{x}}d\mathbf{x}\right\}$$
$$hT_{1,0}(\mathbf{k}_r,\mathbf{k}_i) = -2ik_1\beta_{1z}\mathrm{R}Z(\mathbf{k}_r-\mathbf{k}_i), \tag{28}$$

where $Z(\mathbf{k}_r-\mathbf{k}_i)$ is the Fourier transform of the surface profile $z(\mathbf{x})$. From Eq. (14),

$$S\varsigma_j\beta_{1j} = \beta_{1j}\frac{\partial}{\partial x_j}z(\mathbf{x}),$$

which gives

$$ST_{0,1}(\mathbf{k}_r,\mathbf{k}_i) = \frac{(1-\rho)(1+\mathrm{R})}{\rho\beta_{rz}+\kappa\beta_{tz}} \times\left\{\frac{1}{(2\pi)^2}\int\int_x \frac{\mathbf{k}_i\cdot\nabla z(\mathbf{x})}{k_1}e^{i(\mathbf{k}_r-\mathbf{k}_i)\cdot\mathbf{x}}d\mathbf{x}\right\}. \tag{29}$$

Inserting the Fourier transform

$$z(\mathbf{x}) = \int\int_{\mathbf{k}_r'-\mathbf{k}_i'} Z(\mathbf{k}_r'-\mathbf{k}_i')e^{-i(\mathbf{k}_r'-\mathbf{k}_i')\cdot\mathbf{x}}d(\mathbf{k}_r'-\mathbf{k}_i')$$

into Eq. (29) yields

$$ST_{0,1}(\mathbf{k}_r,\mathbf{k}_i) = -i\frac{\mathbf{k}_i\cdot(\mathbf{k}_r-\mathbf{k}_i)}{k_1}\frac{(1-\rho)(1+\mathrm{R})}{\rho\beta_{rz}+\kappa\beta_{tz}} \times Z(\mathbf{k}_r-\mathbf{k}_i). \tag{30}$$

For backscatter $\mathbf{k}_r=-\mathbf{k}_i$, $\beta_{rz}=-\beta_{1z}=\sin\theta$, $\mathbf{k}_i^2=\cos^2\theta$, and $\beta_{tz}=[1-(\cos\theta/\kappa)^2]^{1/2}$. When $T_r=hT_{1,0}+ST_{0,1}$ is inserted into Eq. (16), $C(\mathbf{k}_r,-\mathbf{k}_r)$ is obtained for the first order in T. From $C(\mathbf{k}_r,-\mathbf{k}_r)$, the first order perturbation, backscatter cross-section is determined

$$\sigma_p(\theta) = 4k_a^4\sin^4\theta|Y(k_a,\theta)|^2W(2k_a\cos\theta), \tag{31}$$

where

$$Y(k_a,\theta) = \frac{(\rho-1)^2\cos^2\theta+\rho^2-\kappa^2}{[\rho\sin\theta+\sqrt{\kappa^2-\cos^2\theta}]^2} \tag{32}$$

is a modified reflection coefficient that agrees with the result obtained by both Kuo and Moe.[20,21] Substituting the bulk wavenumber and bulk density for a porous soil surface yields the modified reflection coefficient in terms of the surface impedance

$$Y(k_a,\theta) = \frac{[Z-k_a/k_b]^2\cos^2\theta+Z^2-1}{[Z\sin\theta+\sqrt{1-(k_a\cos\theta/k_b)^2}]^2}. \tag{33}$$

The contribution of the impedance to the backscatter strength is then given by

$$S_s(Z,k_b:k_a,\theta) = 20\log|Y(Z,k_b:k_a,\theta)|. \tag{34}$$

Analysis of the ground impedance term given in Eq. (34) shows the sensitivity of acoustic backscatter to impedance effects. The effective flow resistivity spans the broadest range of values for the soil pore properties, from 1000 mks rayls/m for snow to $O(10^6)$ for wet loess surfaces.[22] The contribution of the effective flow resistivity dominates the contribution of the other soil properties for acoustically medium to hard surfaces. Acoustically medium to hard surfaces are defined here as surfaces with an effective flow resistivity of 300 000-mks rayls/m or higher. Typical values range between 1 and 10 for tortuosity and between 0.3 and 0.6 for porosity in soils. The values for tortuosity and porosity will be less than 0.01% of flow resistivity values for acoustically harder surfaces.

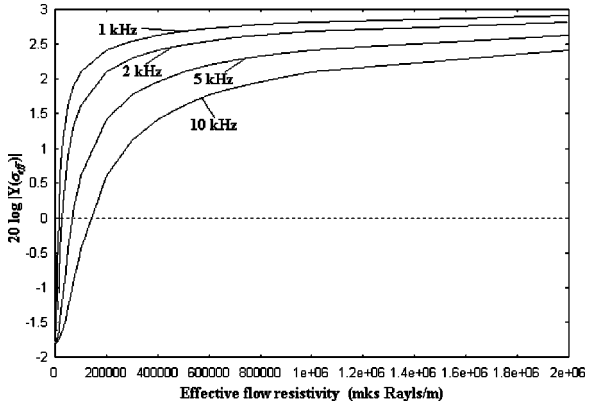


FIG. 5. Contribution of the modified reflection coefficient, $Y(k,\theta)$, versus effective flow resistivity for various frequencies and at a graze angle of 65 degrees.

Figure 5 shows the scatter strength contribution of the modified reflection coefficient for various frequencies versus the effective flow resistivity. The important thing to note is that for frequencies of 10 kHz and below the variation in the contribution of the modified reflection coefficient is less than 1.5 dB for effective flow resistivities of 300 000-mks rayls/m and above. Flow resistivities of 300 000-mks rayls/m and above represent acoustically medium to acoustically hard surfaces. The fact that the contribution of the modified reflection coefficient varies little over these flow resistivity values indicates that the reflection coefficient contribution can be evaluated with minimal error by estimating an appropriate flow resistivity.

The contribution of the modified reflection coefficient is found by choosing likely pore parameters for the surface and evaluating it at a frequency and graze angle. The modified reflection coefficient contribution can then be used to determine the power spectrum of roughness by inverting the backscatter measurement from the perturbation backscatter cross-section. Scatter strength measurements may be inverted to evaluate a point on the power spectrum at some roughness wavenumber

$$S_s(k_a,\theta) = 10\log[4k_a^4\sin^4\theta|Y(k_a,\theta)|^2W(2k_a\cos\theta)], \tag{35}$$

$$10\log W(2k_a\cos\theta) = S_s(k_a,\theta) - 10\log[4k_a^4\sin^4\theta|Y(k_a,\theta)|^2]. \tag{36}$$

If the chosen pore parameters are incorrect, then error will be introduced into the predicted roughness measurement. Since the contribution of the reflection coefficient is dominated by the flow resistivity for acoustically harder surfaces, errors

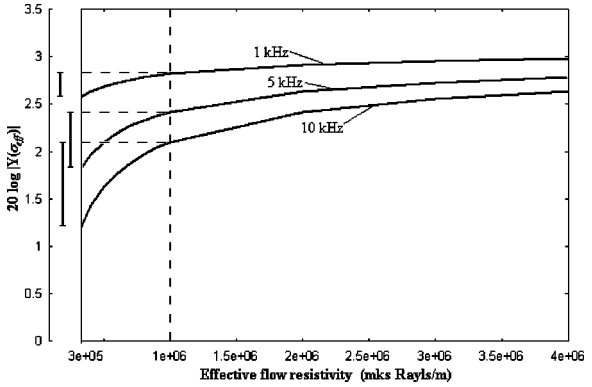


FIG. 6. Range of modified reflection coefficient contribution per range of flow resistivity values for three frequencies.

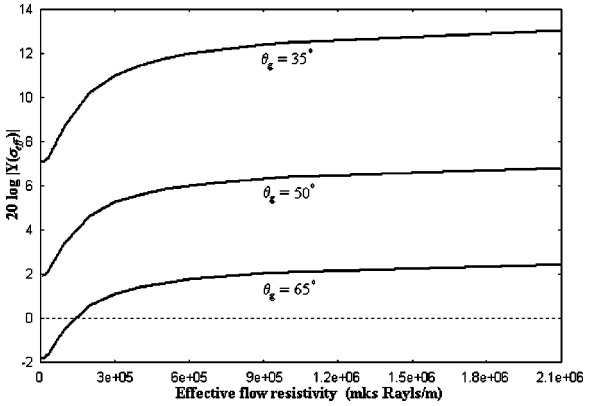


FIG. 7. Contribution of the modified reflection coefficient at various graze angles for 10-kHz signal.

from incorrect choices of tortuosity and porosity have negligible effect. For most outdoor soil surfaces, the frequencies used will be at 10 kHz and below. Assuming an acoustically harder surface and nothing more, the range of choices for the flow resistivity are still very broad. Figure 6 illustrates the range of decibel contribution of the modified reflection coefficient for the range of flow resistivities assumed for an acoustically harder surface. At 5 kHz, choosing an effective flow resistivity of 1×10^6-mks rayls/m, the maximum error introduced into the contribution of the modified reflection coefficient for acoustically medium to hard surfaces is at most about $\frac{1}{2}$ dB. An effective flow resistivity of 1×10^6-mks rayls/m is chosen because it is the middle value within the range of decibel contributions for flow resistivities of 300 000-mks rayls/m and greater. For typical soil surfaces with the rms height calculated for a 60×60-cm^2 section, $\frac{1}{2}$-dB error in the absolute level of the roughness power spectrum translates to about 5.7% error in the rms height calculation. As the frequency is increased, the potential for error from the choice of resistivity of 1×10^6-mks rayls/m increases such that at 10 kHz the maximum error is almost 1 dB to the modified reflection coefficient contribution. At lower frequencies, however, the potential for error from flow resistivity estimations is decreased. At 1 kHz, for example, the error to the modified reflection coefficient is at most about $\frac{1}{4}$ dB for a choice of 1×10^6-mks rayls/m for the flow resistivity. The fact that lower frequencies have less potential for error from the choice of flow resistivity means that mapping out the roughness power spectrum using backscatter may be more accurately accomplished by weighting the measurements taken at lower frequency.

The contribution of the modified reflection coefficient to the overall scatter strength is also affected by the angle of grazing. Figure 7 displays the contribution of the modified reflection coefficient at several graze angles for 10 kHz. The overall effect of the change in graze angle is to raise the absolute level of the reflection coefficient contribution with a slight increase in the slope of the curve for higher flow resistivities. For lower frequencies, the increase in slope of the decibel contribution of the modified reflection coefficient for the higher flow resistivities is even smaller. The impedance effects seen through the contribution of the modified reflection coefficient have a larger effect upon the backscatter strength for smaller graze angles. However, the ability to factor out the impedance effects from the roughness effects rests in the fact that the contribution of the modified reflection coefficient varies little for medium to acoustically hard surfaces. Even for smaller graze angles, errors in choice of flow resistivity yield minimal error in the modified reflection coefficient contribution for medium to acoustically hard surfaces.

It has been shown, theoretically, that the roughness of a porous soil surface can be measured independently from the surface impedance effects for acoustically medium to hard surfaces. The ability to measure the roughness independent from the effects of the pore properties for acoustically medium to hard surfaces using an assumed effective flow resistivity will be examined by experiment. First, backscatter measurements should be taken in the frequency range of 1–10 kHz with different graze angles. A value for the effective flow resistivity, $\sigma_{\text{eff}}=1\times10^6$-mks rayls/m, is assumed in order to calculate the approximate contribution of the modified reflection coefficient to the scatter strength. The roughness power spectrum, evaluated at the roughness wavenumber corresponding to the frequency and graze angle, can be determined by use of Eq. (36) with the modified reflection coefficient and the scatter strength. From the power spectrum, the rms height and correlation length can be calculated. Values for the flow resistivity can be independently measured using a Leonard's apparatus,[23] the method of Stinson and Daigle[24] or a probe microphone technique.[25] From the measured values of the effective flow resistivity, the actual contribution of the modified reflection coefficient can be determined. From the actual value of the modified reflection coefficient the power spectrum, rms height and correlation length can be determined. The effective error in the rms height and correlation length calculation can then be determined by comparing the values obtained with the assumed flow resistivity and the measured flow resistivity.

V. CONCLUSIONS

The scales of roughness for outdoor soil surfaces are typically centimeter in size, which implies that the wavelengths needed to characterize the roughness also be of centimeter size in order to meet the restriction of Eq. (9). Wave-

lengths of this size correspond to frequencies of 10 kHz and below in air. The contribution of the modified reflection coefficient is small and little error is induced by the use of an assumed flow resistivity for frequencies less than 10 kHz. Furthermore, typical weathered soil surfaces have effective flow resistivities of 300 000-mks rayls and greater. For weathered agricultural surfaces an assumed choice of effective flow resistivity, $\sigma_{\text{eff}}=1\times10^{6}$-mks rayls/m, results in minimal error to the value of the modified reflection coefficient. With the maximum possible error of $\frac{1}{2}$ dB to the modified reflection coefficient contribution for a frequency of 5 kHz, the error to the rms height calculation is 5.7%. The error to the rms height is further reduced by taking backscatter measurements at lower frequencies which will have less decibel error to the modified reflection coefficient for the assumed choice of flow resistivity.

Future work will consist of using acoustic backscatter techniques to map out the power spectrums of porous soil surfaces. Presently, the laser profile techniques used to obtain roughness statistics provides more information about a rough surface, however, scan time is extremely long. The acoustic backscatter technique should provide a quick means to obtain the rms height and correlation length statistics for a rough porous surface. The acoustic backscatter technique can then be used in combination with the forward scatter techniques to provide both roughness characterization and pore property characterization for a soil surface.

ACKNOWLEDGMENTS

This research was supported by grants from the USDA through the National Sedimentation Laboratory. We thank Dr. Matt Römkens for his direction in this work.

[1] K. Attenborough, ''Acoustical characteristics of rigid fibrous absorbents and granular media,'' J. Acoust. Soc. Am. **73**, 785–799 (1983).
[2] J. M. Sabatier, H. Hess, W. P. Arnott, K. Attenborough, M. J. M. Romkins, and E. H. Grissinger, ''*In situ* measurements of soil physical properties by acoustical techniques,'' Soil Sci. Soc. Am. J. **54**, 658–672 (1990).
[3] J. M. Sabatier, R. Raspet, and C. K. Fredrickson, ''An improved procedure for the determination of ground parameters using level difference measurements,'' J. Acoust. Soc. Am. **94**, 396–399 (1993).
[4] K. Attenborough, S. I. Hayek, and J. M. Lawther, ''Propagation of sound above a porous half-space,'' J. Acoust. Soc. Am. **68**, 1493–1501 (1980).
[5] K. Attenborough, ''Review of ground effects on outdoor sound propagation from continuous broadband sources,'' Appl. Acoust. **24**, 289–319 (1988).
[6] K. Attenborough and S. Taherzadeh, ''Propagation from a point source over a rough finite impedance boundary,'' J. Acoust. Soc. Am. **98**, 1717–1722 (1995).
[7] M. S. Howe, ''On the long range propagation of sound over irregular terrain,'' J. Sound Vib. **77**, 83–94 (1985).
[8] 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).
[9] J. P. Chambers, R. Raspet, and J. Sabatier, ''Incorporating the effects of roughness in outdoor sound propagation models,'' in *Proceeding of Noise-Con '96* (Conference on Noise Control Engineering, Seattle), pp. 905–910 (1996).
[10] M. J. M. Römkens, J. Y. Wang, and R. W. Darden, ''A Laser Microreliefmeter,'' Trans. *ASAE* **31**(2), 408–413 (1988).
[11] M. J. M. Römkens and J. Y. Wang, ''Roughness and sealing effect on soil loss and infiltration on a low slope,'' Advances in GeoEcology **31**, 589–595 (1998).
[12] A. N. Kolmogorov, ''On the logarithmic normal distribution rules of dimensions of particles by grinding,'' Mathematika **31**, 99–101 (1971).
[13] D. R. Jackson, D. P. Winebrenner, and A. Ishimaru, ''Application of the composite roughness model to high-frequency bottom backscattering,'' J. Acoust. Soc. Am. **79**, 1410–1422 (1986).
[14] R. J. Urick, *Principals of Underwater Sound* (McGraw-Hill, New York, 1983).
[15] J A Ogilvy, *Theory of Wave Scattering from Random Rough Surfaces* (Adam Hilger, Bristol, England, 1991).
[16] D. R. Jackson, K. B. Briggs, K. L. Williams, and M. D. Richardson, ''Tests of models for high-frequency seafloor backscatter,'' IEEE J. Ocean Eng. **21**, 458–470 (1996).
[17] D. R. Jackson and K. B. Briggs, ''High frequency bottom backscattering: Roughness versus sediment volume scattering,'' J. Acoust. Soc. Am. **92**, 962–977 (1992).
[18] P. D. Mourad and D. R. Jackson, ''High frequency sonar equation models for bottom backscatter and forward loss,'' in *Proceedings of Oceans '89* (Marine Technology Society and IEEE Oceanic Engineering Society, Seattle, 1989), pp. 1168–1989.
[19] E. I. Thorsos and D. R. Jackson, ''The validity of the perturbation approximation for rough surface scattering using a Gaussian roughness spectrum,'' J. Acoust. Soc. Am. **86**, 261–277 (1989).
[20] E. Y. T. Kuo, ''Wave scattering and transmission at irregular surfaces,'' J. Acoust. Soc. Am. **36**, 2135–2142 (1964).
[21] J. E. Moe and D. R. Jackson, ''First-order perturbation solution for rough surface scattering cross section including the effects of gradient,'' J. Acoust. Soc. Am. **96**, 1748–1754 (1994).
[22] K. Attenborough, ''Ground parameter information for propagation modeling,'' J. Acoust. Soc. Am. **92**, 418–427 (1992).
[23] R. W. Leonard, ''Simplified flow resistance measurements,'' J. Acoust. Soc. Am. **17**, 240–241 (1946).
[24] M. R. Stinson and G. A. Daigle, ''Electronic system for the measurement of flow resistance,'' J. Acoust. Soc. Am. **83**, 2422–2428 (1988).
[25] J. M. Sabatier, D. C. Sokol, C. K. Fredrickson, M. J. M. Römkens, E. H. Grissinger, and J. C. Shipps, ''Probe microphone instrumentation for determining soil physical properties: testing in model porous materials,'' Soil Tech. **8**, 259–274 (1996).

Point load wave excitation in multi-layered solids: Experiments and model verification

Chunnan Zhou
Center for Quality Engineering and Failure Prevention, Northwestern University, Evanston, Illinois 60208

John S. Popovics
Department of Civil and Architectural Engineering, Drexel University, Philadelphia, Pennsylvania 19104

(Received 28 August 2000; revised 28 December 2000; accepted 14 February 2001)

Studies of transient wave propagation in a single plate, a two-layer plate structure, and a two-layer-on-a-half-space structure are reported. Experimental results are used to verify the accuracy of a new model for wave propagation in layered solids. The experiments are performed using point load excitation and point detection at the surface of the layered structure. The basis of the new wave propagation model is first introduced. Then, experiments performed on a single plate are described. Results from different wave sources, plate materials, and wave sensors are compared with each other and with computed results from the model. The transient source function for a pencil-lead-break wave source is obtained. Next, the results of experiments performed on a two-layer plate structure (aluminum bonded to stainless steel) are reported. The cases of stainless steel atop aluminum and aluminum atop stainless steel are considered. The effects of the second layer on the transient displacement at the top surface are discussed. Finally, a two-layer-on-a-half-space structure consisting of a stainless steel plate as the top layer, an aluminum plate as the second layer and a thick acrylic resin block as the half-space is studied. The obtained experimental transient measurements agree with the computed results for all cases, thus verifying the accuracy of the model. [DOI: 10.1121/1.1367244]

PACS numbers: 43.20.Gp, 43.58.Ta [DEC]

I. INTRODUCTION

A new model for wave propagation in layered solids owing to point source excitation has been developed.[1] The work is motivated by interest in the nondestructive evaluation (NDE) of multi-layered structures by means of mechanical waves. On a large scale, an example is a concrete airport pavement (typically consisting of two layers, each with distinct elastic constants and thickness, overlaying a half-space) where the waves are generated by a controlled nondamaging impact event. On a very small scale an example is a structure of thin films deposited on a body (to improve the hardness and the resistance to wear) where the waves are generated by a laser source. Much work has been carried out toward understanding the response of multi-layered structures to localized sources by researchers in seismology.[2–4] Similar work has been carried out to consider guided ultrasonic waves in single isotropic and anisotropic plates for NDE purposes.[5] In particular matrix techniques, which provide the far-field characteristics of plane wave propagation in arbitrary multi-layered structures, have been used effectively.[6] However, the new model is unique in that it uses an exact formulation to directly compute the transient near-field response (within a radius of ten top layer thicknesses) at a point owing to point source excitation in a structure consisting of two layers on a half-space. The model accounts for all possible modes (true and leaky) and is applicable to all layer thicknesses and wave frequencies. The model thereby enables evaluation of different nondestructive inspection strategies and provides an enhanced understanding of the propagation of mechanical waves in the described layered structures.

A summary of the new model is presented now. The assumed structure of the model is shown in Fig. 1. An isotropic homogeneous layer of thickness $2h$ lies atop an isotropic homogeneous layer of thickness $2h_2$, overlaying an isotropic homogeneous half-space. Perfect continuity of tractions and displacements across all layer interfaces are assumed. The origin of the Cartesian coordinate system is at mid-depth of the top layer. A point force source f_j of step function time dependence is located within the top layer at point $(0,0,z_0)$.

The model assumes an infinitesimal stress-displacement relation, and introduces the Heaviside Green's function $\mathbf{G}$ (a 3×3 matrix with components G_{ij}) defined by $u_i = G_{ij} f_j$, where u_i is the displacement. The expression applies to the top layer, the second layer and the half-space. Generalized ray theory[7] and Willis' inversion method[8] are used to model the transient waves. The boundary conditions are defined in terms of the Green's function. Reflection and transmission coefficients between each interface are obtained, and a counting scheme that monitors all rays arriving at the detector at the same time is formulated. The final output is a time domain dynamic Green's tensor for a point in the top layer. The model, integration technique, and the computation procedure are in a form suitable for parametric computation; a computer program based on the theory has been developed. The thickness, density, and wave propagation velocities (both longitudinal and transverse waves) of each layer and the locations of the source and receiving points are needed as input to the model. The output of the model is provided in two separate files. In the first output file, the transient displacement (time domain signal) within a given time range

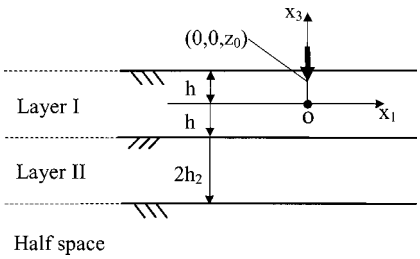


FIG. 1. Assumed two-layer-on-a-half-space structure. A point load is applied in layer I at $x_3 = z_0$.

owing to a perfect step load forcing function is given. The behavior of transient waves for arbitrary extended source distributions in space and time can then be obtained from this response by integration and convolution, respectively. The second output file lists the arrival time for each individual defined wave ray between the source and receiver; the model provides arrival sequences, arrival times and the type (direct wave, reflected longitudinal, or transverse body wave or head wave) of each possible ray. The output from the program accounts for all possible expected wave arrivals in the two-layer-on-a-half-space structure. The detailed derivation of the model is published.[1]

A preliminary verification of the model has been carried out by theoretical check cases, including a reciprocity check and some simplified case checks.[1] The emphasis of this paper is experimental verification of the model. Several layers and test configurations are considered.

II. SINGLE PLATE EXPERIMENTS

When both the second layer and the half-space are assumed to be vacuum, the model simplifies to the single plate case. In the model, vacuum is simulated by setting the wave propagation velocities and densities of those layers to be very small numbers. The output of the model is compared with experimental data obtained from single free plates. In the experiments, a wave source, either ball drop or pencil-lead-break excitation, is applied to a point on the surface of a plate specimen. The specimens are 305 mm×305 mm plates with different thickness, and are comprised of different materials. Detail about the specimens is given in Table I, where V_L is the longitudinal wave velocity and V_T the transverse wave velocity. V_L and V_T of each plate were obtained experimentally from ultrasonic pulse echo measurements using 5-MHz contact transducers. The results were compared to standard velocities for each material[9] and the experimentally obtained velocities compared favorably with those published. Different waves, including body waves and surface waves, are generated upon the application of the source. The wave motion at another point on the same surface (same-side configuration) is monitored with a sensor. The signal from the sensor is sent to a signal conditioner (amplifier) before being digitized by an oscilloscope. The signal data are sent to a personal computer via a GPIB interface, and then processed by a computer program within the Labwindows environment.

Preliminary experiments were carried out on aluminum and stainless steel plates, where surface-mounted miniature accelerometers sensed the generated waves. Two different wave sources were investigated: the impact of 2-mm and 5-mm diameter solid stainless steel spheres dropped onto the plate and a pencil-lead-break event. A sphere impacting on a half-space is known to generate a transient force input (source function) approximately as a half-cycle sine pulse.[10] For the pencil-lead-break source, 0.5-mm and 0.3-mm leads with nominal hardness of 2H, H, HB, and B were used. A force is applied on the pencil lead, where the lead tip is in contact with the specimen surface. The force is increased gradually until the lead suddenly breaks and waves are generated in the plate. The pencil-lead-break source is known to generate a source function approximately as an unloading step function.[11] Since the direct output of the model is the displacement response for an ideal step function point load, differentiation of the output with respect to time gives the displacement response for a delta function input source. The theoretical output for either the sphere drop or the pencil lead source can be calculated by convolving the time differentiated program output with the assumed source function.

In the preliminary experiments miniature contact accelerometers, with a linear response range up to about 75 kHz, were used as point sensors. In order to compare the experimental output with that predicted by the model, the accelerometer response must be doubly integrated with respect to time or the model output doubly differentiated. The experimental results with both the sphere drop and pencil-lead-break sources were not similar to the computed results. Furthermore, experimental conditions were nominally identical impact event to impact event, yet the resulting repeated signals were inconsistent. Because of the observed experimental difficulties, sphere drop sources were not used in subsequent experiments. On the other hand, the signals from repeated lead breaks, particularly using the 0.3-mm 2H lead, were very consistent. However, the frequency content of the waves generated by the 0.3-mm 2H pencil-lead source was much broader (0 to 1 MHz) than the linear response range of the accelerometer (0 to 75 kHz). In subsequent experiments, the 0.3-mm 2H lead break source only was used and more appropriate wave sensing techniques investigated.

TABLE I. Properties of specimens.

Label	Material	Thickness (mm)	V_L (m/s)	V_T (m/s)	Density (g/cm^3)
A1	aluminum	12.7	6410	3170	2.77
A2	aluminum	9.7	6410	3170	2.77
S1	stainless steel	12.7	5780	3150	8.03
S2	stainless steel	8.0	5780	3150	8.03
P1	acrylic resin	52	2730	1430	1.15

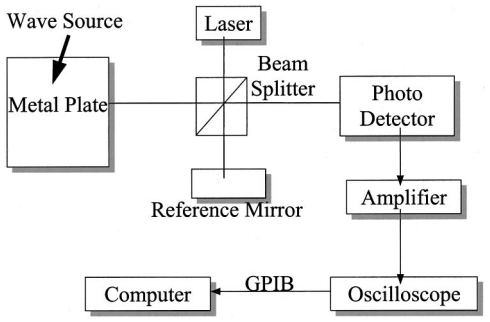


FIG. 2. Block diagram of the experimental setup for the laser interferometer.

A. Experiments using a laser interferometer

Since conventional sensors were deemed insufficient, a broad bandwidth detection system was tried. Specifically, a stabilized Michelson interferometer with a 10-mW HeNe laser was used. The bandwidth of the detection system is about 20 MHz, and the dynamic response of laser interferometers is linear over all the bandwidth.[12] An illustration of the test setup is shown in Fig. 2. A light beam is generated by the laser and sent to a beam splitter, which splits the light into two equal beams. One of the split beams is sent to the surface of the test plate. The light beam impinges normally on the surface of the plate, reflects back, and travels through the beam splitter to the photodetector. The other light beam transmits through the beam splitter and impinges normally on a reference mirror. This light beam reflects back to the beam splitter and is sent to the photodetector. The two beams, one reflected from the metal plate and the other from the reference mirror, interfere with each other at the photodetector. This interference, which is transformed to an electrical signal by the photodetector, is directly related to the absolute out-of-plane displacement at a point on the surface of the test plate. A very reflective (polished) specimen surface is required for appropriate optical sensitivity with this set up. Tests performed on the untreated metal plates yielded no usable data. Aluminum and a stainless steel plates (specimens A1 and S1) were polished and then used in the laser tests. Experiments were performed using the same side configuration with source-detector spacing ranging from 5 mm to 25 mm.

The results obtained from the aluminum plate are better than the ones for the stainless steel plate because the polished aluminum is more reflective, but all the experimental results show good agreement with predictions computed by the model. However, the experimentally obtained signals contain a significant amount of noise. A typical response for steel plate S1 with a source–receiver spacing of 10 mm is shown as a semi-dashed line in Fig. 3. The computed prediction is plotted in the same figure as a solid line. In the figure, the results are shifted with respect to time so that the large negative peaks align to the same point; the large negative peak in the signal is known to coincide with the arrival of the Rayleigh surface wave. This shift was necessary since the time of application of the point source could not be determined accurately in the experimental signal when lead breaks were used. In addition, amplitudes of the theoretical and experimental results are normalized with respect to their value 4 μs after the Rayleigh wave arrival. This was necessary since the amplitude of lead break sources could not be measured and are known to vary among repeated nominally identical applications although the overall shape of the forcing function remains consistent.[11] The results in Fig. 3 show that the experimental and computed curves have somewhat similar shape except at times close to the arrival of the Rayleigh wave. The computed response assumes a source function in the form of an ideal step function. The pencil-lead source is similar to but not identical to an ideal step function, however, so some difference between the experimental and theoretical shapes is expected.

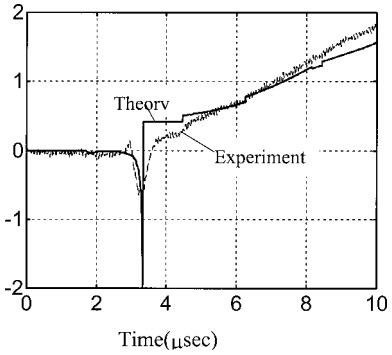


FIG. 3. Comparison of displacement measurements from a stainless steel plate detected with a laser interferometer and the computed prediction. The signals have been shifted and normalized. The source-detector spacing is 10 mm.

The individual wave arrivals in the experimental curve are more clearly seen after application of a 2-MHz low pass digital Butterworth filter, which removes the high-frequency noise from the signal. The experimental data are shown in Fig. 4 together with similarly filtered computed data. The calculated ray arrival times show that point 1 in the figure corresponds to the arrival of a LL wave, a longitudinal wave that emanates from the source, propagates through the first layer and reflects at the bottom (first interface) as a longitudinal wave and propagates back through the first layer to the sensor. Point 2 corresponds to the arrival of LT and TL waves, which are waves that are reflected once at the bottom and convert to the other body wave mode after reflection. The agreement at points 1 and 2 between the theoretical and experimental data is good, but part of the experimental response (indicated as point 3) is not predicted by the model.

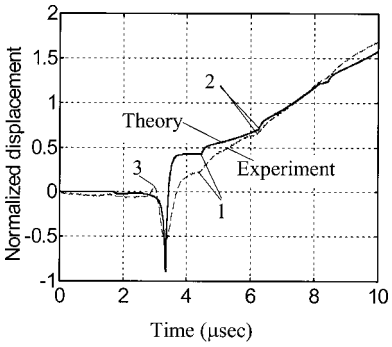


FIG. 4. Replot of Fig. 3 after application of a 2-MHz low pass filter, shifting, and normalization.

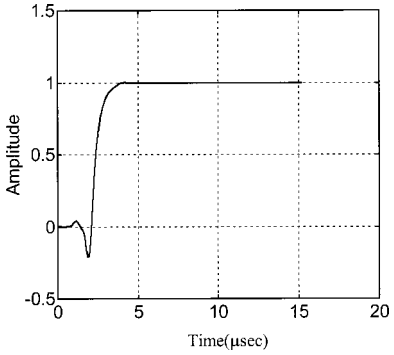


FIG. 5. The forcing function for a pencil-lead-break on an aluminum plate.

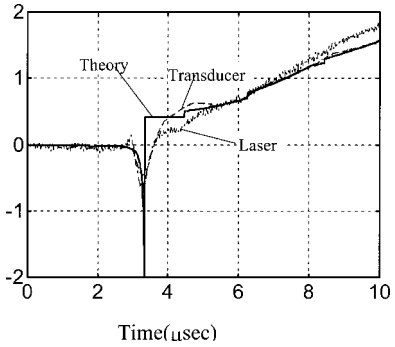


FIG. 6. Comparison between experimentally obtained results using the laser interferometer, the displacement transducer, and the computed prediction. The specimen is a 12.7-mm-thick stainless steel plate. Source–receiver spacing is 10 mm.

Since the laser interferometer produces a linear response across a broad range of frequencies, we are able to obtain experimental data and then use deconvolution to back-calculate the source function that generated the waves. This obtained source function may then be used in the model for subsequent computations. To perform the deconvolution, all wave arrivals after the Rayleigh wave arrival were removed from both the computed and experimental 2-MHz low pass filtered signals. Then, zeros were added to the end of the shortened signals so that the peak corresponding to the Rayleigh wave arrival was centered in the time domain signal. Next, the derivative with respect to time of both the computed and experimental data were determined. The Fourier transform of the derivative of the experimental signal was then divided by that of the computed signal. The inverse Fourier transform of the complex-valued quotient was then computed to obtain a time-dependent function that represents the derivative of the source function. By integration, the source function was obtained.

The shape of the source function for the 2H pencil-lead-break on aluminum, as obtained by the described deconvolution procedure, is plotted in Fig. 5. The source function for the 2H pencil lead break on steel yields a nearly identical response. The vertical axis is normalized so that the amplitude of the step is unity, the horizontal axis is the actual time. The figure shows there are some ripples before the arrival of the sharp step in the source function. Thus the pencil-lead-break does not provide an ideal step function source, and this may cause the shape differences between the computed and experimental signals, for example the occurrence of point 3 in Fig. 4.

Since the use of the laser interferometer requires the specimen surface to be highly reflective, the apparatus is not portable, and the obtained signals contained disturbing high-frequency noise, an alternative sensor was investigated to replace the laser interferometer. The following sub-section describes experiments on the single plate using a contact displacement transducer.

B. Experiments using a displacement transducer

A conical displacement point contact transducer with a broad frequency response was investigated. The output voltage of the transducer is closely related to the displacement normal to the surface of the specimen. This type of sensor has been used for acoustic emission and ultrasonic experiments.[13] The transducer consists of a specially designed brass backing to which a PZT5 conical element is attached. The transducer is a three-point device. Two leveling screws and the active element make up the three legs. The transducer requires that the return electrical path to the amplifier be supplied. This is easily satisfied in our experiments because the transducer sits on an electrically conductive metal plate. The nominal response range of the transducer is 10 kHz to 1.2 MHz, which covers the typical frequency range of the pencil-lead-break source.[13,14] To obtain optimal coupling, petroleum jelly was placed on the transducer surface that contacts the plate.

The results obtained with the displacement transducer show good agreement with the computed predictions for both aluminum and stainless steel plates. The responses from the transducer and that from the laser interferometer obtained from steel plate S1 are plotted together along with the theoretical response in Fig. 6. The source–receiver spacing is 10 mm. As before, the horizontal axis is time and the vertical axis is the displacement normalized with respect to point 4 μs after the Rayleigh wave arrival. In the figure, the solid line is the response calculated by the model assuming an ideal step forcing function, the dashed line the response measured by the transducer, and the dotted line the response measured by the laser interferometer. The signal-to-noise ratio of the signal from the laser interferometer is lower than that for the conical transducer, which illustrates the excellent sensitivity of the transducer. In fact, the amplitude of the experimental data from the transducer is approximately ten times that from the laser interferometer. The good agreement between the transducer and the laser responses further illustrates the sufficiently broad frequency range of the transducer.

The displacement transducer is sensitive and broadband, allowing direct comparison between experimental and theoretical data. The transducer is portable and can be used on almost any surface without surface preparation. Thus the displacement transducer will be used in subsequent experiments in place of the laser interferometer.

III. TWO-LAYER PLATE EXPERIMENTS

When the half-space is assumed to be vacuum, the model simplifies to the two-layer case. In the model, vacuum

is simulated by setting the wave velocities and density to be very small numbers. In this section, the model is experimentally verified for the two-layer case. Aluminum plate A2 and steel plate S2 were bonded together for the experiments. In order to simulate the welded boundary conditions assumed by the model, the plates were bonded with a high-strength epoxy (Sikadur 32 Hi-Mod). To prepare the metal plates for bonding, all bond surfaces were sanded to a white metal finish and then cleaned. A thin film of epoxy was applied to the bond surfaces, as specified by the epoxy manufacturer. The plates were then aligned and compressed with a 100-kg dead-load across the entire surface. The plates were compressed for a period of 24 h, while the epoxy hardened fully. The thickness of the resulting epoxy bond layer may be measured from the difference of individual plate thickness from the bonded plate structure thickness. These thicknesses were measured with a set of calipers at several locations. No difference in the values was found, so the actual bond layer must on average be less than 0.05 mm, which is the precision provided by the calipers.

In order to assess the bond condition at the interface between the plates, an ultrasonic C-scan image of the interface was generated. The C-scan apparatus consists of a stepper motor controller, a pulse generator, a focused immersion transducer (10 MHz, 13 mm diameter with 25-mm focal length), a signal amplifier and a computer. The data collection was performed using a commercially available system.[15] The data were collected in the pulse echo configuration, with the transducer and the specimen placed in a water bath. By analyzing the obtained C-scan image, sizable regions of continuous good bonding were found.[16] Experiments were performed over these well-bonded regions.

In the experiments, a 0.3-mm, 2H pencil-lead-break was used to generate the waves and the displacement transducer was used to receive the waves. The same-side configuration was used with varying source–receiver spacing.

A. Experimental results: Stainless steel atop aluminum

The case of the stainless steel plate atop aluminum plate is considered first. The experimental results show good agreement with those given by the theoretical model. As an example, the theoretical and experimental results are shown in Fig. 7 for a source–receiver spacing of 20 mm. The computed prediction in Fig. 7 uses the pencil-lead-break source function from aluminum described in the previous section. Since the source lead break functions from aluminum and steel are nearly identical, only one source function (from aluminum) is used in the subsequent computations. In Fig. 7 the horizontal axis is the actual time in microseconds, where the signals are shifted to align with the respective Rayleigh wave arrivals. The vertical axis is the displacement normalized with respect to the point 4 μs after the Rayleigh wave arrival. The solid line shows the experimental results while the dashed line shows the computed results. The theoretical and experimental results agree closely, as shown from this figure. Thus the model gives good results for the two-layer case when stainless steel is atop aluminum.

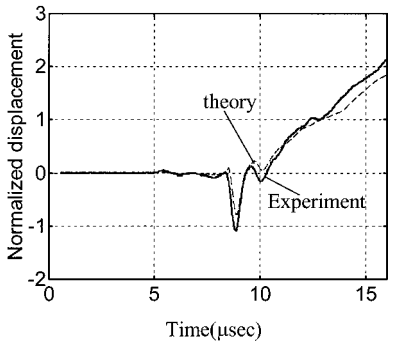


FIG. 7. Comparison of computed and experimental two-layer signals for the case of steel atop aluminum. The computed signals were obtained by convolving the output of the model with the pencil-lead-break source function. The source-detector spacing is 20 mm.

The model defines altogether 70 rays that propagate between the detector and the source within the time range presented in Fig. 7. Note that there may be several different combinations of rays arriving at the detector at the same time. To illustrate the effect of the bonded second layer, the theoretical prediction for the two-layer model and that for the single top plate (steel) are presented together in Fig. 8 for a source–receiver spacing of 20 mm. In order to clearly show the individual wave arrivals, the computed results in Fig. 8 assume the ideal step function input source. The horizontal axis is the time in microseconds and the vertical axis is the computed displacement in mm. The solid line represents the displacement for the two-layer case and the dashed line for the single plate case. We can see that the two curves are similar but not identical. The difference is due to the existence of the acoustically softer second layer, the aluminum plate, which contributes some additional reflections to the signal.

To see in more detail how the second layer affects the displacements, Fig. 8 is replotted to show the individual arrivals of waves that travel through the second layer in Fig. 9. We note that in the time range of 9–10 μs, there are three obvious arrivals in the steel atop aluminum configuration that do not exist in the single steel plate (marked as 1, 2, and 3 in Fig. 9). By checking the wave ray arrival times provided by the model, one can determine that point 1 corresponds to the $L_{\mathrm{I}}L_{\mathrm{II}}L_{\mathrm{II}}L_{\mathrm{I}}L_{\mathrm{I}}L_{\mathrm{I}}$ and $L_{\mathrm{I}}L_{\mathrm{I}}L_{\mathrm{I}}L_{\mathrm{II}}L_{\mathrm{II}}L_{\mathrm{I}}$ wave arrivals (arrival time is 9.178 μs). The travel path of the $L_{\mathrm{I}}L_{\mathrm{II}}L_{\mathrm{II}}L_{\mathrm{I}}L_{\mathrm{I}}L_{\mathrm{I}}$ ray is that a longitudinal wave emanating from the source, trans-

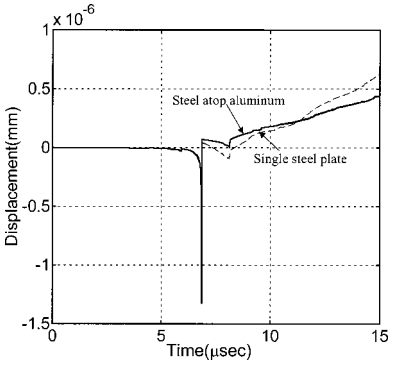


FIG. 8. Comparison of the computed results for the case of steel atop aluminum (two-layer plate) with that of the single steel plate.

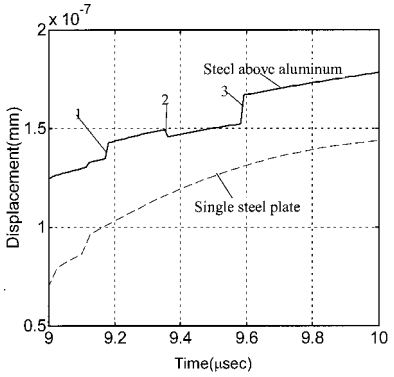


FIG. 9. Re-plot of Fig. 8 in the time range 9 μs to 10 μs.

mitted into the second layer as a longitudinal wave, reflected as a longitudinal wave, transmitted into the top layer and reflected at the top surface, reflected again at the interface between the first and the second layer and received by the receiver as a longitudinal wave. The travel path of $L_{\mathrm{I}}L_{\mathrm{I}}L_{\mathrm{I}}L_{\mathrm{II}}L_{\mathrm{II}}L_{\mathrm{I}}$ is similarly defined, where the subscript I indicates propagation in the top layer and subscript II the second layer. Point 2 coincides with the arrival of the $L_{\mathrm{I}}L_{\mathrm{II}}L_{\mathrm{II}}L_{\mathrm{II}}L_{\mathrm{II}}L_{\mathrm{I}}$ wave (arrival time is 9.35 μs), and point 3 the coincidental arrivals of the $L_{\mathrm{I}}L_{\mathrm{II}}T_{\mathrm{II}}T_{\mathrm{I}}$, $L_{\mathrm{I}}T_{\mathrm{II}}L_{\mathrm{II}}T_{\mathrm{I}}$, $T_{\mathrm{I}}L_{\mathrm{II}}T_{\mathrm{II}}L_{\mathrm{I}}$, and $T_{\mathrm{I}}T_{\mathrm{II}}L_{\mathrm{II}}L_{\mathrm{I}}$ rays (arrival time of 9.58 μs), where T represents transverse waves. These wave paths travel through the second layer, but waves that propagate only in the first layer are also affected owing to the change of the reflection coefficients between the top layer and the second layer. Nevertheless, the responses are very similar and differences between the two cases are not readily seen when the convolved theoretical or experimental responses are directly compared. Thus the acoustically softer second layer has a small, and possibly insignificant from a practical point of view, effect on the response within the considered time range.

B. Experimental results: Aluminum atop stainless steel

Experimental data from the two-layer case when the aluminum plate is atop the stainless steel plate are considered now. The obtained results show that the agreement between the experimental and theoretical results is not as good as that for the case of steel atop aluminum. Nevertheless, the agreement between theory and experiment is still reasonably good. Figure 10 shows the comparison between the theoretical prediction using the pencil-lead-break source function and the experimental result.

In Fig. 10, the horizontal axis is the time in microseconds, the vertical axis is the displacement normalized with respect to the point 4.5 μs after the Rayleigh wave arrival. The solid line shows the experimental data and the dashed line shows the computed results. The experimental results are shifted with respect to the theoretical results so that the peaks corresponding to the Rayleigh wave arrivals align with each other. Since the agreement between the experimental and theoretical results is reasonably good, the model appears to give good results for the two-layer case.

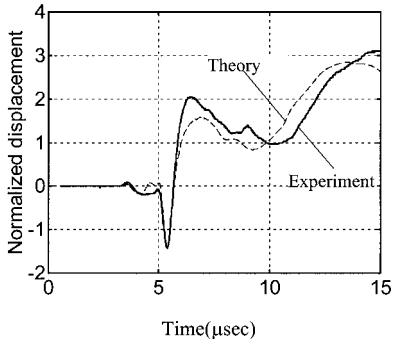


FIG. 10. Comparison of computed and experimental two-layer signals for the case of aluminum atop steel. The computed signals were obtained by convolving the output of the model with the pencil-lead-break source function. The source-detector spacing is 10 mm.

The effects of the bonded second layer to the overall wave propagation behavior can be seen in Fig. 11, the comparison of the theoretical predictions for the single aluminum plate when the source–receiver distance is 10 mm and that of the aluminum atop steel two-layer structure. For clarity, both theoretical predictions assume an ideal step function input source. In the figure, the horizontal axis is the time in microseconds and the vertical axis is computed displacement in mm. The solid line represents the two-layer aluminum atop stainless steel structure and the dashed line the single aluminum plate. The two curves are very different, clearly showing the significant contribution from an acoustically stiffer underlying layer. Unlike the case of steel atop aluminum considered previously, significant differences are seen when the theoretically predicted responses of the single steel plate and aluminum atop steel cases are directly compared.

IV. TWO-LAYER-ON-A-HALF-SPACE EXPERIMENTS

The full capacity of the model is verified by experiments on a two-layer-on-a-half-space specimen. A thick acrylic resin plate was used to simulate the half-space. The acrylic plate can be considered to be a half-space if the time required for a longitudinal wave to travel across the thickness of the plate and back after reflection is more than the total sampling time. In this case, the time for an L-wave to travel through

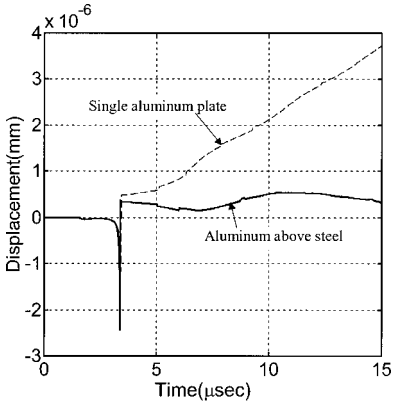


FIG. 11. Comparison of the theoretical prediction for aluminum atop steel (two-layer) with that for the single aluminum plate.

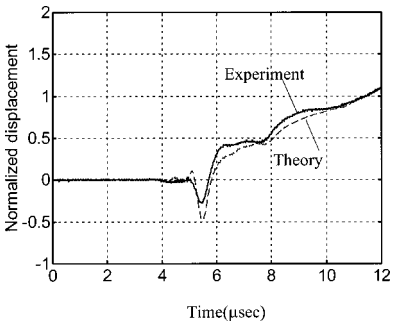


FIG. 12. Comparison between the computed prediction (pencil-lead-break source function) and experimentally obtained signal for the two-layer-on-a-half-space structure. The source–receiver spacing is 10 mm.

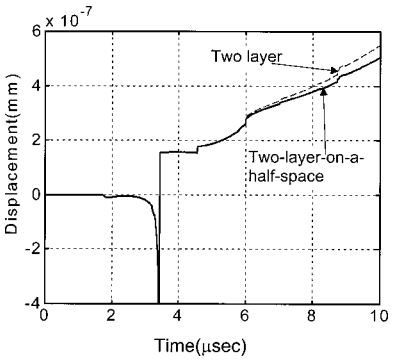


FIG. 13. Comparison between the computed displacement for the stainless steel, aluminum and acrylic resin two-layer-on-a-half-space structure and the corresponding displacement for the two-layer structure of stainless steel atop aluminum. Source-receiver spacing is 10 mm. The vertical axis is cut so that the displacement range is -4×10^{-7} to 6×10^{-7} mm.

the plane thickness and back is 67.7 μs while the maximum sampling time is 20 μs. Thus the acrylic plane does indeed act as a half-space for these experiments.

The two-layer structure used in the previous experiments, plate A2 bonded to plate S2, was bonded to the acrylic resin plate using the high-strength epoxy. The two-layer on a half-space structure is stainless steel on top, aluminum in the middle, and acrylic resin as the half-space. All bond thicknesses were less than 0.05 mm.

The conditions of the bonds were evaluated with an ultrasonic pulse echo C-scan test, similar to that done earlier. To obtain improved imaging of the aluminum-acrylic interface the structure was ultrasonically insonified through the acrylic resin, wherein a 2.25-MHz, 13-mm diameter focussed transducer was used, with a focal length of 37 mm. This lower frequency allows the ultrasound to pass through the thick and attenuating acrylic resin plate. Inspection of the C-scan images for the interfaces between the stainless steel and aluminum and that between the aluminum and the acrylic resin indicates a common region of good bonding for both interfaces.[16] Subsequent experiments were performed over this region. The 2H pencil-lead-break source was applied to the surface of the top stainless steel plate. The displacement transducer was placed on the same surface of the plate, separated by a certain distance from the source.

In all cases, the experimental results are similar to those predicted by the model. The experimental results and the associated computed prediction are shown in Fig. 12 for a source–receiver spacing of 10 mm. The solid line is the experimental measurement and the dashed line the computed prediction using the obtained pencil-lead-break source function. Both the theoretical and the experimental results have been normalized with respect to the displacement 6 μs after the Rayleigh wave arrival. The experimental results are shifted with respect to the theoretical results so that the peaks corresponding to the Rayleigh wave arrivals align with each other. From Fig. 12 we can see that the theoretical and experimental results are in excellent agreement, thus verifying the accuracy of the model.

The model indicates that 23 distinct wave ray arrivals contribute to the displayed signal, including one head wave ($T_{\mathrm{I}}L_{\mathrm{II}}T_{\mathrm{I}}$) and ten rays that involve body waves reflecting from the aluminum-acrylic interface. However, the amplitude of the other wave arrivals are dwarfed by that of the Rayleigh wave. Interestingly, there are no head waves that pass along the interface with the half-space. This is because the longitudinal wave velocity of the acrylic resin is less than the transverse velocity of the stainless steel plate, and that condition precludes the existence of head waves.[17]

To see the contribution of the half-space to the transient motion, the computed displacement data for the two-layer-on-a-half-space structure and the data obtained for the two-layer structure of the steel atop aluminum are presented in Fig. 13. In this figure, the solid line is the computed prediction for the two-layer-on-a-half-space structure and the dashed line that for the two-layer structure without the half-space. The vertical axis is modified so that the displayed displacement range is between $(-4)\times10^{-7}$ and 6×10^{-7} mm. We can see that there is no difference between the two results until about 6 μs. By checking the ray arrival time, it is determined that this time corresponds to the arrival of the $L_{\mathrm{I}}L_{\mathrm{II}}L_{\mathrm{II}}L_{\mathrm{I}}$ wave. This wave is the first wave reflected by the interface between the second layer and the half-space, and thus the first ray affected by the acrylic resin half-space. Similar results are obtained for other source–receiver spacings.

V. SUMMARY

After a brief introduction to a new model for wave propagation in layered structures, experimental results from a single plate were reported. Measurements were made with several wave sources and sensors. The source function for a pencil-lead-break wave source was obtained and conclusions concerning the performance of each sensor were presented. Experiments were then performed on a two-layer plate specimen, where the cases of stainless steel atop aluminum and aluminum atop stainless steel were investigated. The experimental results showed good agreement with that predicted by the model for the case of steel atop the aluminum. In the case of aluminum atop the steel, the agreement between theory and experiment was reasonable but not as good. Measurements were also carried out on a two-layers-on-a-half-space structure, consisting of a stainless steel plate as the top layer, an aluminum plate as the second layer, and an acrylic resin block as the half-space. A region of good bonding between all interfaces was determined from ultrasonic C-scan tests.

Point source wave propagation experiments were performed over this region of good bonding. The transient displacement measurements showed good agreement with that predicted by the model, verifying the accuracy of the model for all layer configurations. Selected characteristics of wave propagation behavior in layered structures were illustrated using computed results from the model.

ACKNOWLEDGMENTS

This paper was prepared in the course of research sponsored by the Office of Naval Research under Contract No. N00014-89-J-1362. Helpful comments by J. D. Achenbach and assistance with the laser setup from Dr. Todd Murray are gratefully acknowledged.

[1] C. Zhou, N. N. Hsu, J. S. Popovics, and J. D. Achenbach, ''Response to a suddenly applied point force of two layers overlaying a half-space,'' Wave Motion **31**, 255–272 (2000).
[2] W. M. Ewing, W. S. Jardetzky, and F. Press, *Elastic Waves in Layered Media* (McGraw-Hill, New York, 1957).
[3] J. H. Rosenblum, ''The long-time response of a layered elastic medium to explosive sound,'' J. Geophys. Res. **65**, 2071–2084 (1960).
[4] F. Gilbert, ''Propagation of transient leaking modes in a stratified elastic wave guide,'' Rev. Geophys. **2**, 123–153 (1964).
[5] D. E. Chimenti, ''Guided waves in plates and their use in materials characterization,'' Appl. Mech. Rev. **50**, 247–284 (1997).
[6] M. J. S. Lowe, ''Matrix techniques for modeling ultrasonic waves in multilayered media,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**, 525–542 (1995).
[7] Y. H. Pao and R. R. Gajewski, ''The generalized ray theory and transient response of layered elastic solids,'' in *Physical Acoustics Volume XIII*, edited by R. N. Thurston (Academic, New York, 1977), Chap. 6.
[8] J. R. Willis, ''Self-similar problems in elastodynamics,'' Philos. Trans. R. Soc. London, Ser. A **274**, 435–491 (1973).
[9] J. Krautkramer and H. Krautkramer, *Ultrasonic Testing of Materials*, 4th ed. (Springer, Berlin, 1990).
[10] W. Goldsmith, *Impact: The Theory and Physical Behavior of Colliding Solids* (Edward Arnold, London, 1960), pp. 82–91.
[11] F. R. Breckenridge, T. M. Proctor, N. N. Hsu, S. E. Fick, and D. G. Eitzen, ''Transient sources for acoustic emission work,'' in *Progress in Acoustic Emission V, Proceedings of the 10th International Acoustic Emission Symposium*, Sendai, Japan, 1990.
[12] J. W. Wagner, ''Optical detection of ultrasound,'' in *Physical Acoustics Volume XIX*, edited by E. Pierce (Academic, New York, 1990), pp. 201–266.
[13] T. M. Proctor *et al.*, ''More recent improvement on the NBS conical transducer,'' J. Acoust. Emiss. **5**, 134–142 (1986).
[14] F. R. Breckenridge, ''Acoustic emission transducer calibration by means of a seismic surface pulse,'' J. Acoust. Emiss. **1**, 87–94 (1982).
[15] The hardware and software (FLEXScan-C) used in the C-scan measurements are products of Sonix, Inc., Springfield, Virginia.
[16] C. Zhou, ''Point load excitation of a multi-layered structure: ray theory and experimental verification,'' Ph.D. Dissertation, Northwestern University, Evanston, IL (1999).
[17] V. Cerveny and R. Ravindra, *Theory of Seismic Head Waves* (University of Toronto Press, Toronto, 1970).

Time-frequency representations of Lamb waves

Marc Niethammer and Laurence J. Jacobs[a)]
School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0355

Jianmin Qu and Jacek Jarzynski
G. W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0405

(Received 26 July 2000; revised 20 December 2000; accepted 21 January 2001)

The objective of this study is to establish the effectiveness of four different time-frequency representations (TFRs)—the reassigned spectrogram, the reassigned scalogram, the smoothed Wigner–Ville distribution, and the Hilbert spectrum—by comparing their ability to resolve the dispersion relationships for Lamb waves generated and detected with optical techniques. This paper illustrates the utility of using TFRs to quantitatively resolve changes in the frequency content of these nonstationary signals, as a function of time. While each technique has certain strengths and weaknesses, the reassigned spectrogram appears to be the best choice to characterize multimode Lamb waves. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1357813]

PACS numbers: 43.20.Mv, 43.35.Cg [DEC]

I. INTRODUCTION

There have been significant advances in the field of signal processing since the development of the fast Fourier transform (FFT) in the mid 1960's. Current research is primarily concentrated on applications for microelectronics and telecommunications, although most signal-processing techniques are general enough in nature to have potential applications in a variety of fields. Recent work in the area of time-frequency representations (TFRs) such as the spectrogram, the scalogram, the Wigner–Ville distribution, and the Hilbert spectrum shows great promise for applications in nondestructive evaluation. Of particular interest (and importance) is the use of TFRs to interpret ultrasonic guided waves. This class of ultrasonic signals can be extremely complicated, exhibiting dispersion and containing multiple modes.

This paper compares the effectiveness of four candidate TFRs to characterize Lamb waves—guided ultrasonic waves that propagate in plates. Lamb waves have received extensive attention since the study by Mindlin.[1] Recent experimental work has shown that it is possible to obtain a plate's dispersion relationship by using the two-dimensional Fourier transform (2D-FT) to operate on multiple, equally spaced waveforms.[2,3] Unfortunately, the need for exact, spatially sampled data restricts the practicality of the 2D-FT for some inspection applications. In contrast, TFRs require only a single signal. Recently, Prosser *et al.*[4] used the smoothed Wigner–Ville distribution to determine the Lamb modes of numerically simulated waveforms in an aluminum plate. They also consider real experimental data for a composite plate and identify the s_0 and the a_0 Lamb modes for frequencies below 500 kHz. Hayashi *et al.*[5] determined the thickness and the elastic properties of thin metallic foils (thickness of less than 40 μm) by calculating the group velocity of a single mode (the a_0 up to 3.5 MHz) using the wavelet transform of laser-generated and -detected Lamb waves. Holland *et al.*[6] used a reassigned, smoothed Wigner–Ville distribution to examine synthetic waves in a 10-mm-thick steel plate with excellent results. Niethammer *et al.*[7] showed that the reassigned spectrogram is capable of distinguishing multiple, closely spaced Lamb modes in the ultrasonic frequency range.

The current study extends[7] by evaluating the suitability of the reassigned spectrogram, the reassigned scalogram, the smoothed Wigner–Ville distribution, and the Hilbert spectrum to extract the dispersion curves of a plate from a single time-domain signal. The objective of this paper is to establish the robustness of this collection of TFRs by comparing their ability to resolve the dispersion relationships for Lamb waves generated and detected with optical techniques (laser source and interferometric detector).

II. BRIEF REVIEW OF TFRs

A transient time-domain signal, together with its Fourier transformed spectrum, does not provide enough information for applications that require an understanding of how a signal's frequency changes *as a function of time*. Note that the Fourier transform is essentially limited to stationary signals—signals that have the same frequency content for all times. In contrast, nonstationary signals require signal-processing methods that can quantitatively resolve changes in frequency content, as a function of time. A large number of TFRs have been developed to analyze nonstationary signals, many of which are subsumed in the general framework of Cohen's class.[8]

This section provides a brief review of TFRs to furnish a common foundation and assist in understanding why certain TFRs are more effective in this application—it is not intended to be a comprehensive review, and references that contain technical details and mathematical derivations are provided for an interested reader.

[a)]Author to whom correspondence should be addressed. Electronic mail: laurence.jacobs@ce.gatech.edu

The short-time Fourier transform (STFT) divides a time-domain signal into a series of small overlapping pieces; each of these pieces is windowed and then individually Fourier transformed.[8] The STFT of a function $s(t)$ is defined as

$$S_1(\omega,t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-i\omega\tau}s(\tau)h(\tau-t)\mathrm{d}\tau, \tag{1}$$

where $h(t)$ is a window function. The energy density spectrum of an STFT is defined as $E_1(\omega,t)=|S_1(\omega,t)|^2$ and called a spectrogram.

Instead of a fixed window function, the wavelet transform (WT) uses time-frequency atoms or wavelets. The WT of a function $s(t)$ is given by[9]

$$S_2(a,b)=\frac{1}{\sqrt{a}}\int_{-\infty}^{\infty}s(t)\psi\left(\frac{t-b}{a}\right)\mathrm{d}t, \tag{2}$$

where a is the scale, and b the time-shift variable (a and b are also known as the dilation and the translation parameters, respectively). A small a corresponds to a high frequency and vice versa [a dilates the mother wavelet $\psi(t)$, while b simply shifts the wavelet with respect to time without altering the frequency content]. The energy density function of a WT is defined as $E_2(a,b)=|S_2(a,b)|^2$ and is called a scalogram.

Wavelets are derived from a mother wavelet $\psi(t)$ by dilation and translation. This study uses the Gabor wavelet

$$\psi(t)=\frac{1}{\sqrt[4]{\pi}}\sqrt{\frac{\omega_0}{\gamma}}e^{-1/2(\omega_0 t/\gamma)^2+i\omega_0 t}, \tag{3}$$

which provides an excellent compromise between time and frequency resolution, because it is based on a Gaussian envelope (which guarantees the best possible time-frequency resolution).[9,10] Note that the variable γ controls the sharpness of the Gaussian envelope in the time domain.

Since the WT decomposes a signal into wavelet components (and not into sine components like the Fourier transform), there is not a direct map from wavelet scale, a, to frequency, ω. It is possible, however, to compute the WT of a sine, and then calculate the relationship between scale and frequency by determining the value a for which a scalogram reaches its maximum.[11] Finally, if broadband signals (such as laser-generated Lamb waves) are analyzed with the WT, the higher frequencies tend to have lower energies. As a result, it is advantageous to amplify high-frequency components by multiplying the WT with $(1/\sqrt{a})$[12]—this research uses the $1/\sqrt{a}$ scaled Gabor WT.

While the STFT is the Fourier transform of small, overlapping, windowed pieces of a time signal $s(t)$, the Wigner–Ville distribution (WVD) is the Fourier transform with respect to τ of $s(t+[\tau/2])s^*(t-[\tau/2])$ [where $s^*(t)$ is the complex conjugate of $s(t)$], or[8]

$$S_3(\omega,t)=\int_{-\infty}^{\infty}s\left(t+\frac{\tau}{2}\right)s^*\left(t-\frac{\tau}{2}\right)e^{-i\omega\tau}\mathrm{d}\tau. \tag{4}$$

As a result, the WVD is a measure of the signal's local time-frequency energy. An advantage of the WVD is that it can exactly localize sines or Dirac impulses; this is not the case for the spectrogram and the scalogram. Unfortunately, this property means that the WVD is real, but not necessarily positive [the only signal class that leads to a strictly positive WVD is $s(t)=(\alpha/\pi)^{1/4}e^{-(1/2)\alpha t^2+(1/2)i\beta t^2+i\omega_0 t}$], and that it suffers from signal interference. For example, the WVD can perfectly resolve a single chirp, but a signal consisting of two chirps will be decomposed into the two chirps themselves *plus an additional interference term*.

To avoid this interference, the WVD is often replaced by the smoothed WVD

$$S_4(t,f)=\int\int_{-\infty}^{\infty}G(t-t',f-f')S_3(2\pi f',t')\mathrm{d}t'\,\mathrm{d}f', \tag{5}$$

which filters the original WVD of Eq. (4) with a two-dimensional filter, G.[8] An unfortunate side effect of this smoothing is the introduction of time-frequency smearing which nullifies the WVD's property of exact localization of sines and impulses. This study uses the Gaussian filter

$$G(t,f)=\sqrt{\frac{1}{\alpha\beta}}e^{-(t^2/\alpha)-4\pi^2(f^2/\beta)}. \tag{6}$$

The value of the $\alpha\beta$ term in Eq. (6) controls both the positivity ($\alpha\beta\geqslant 1$ guarantees positivity) and signal localization of the resulting smoothed WVD. While the choice of $\alpha\beta \geqslant 1$ will guarantee positivity, the resulting distribution becomes a spectrogram and no longer possesses the WVD's advantages of signal localization.[13] As a result, it is critical to use a filter with $\alpha\beta<1$ in applications where superior signal localization (compared to the spectrogram) is desired.

The final TFR considered in this study is based on the empirical mode decomposition (EMD), that is used to generate a set of intrinsic mode functions (IMFs). Huang *et al.*[14,15] propose a method with which a complicated time-domain signal is decomposed into a finite number of IMFs that admit well-behaved Hilbert transforms. By using the Hilbert transform, the IMFs produce instantaneous frequencies *as functions of time*, that enable the identification of a signal's embedded structure (its modes). The resulting energy-time-frequency distribution is called the Hilbert spectrum, and is comparable to the spectrogram, scalogram, or WVD. This decomposition (the IMFs) is based on local properties of the signal itself, and not an artificial ''external'' function, so the instantaneous frequencies are (ideally) physically meaningful.

The decomposition of a signal, $s(t)$, into its IMFs is accomplished with a ''sifting'' process that uses the signal itself as the basis for the decomposition. Sifting empirically identifies the signal's intrinsic oscillatory modes by their characteristic time scales, and then decomposes the signal accordingly. The sifting process starts by first determining the maxima and minima of $s(t)$, and then connecting these maxima and minima with cubic splines (spline envelopes). In order to avoid (possible) erroneous behavior, such as wide swings of the spline envelopes at the signal's endpoints (both the right and left boundaries), ''artificial boundary conditions'' must be introduced. Huang *et al.*[14,15] propose the addition of characteristic waves at both ends of the signal. The sifting process (in brief) calculates the mean value, $m(t)$, of

a point on the maxima and minima spline envelopes and subtracts it from the original signal, $s(t)$, to obtain the first IMF candidate, or $h_1(t)=s(t)-m(t)$.[10,15] If h_1 is an IMF (h_1 is an IMF if the number of extrema and the number of zero crossings differs by at most 1, and if the mean value of h_1 defined by its spline envelope is zero), it is added to the list of calculated IMFs. If it is not, this step is repeated for k siftings—with $h_1, h_{12}, \ldots, h_{1k}$ leading (after k siftings) to the IMF, $h_{1k}(t)=h_{1(k-1)}(t)-m_{1k}(t)$. Once it meets the IMF definition, h_{1k} is subtracted from the original signal, producing the residue $r_1(t)=s(t)-h_{1k}(t)$. The complete process is then repeated with r_1 replacing the original signal to calculate the next IMF. The search for IMFs is stopped when a predetermined number of IMFs are calculated, or when r_n becomes monotonic—see the next section for a demonstration of the sifting process.

Having decomposed the signal $s(t)$ into n IMF components and a residue r_n, take the Hilbert transform of every IMF component, and add the (respective) Hilbert transform to every IMF to produce an analytic signal, $s_a(t)$. Next, determine the phase ϕ by

$$\phi(t)=\arctan\left(\frac{\mathcal{I}s_a(t)}{\Re s_a(t)}\right), \tag{7}$$

and phase unwrapping. The instantaneous (angular) frequency ω is computed by differentiating the phase ϕ, with respect to time. The Hilbert spectrum is then defined as the representation of the calculated amplitudes of the analytic signals of the IMFs, as a function of time and instantaneous frequency—a time-frequency representation of $s(t)$.

TFRs suffer from the Heisenberg uncertainty principle, making it impossible to *simultaneously* have perfect resolution in both time and frequency. The time-frequency resolution of a spectrogram depends only on the window size and type and is independent of frequency.[7,8] A wide window gives better frequency resolution, but worsens the time resolution, whereas a narrow window improves time resolution but worsens frequency resolution. In contrast, the scalogram tiles the time-frequency plane in an irregular fashion, resulting in a frequency-dependent time-frequency resolution.[9] The WT of small frequency values provides good frequency resolution, but the time resolution is bad. On the other hand, the WT of large frequency values provides poor frequency resolution, but the time resolution is good. The WVD automatically satisfies the uncertainty principle of a signal (as seen in the perfect localization of impulses and sines), but interference terms appear in multicomponent signals.[8,16] Smoothing the WVD with a Gaussian filter removes the interference terms, but changes the uncertainty of the resulting TFR—its uncertainty is quantified by the standard deviations (time and frequency) of the Gaussian filter, noting that the smoothed WVD becomes spectrogram-like with a Gaussian filter that guarantees positivity ($\alpha\beta \geqslant 1$). In contrast, the time-frequency accuracy of the Hilbert spectrum is dependent on the accuracy of the EMD—if the decomposition into IMFs does not capture the signal's real behavior, the resulting Hilbert spectrum will not give precise time-frequency results. Since the EMD is an empirical method, it is not possible to analyze an EMD's time-frequency resolution in as rigorous a fashion as for the spectrogram, scalogram or WVD.

It is possible to improve the time-frequency resolution of a TFR with the reassignment method;[17] the reassignment method improves the time-frequency resolution of a TFR by concentrating its energy at a center of gravity. Auger and Flandrin[18] provide a computationally efficient way to compute the reassigned values for the spectrogram and the scalogram, first proposed in Ref. 17. For example, calculating the reassigned spectrogram amounts to the calculation of the reassigned coordinates $(\hat{t},\hat{\omega})$ for each time-frequency pair (t, ω) in the original spectrogram, where

$$\hat{t}=t-\Re\left(\frac{S_{1_{\mathcal{T}h}}(x,t,\omega)\cdot\overline{S_{1_h(x,t,\omega)}}}{|S_{1_h}(x,t,\omega)|^2}\right), \tag{8}$$

and

$$\hat{\omega}=\omega-\mathcal{I}\left(\frac{S_{1_{\mathcal{D}h}}(x,t,\omega)\cdot\overline{S_{1_h(x,t,\omega)}}}{|S_{1_h}(x,t,\omega)|^2}\right). \tag{9}$$

S_{1_h}, $S_{1_{\mathcal{T}h}}$, and $S_{1_{\mathcal{D}h}}$ are the short-time Fourier transforms with window functions $h(t)$, $t\cdot h(t)$, and $[\mathrm{d}h(t)/\mathrm{d}t]$, respectively. The reassigned spectrogram is then given by

$$E_{1_{ra}}(\omega',t')=\frac{1}{2\pi}\int\int E_1(\omega,t)\,\delta(t'-\hat{t}(t,\omega)) \cdot\delta(\omega'-\hat{\omega}(t,\omega))\,\mathrm{d}t\,\mathrm{d}\omega, \tag{10}$$

where $\delta(t)$ is the Dirac impulse, and the integration is performed over the range of all t and ω. In summary, the reassigned spectrogram requires the calculation of three short-time Fourier transforms (with three different window functions). These transforms are then used to move each value of the spectrogram, $E_1(\omega,t)$ [at (t,ω)] to its reassigned coordinates, $(\hat{t},\hat{\omega})$, calculated with Eqs. (8) and (9). Finally, Eq. (10) is used to sum up values assigned to the same $(\hat{t},\hat{\omega})$ bin. Note that the reassignment method is not restricted to a specific TFR, but can be applied to any time-frequency shift invariant distribution of Cohen's class.[8] The Hilbert spectrum, however, does not meet this requirement.

III. APPLICATION OF TFRs TO LASER-GENERATED/DETECTED LAMB WAVES

Broad-bandwidth Lamb waves are generated with a Nd:YAG laser (4–6-ns pulse) and measured with a high-fidelity (resonance-free) laser interferometer over a wide frequency range (200 kHz to 10 MHz).[3] Figure 1 shows a (transient) time-domain signal with a propagation distance of 11 cm measured in a 0.93-mm aluminum plate. The Nd:YAG laser fires at $t=0$ and generates a Lamb wave at the source location. (Note that the electromagnetic discharge of the Nd:YAG's firing causes a spurious noise spike at $t=0$.) The signal in Fig. 1 is discretized with a sampling frequency of 100 MHz, low-pass filtered at 10 MHz, and represents an average of 100 Nd:YAG shots to increase the signal-to-noise ratio (SNR). It is important to note that the broad bandwidth, high fidelity, and high SNR (due to signal averaging) of this

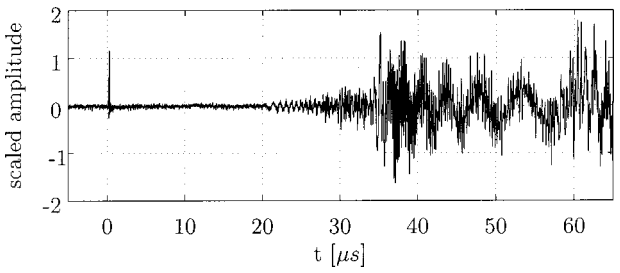


FIG. 1. Time-domain signal—multimode Lamb wave.

time-domain signal are crucial properties, making it an ideal signal to compare the effectiveness of the four candidate TFRs.

Assessment of the accuracy of the dispersion curves obtained with the candidate TFRs requires analytical results of the Rayleigh–Lamb frequency spectrum.[1] Solution of the Rayleigh–Lamb spectrum provides dispersion curves in the frequency-wave-number (f, k) domain, whereas a TFR maps a time-domain signal into the time-frequency domain. The group velocities for each of the different modes are determined by numerically differentiating f with respect to k to obtain the analytical dispersion curves in the time-frequency domain.[10]

Since Niethammer *et al.*[7] provide details on characterizing Lamb waves with the reassigned spectrogram, this paper briefly presents the results from this TFR, and only for comparison purposes. Figure 2 shows a contour plot of the square root of the reassigned spectrogram (384-point Hanning window) of the time-domain signal in Fig. 1, together with the analytically obtained dispersion curves as solid lines (all of the subsequent TFR plots include these analytical curves). The reassigned spectrogram provides a crisp definition of the individual modes, and these experimental modes are localized to the analytical curves. There is excellent definition of seven modes (s_0–s_2 and a_0–a_3) through a wide frequency range (up to 10 MHz). Note a general (and significant) advantage of TFRs in interpreting multimode Lamb waves—they enable the clear identification of the arrival time of the different modes (e.g., s_0 at 21 μs or a_0 at 35 μs). Finally, the

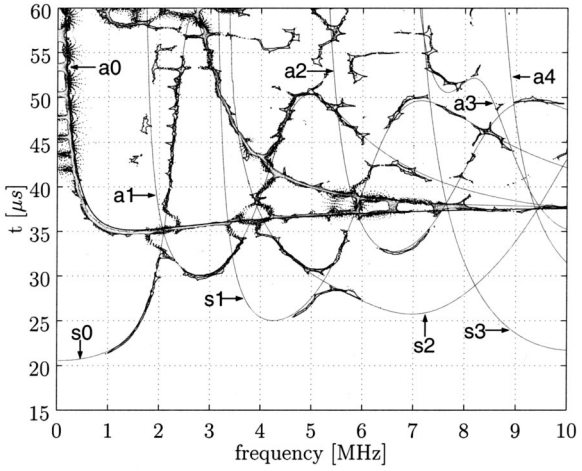


FIG. 2. Reassigned spectrogram of multimode Lamb wave, plus analytical solution (solid lines).

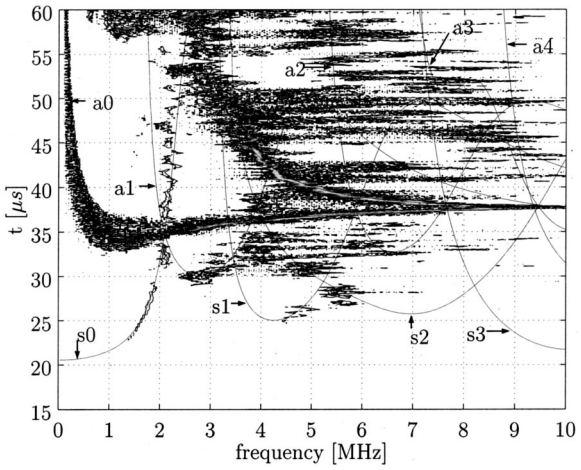


FIG. 3. Reassigned scalogram of multimode Lamb wave, plus analytical solution (solid lines).

broken lines above 50 μs are most likely caused by reflections from the plate's boundaries.

Figure 3 shows the square root of a reassigned scalogram of the same time-domain signal (Fig. 1) calculated with a Gabor wavelet. Although the time resolution at high frequencies is very good, there is not enough frequency resolution to separate the different modes at the high frequencies (e.g., above 2 MHz). The reassigned scalogram is effective in resolving the a_0 mode up to 10 MHz—an important feature for some applications. The WT has proven to be effective in many diverse applications (such as the detection of radar chirps), but a common theme in most of these applications is the need for good time resolution at high frequencies, and good frequency resolution at low frequencies. Figure 3 clearly shows that this attribute of the WT is not advantageous for resolving multiple, broadband Lamb modes. Note that an additional portion of this research[10] examines the Mexican-hat WT and shows its reassigned scalogram has even less resolution than the Gabor WT—the poor resolution in Fig. 3 is not due to improper selection of a mother wavelet, but instead is an inherent property of the WT.

The WVD and the smoothed WVD of the same time-domain signal (Fig. 1) are shown in Figs. 4 and 5, respectively. The results are stored in a 601×600 matrix, where columns correspond to a specific time and rows to a specific frequency (the frequency axis is equally spaced between 0 and 10 MHz, and the time axis has 0.1-μs discretization steps). The WVD of Fig. 4 is smoothed with a 21×21 Gaussian filter (which smoothes over a domain of 350 kHz $\times2.1$ μs) to create the smoothed WVD of Fig. 5. The selection of this particular filter size is based on ''visual'' comparisons, and note that it has $\alpha\beta<1$, so the resulting distribution is not positive (the negative values are set to zero in Fig. 5), but it provides good signal localization. Figure 4 shows that the WVD does a fairly good job in resolving the a_0 and the s_0 modes (they are both very localized) over a large frequency range—a_0 through the entire 10 MHz, and s_0 between 3 and 10 MHz. Unfortunately, all the other modes are obscured by interference terms and are not visible. The smoothed WVD of Fig. 5 provides a very good representation of the individual modes, with the s_0 and a_0 modes

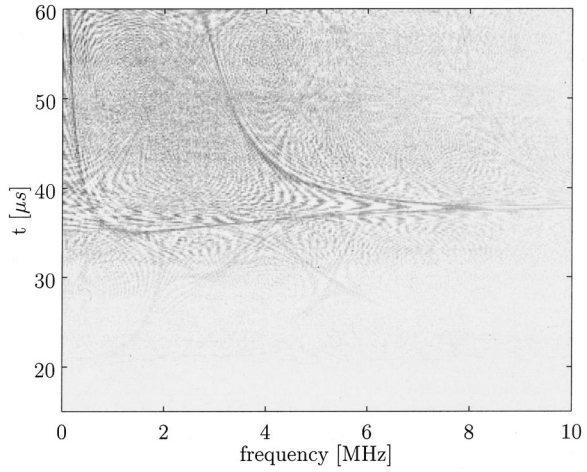


FIG. 4. Wigner–Ville distribution of multimode Lamb wave, plus analytical solution (solid lines).

clearly visible through the entire frequency bandwidth (to 10 MHz), the a_1 mode appears from 2 to 5 MHz, and traces of the s_1, s_2, and a_2 modes are evident. There is a general lack of time-frequency resolution (clarity) in the smoothed WVD. For example, it is difficult to positively identify the individual modes for frequencies above 5 MHz and times greater than 50 μs. The dispersion curves developed with the smoothed WVD (Fig. 5) render a mode resolution of about the same quality as the un-reassigned spectrogram.[7] It is important to note that it is possible to use a reassignment algorithm on the smoothed WVD[6,17] (the smoothed WVD of Fig. 5 has *not* been reassigned), but unlike the reassigned spectrogram and scalogram, the reassignment calculations for the smoothed WVD are computationally intensive (first requiring the smoothing procedure), difficult to implement, and time consuming (computationally).

The first step for the Hilbert spectrum is to decompose the time-domain signal of Fig. 1 signal into IMFs. Figure 6 shows the first seven IMFs, noting that the sifting process is stopped when a monotonic IMF (the 10th) occurs. Figure 6 shows that basically only the first 5 IMFs contain ''useful'' information, since IMFs above 5 are too low in frequency

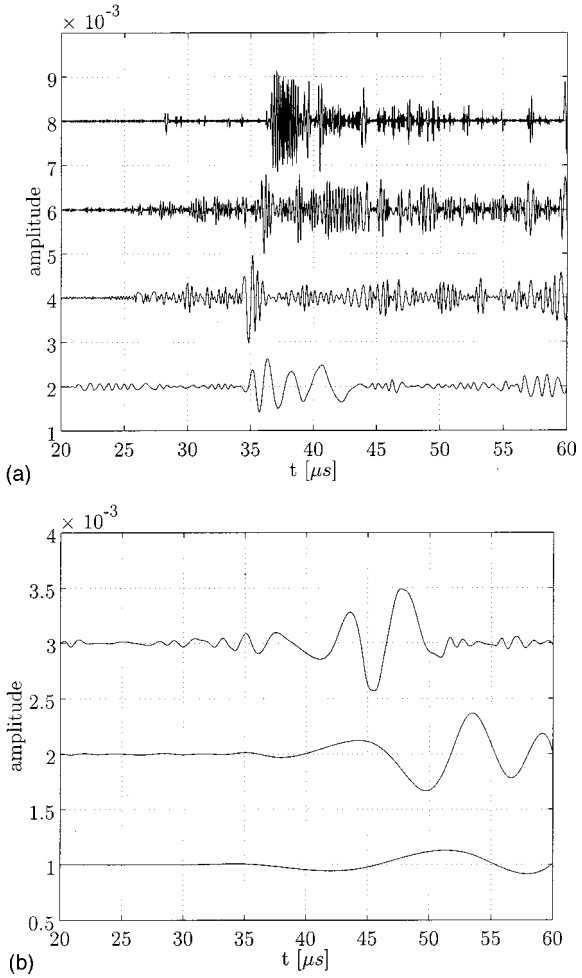


FIG. 6. (a) First four intrinsic mode functions (first to fourth from top to bottom) of multimode Lamb wave. (b) Fifth through seventh intrinsic mode functions (from top to bottom) of multimode Lamb wave.

(100 KHz or less) to be of importance for this application. Figure 7 shows a contour plot of the Hilbert spectrum computed from these IMFs, and it is far from being a ''clean'' representation of the dispersion curves of this signal. Only the a_0 mode is definitively present—a large number of points are clustered around the s_0 mode between 3 and 5 MHz, but

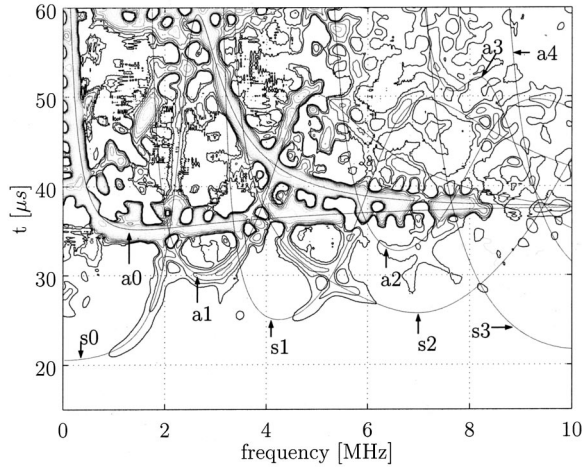


FIG. 5. Smoothed Wigner–Ville distribution of multimode Lamb wave, plus analytical solution (solid lines).

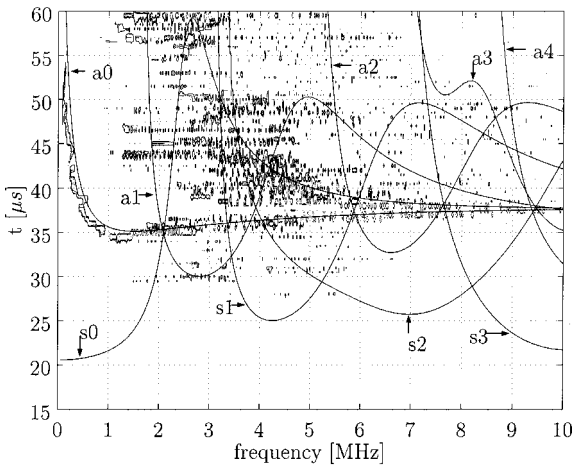


FIG. 7. Hilbert spectrum of multimode Lamb wave, plus analytical solution (solid lines).

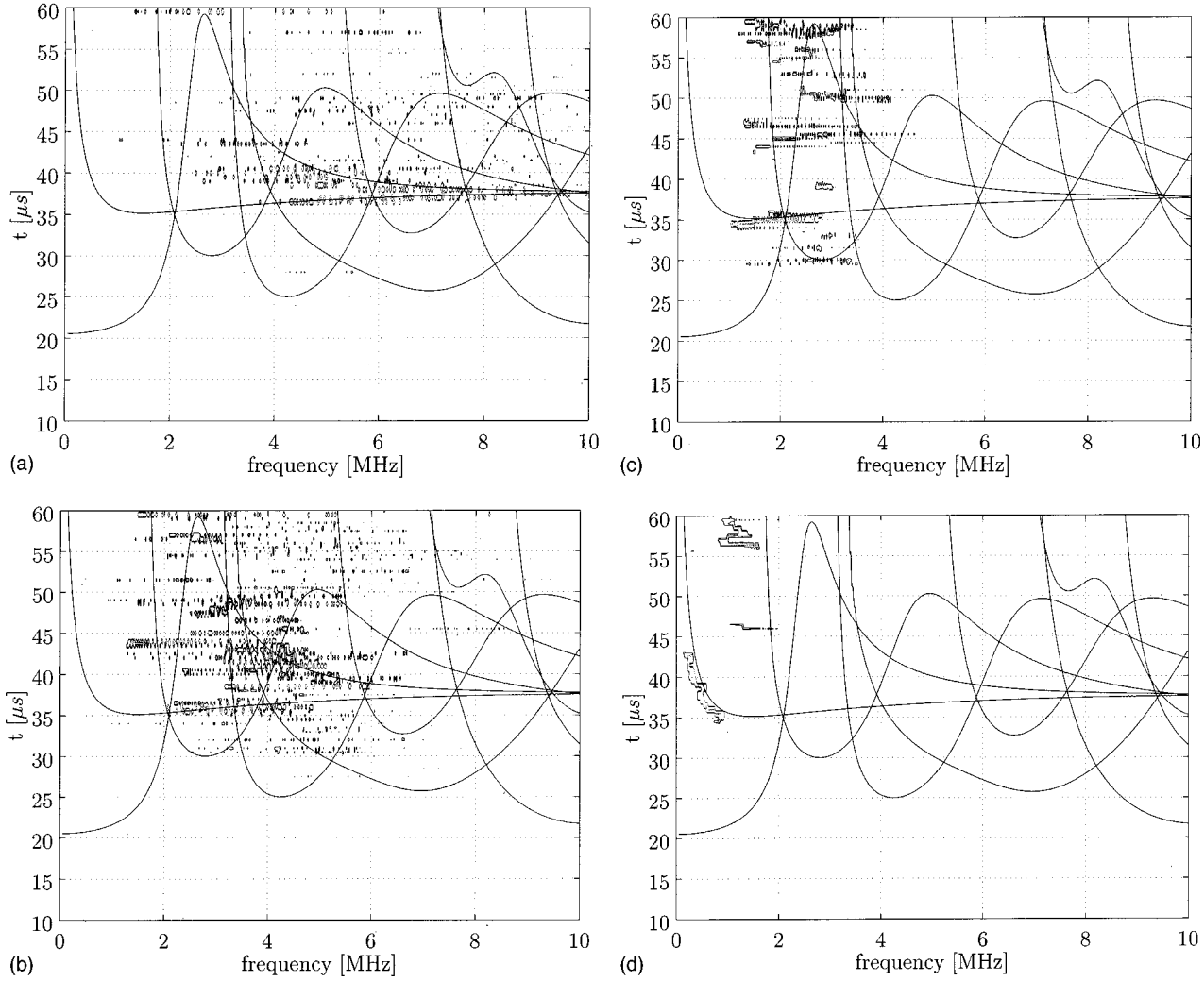


FIG. 8. (a) Hilbert spectra of first IMF, plus analytical solution (solid lines). (b) Hilbert spectra of second IMF, plus analytical solution (solid lines). (c) Hilbert spectra of third IMF, plus analytical solution (solid lines). (d) Hilbert spectra of fourth IMF, plus analytical solution (solid lines).

this mode is not clearly defined. In order to interpret the Hilbert spectrum of Fig. 7, Fig. 8 shows the instantaneous frequencies of the first four IMFs. By analyzing the Hilbert spectra of each IMF, it is possible to determine exactly which IMFs contribute to exactly which features in the Hilbert spectrum. For example, IMF 1 contains most of the a_0 and the s_0 modes between 4 and 10 MHz [Fig. 8(a)], but fails to localize its energy exactly on either mode—the first IMF scatters its energy around both a_0 and s_0. Also note the high level of noise present in IMF 1. IMF 2 contributes to the frequency band from 2 to 6 MHz [Fig. 8(b)]; except for a very small part of the a_0 mode at approximately 3 MHz, there is not an obvious correlation between the analytical modes and the resulting Hilbert spectrum. IMF 3 is quite effective in representing the a_0 mode from 1 to 3 MHz [Fig. 8(c)], but it fails to represent the other modes in the same frequency range. IMF 4 constitutes the a_0 mode from 35 to 45 μs [Fig. 8(d)]. It appears that each IMF specializes in a certain frequency range for these laser-generated/detected Lamb waves. This approach works fine for frequency ranges where there are only a few modes present, but it clearly fails for frequency ranges that contain multiple modes. This failure for multiple modes occurs because there is only one instantaneous frequency for an IMF for any time *t*. As a result, the EMD and Hilbert spectrum cannot resolve the multimode dispersion curves for this plate.

An inherent problem associated with all window-based TFRs (and thus their respective reassigned representations) is the ''ladder-like'' effect that occurs when the distance between two mode lines is less than the uncertainty of their representations. These mode lines will then interfere with each other, resulting in a smeared representation—this behavior is particularly evident when mode lines intersect each other (e.g., the intersection of the s_0, a_0, and a_1 modes at 2 MHz in Figs. 2–5). Image-processing methodologies that take the specific mode structure into account are a possible remedy for this smearing. Note that the comparison between these TFRs is generic enough to be extended to any plate thickness and material. As an example of the general nature of these results, the reassigned spectrogram has been used to successfully resolve the dispersion relationships for multimode circumferential waves in a cylinder.[10]

IV. CONCLUSION

This paper establishes the effectiveness of four candidate TFRs to analyze broadband, multimode ultrasonic

waves. All TFRs are not equally suited to represent every signal type, so a vital contribution of this study is the identification of which TFRs are most effective for analyzing multimode Lamb waves, and why. Overall, this paper illustrates the effectiveness of using TFRs to quantitatively resolve changes in the frequency content of these nonstationary signals, as a function of time. Another strength of TFRs demonstrated in this study is their ability to facilitate the identification of the arrival times of the different modes in a multimode signal.

This research shows that the reassigned spectrogram is extremely effective in localizing multiple, closely spaced Lamb modes in both time and frequency. The reassigned scalogram can only accurately resolve a single mode (the a_0) in this multimode Lamb wave; this deficiency is attributed to the method's insufficient frequency resolution for high frequencies. The WVD is ineffective in this application, while the smoothed WVD provides a very good representation of the individual Lamb modes, while showing a general lack of time-frequency clarity (note that with more smoothing, the WVD becomes similar to a spectrogram). Note that the time-frequency resolution of the smoothed WVD can be improved with a reassignment algorithm (like the reassignment of the spectrogram and scalogram). Unfortunately, the reassignment calculations for the smoothed WVD are intensive (first requiring the smoothing procedure), difficult to implement, and computationally time consuming. As a result, the reassigned spectrogram appears to be a better choice to characterize these multimode Lamb waves. Even though the EMD combined with the Hilbert transform is based on local properties of the signal itself, this study shows that this Hilbert spectrum technique is not capable of extracting the Rayleigh–Lamb frequency spectrum between 200 kHz and 10 MHz, from a laser-generated/detected time-domain signal. The Hilbert spectrum fails because of the plate's intricate mode structure, with multiple modes occupying the same frequency band, and intersecting each other in time and frequency.

ACKNOWLEDGMENTS

This work is supported by the Office of Naval Research M-URI Program ''Integrated Diagnostics'' (Contract Number: N00014-95-1-0539). The Deutscher Akademischer Austausch Dienst (DAAD) provided partial support to Marc Niethammer. The authors thank NASA Goddard for providing them with the EMD/Hilbert Spectrum software and Christoph Eisenhardt for his contributions in making the experimental measurements.

[1] R. D. Mindlin, ''Waves and vibrations in isotropic elastic plates,'' in *Structural Mechanics*, edited by J. N. Goodier and N. J. Hoff (Pergamon, New York, 1960).

[2] D. Alleyne and P. Cawley, ''A two-dimensional Fourier transform method for measurement of propagating multimode signals,'' J. Acoust. Soc. Am. **89**, 1159–1168 (1991).

[3] C. Eisenhardt, L. J. Jacobs, and J. Qu, ''Application of laser ultrasonics to develop dispersion curves for elastic plates,'' J. Appl. Mech. **66**, 1043–1045 (1999).

[4] W. H. Prosser, M. D. Seale, and B. T. Smith, ''Time-frequency analysis of the dispersion of Lamb modes,'' J. Acoust. Soc. Am. **105**, 2669–2676 (1999).

[5] Y. Hayashi, S. Ogawa, H. Cho, and M. Takemoto, ''Noncontact estimation of thickness and elastic properties of metallic foils by wavelet transform of laser-generated Lamb waves,'' Nondestr. Test. Eval. **32**, 21–27 (1999).

[6] S. Holland, T. Kosel, R. Weaver, and W. Sachse, ''Determination of plate source, detector separation from one signal,'' Ultrasonics **38**, 620–623 (2000).

[7] M. Niethammer, L. J. Jacobs, J. Qu, and J. Jarzynski, ''Time-frequency representation of Lamb waves using the reassigned spectrogram,'' J. Acoust. Soc. Am. **107**, L19–L24 (2000).

[8] L. Cohen, *Time-Frequency Analysis* (Prentice-Hall, Englewood Cliffs, NJ, 1995).

[9] S. Mallat, A *Wavelet Tour of Signal Processing* (Academic, New York, 1998).

[10] M. Niethammer, ''Application of Time-Frequency Representations to Characterize Ultrasonic Signals,'' M.S. thesis, Georgia Institute of Technology, Atlanta, 1999.

[11] S. D. Meyers, B. G. Kelly, and J. J. OBrien, ''An introduction to wavelet analysis in oceanography and meterology: With application to the dispersion of Yanai waves,'' Mon. Weather Rev. **121**, 2858–2866 (1993).

[12] D. Casasent and R. Shenoy, ''New Gabor wavelets with shift-invariance for improved time-frequency analysis and signal detection,'' in Proc. SPIE **2762**, Wavelet Applications III, edited by H. Szu, 244–255 (1996).

[13] L. Cohen, ''A Primer on Time-Frequency Analysis,'' in Time-Frequency Signal Analysis, edited by B. Boashash (Longman Chesire, 1992), pp. 3–42.

[14] N. E. Huang, Z. Shen, and S. R. Long, ''A new view of nonlinear water waves: The Hilbert spectrum,'' Annu. Rev. Fluid Mech. **31**, 417–457 (1999).

[15] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, ''The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis,'' Proc. R. Soc. London, Ser. A **454**, 903–995 (1998).

[16] L. Cohen, ''Time-frequency distributions—A review,'' Proc. IEEE **77**, 941–981 (1989).

[17] K. Kodera, R. Gendrin, and C. de Villedary, ''Analysis of time-varying signals with small BT values,'' IEEE Trans. Acoust., Speech, Signal Process. **26**, 64–76 (1978).

[18] F. Auger and P. Flandrin, ''Improving the readability of time-frequency and time-scale representations by the reassignment method,'' IEEE Trans. Signal Process. **43**, 1068–1089 (1995).

Exact solutions for transient spherical radiation

J. A. Hamilton and R. J. Astley[a)]
Department of Mechanical Engineering, University of Canterbury, Private Bag 4800, Christchurch, New Zealand

(Received 13 October 2000; accepted for publication 21 February 2001)

Closed-form, analytic solutions are derived for exterior wave fields generated by the transient, axisymmetric motion of a spherical boundary. These are intended for use as benchmark problems for assessing the accuracy and correctness of transient numerical schemes. Classical Laplace transform methods are used. The derivation of these solutions is presented in sufficient detail to permit their reconstruction by the reader for multipoles of order n where $0 \leqslant n \leqslant 25$. Composite solutions can also be obtained by superposition. A novel solution of this type is presented for the wave field generated by the transient motion of a spherical piston in a spherical baffle. The transient wave fields obtained in this way are shown to be consistent with steady time-harmonic solutions for large times, and with high-resolution transient numerical solutions at finite times. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1365425]

PACS numbers: 43.20.Px, 43.20.Tb, 43.40.Rj, 43.30.Jx [ANN]

I. INTRODUCTION

The accuracy of numerical schemes for predicting unbounded wave fields can be assessed most effectively by the use of canonical test problems for which precise analytic or semi-analytic solutions are available.[1] Such solutions are needed for example to validate the high-order nonreflecting boundary treatments which are frequently imposed at the outer edge of domain-based finite element (FE) schemes. These are characterized by their transparency to radiated or scattered multipole components of a given order. The local boundary operators of Bayliss, Gunzberger, and Turkel[2,3] are designed along these lines, as are nonlocal treatments such as the FE-DtN approach.[4,5] Infinite element formulations[6] based on truncated versions of the ''Atkinson/Wilcox/Holford'' expansion in spherical[7] and spheroidal/ellipsoidal coordinates[8–10] can also be regarded in this way.[11]

Steady, two- and three-dimensional, multipole solutions are readily available as benchmark problems for time-harmonic numerical schemes. They arise naturally as separable solutions of Helmholtz' equation in cylindrical, elliptical, spherical, and spheroidal coordinates, and can be superimposed to generate more complex solutions both for radiation and scattering.[12,13]

Fewer benchmark problems are available for transient computation. D'Alembert's solution for a pulsating sphere excited from rest has been used for this purpose, as has the analogous transient dipole or ''juddering'' sphere[14,15] and the transient sound field generated by a piston in a plane baffle.[10,16] A transient quadrupole solution—obtained from a fast Fourier transform of the time harmonic solution—has also been used.[17]

A number of elastic-acoustic problems permit an exact solution on the surface of the scatterer. The coupled surface response of a fluid-filled spherical elastic shell to a step pressure pulse has been formulated by Geers for use as a canonical problem for underwater shock codes.[18,19] The analogous surface solution for a rigid sphere has been treated by Huang[20] along with the response of an empty elastic spherical shell[21,22] and concentric shells with fluid in-between.[23] Menton predicted the radiation from an elastic spherical shell excited by internal sources[24] and Stepanishen has developed a general solution for axisymmetric elastic shells using eigenvector expansions.[25,26]

''Steady'' time-harmonic solutions can be used to check the validity of numerical schemes at large times. The steady harmonic solution for a piston in a spherical baffle has been used extensively for this purpose.[27,28] Such checks do not of course provide any assurance of the accuracy or correctness of a transient solution at finite elapsed times.

In the current article transient multipole solutions are formulated which are valid at all times and at all points in the region exterior to a sphere. A classical Laplace transform approach is used. Impulsive, stepwise and time-harmonic excitation—starting from rest at time $t=0$—are considered. These solutions are presented in sufficient detail to permit their reconstruction for multipoles of order $n \in [0,25]$. Finally, a novel solution is presented for the sound field generated by the transient motion of a spherical piston in a rigid spherical baffle. This is constructed by superposition of multipole components and forms a transient analogue of the steady harmonic test problem used to check transient computations at large times.[27,28] For all of the exact solutions presented here, close agreement is demonstrated both with steady time harmonic solutions at large times and with accurate high-order numerical schemes at finite elapsed times.

II. THE TRANSIENT EXTERIOR PROBLEM

Figure 1(a) shows the geometry of the problem to be considered. Let Γ be a three-dimensional region of unlimited extent with an internal boundary S. The unit normal on S acting into the region Γ is denoted by $\underline{n}$. Let $u(\underline{x},t)$ and

[a)]Present address: The Institute of Sound and Vibration Research, University of Southampton, Southampton SO17 1BJ, United Kingdom. Electronic mail: rja@isvr.soton.ac.uk

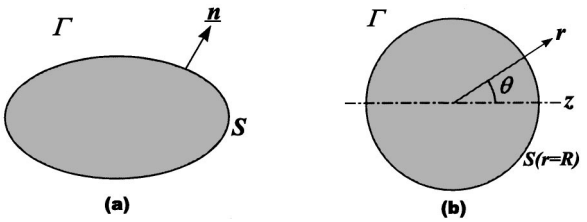


FIG. 1. Problem geometry. (a) General. (b) Spherical.

$a(\underline{x},t)$ denote the unknown acoustical pressure in Γ and the given normal surface acceleration on S. Initial (and past) conditions are given by

$$u(\underline{x},t)=0 \quad \text{and} \quad \dot{u}(\underline{x},t)=0 \quad \text{for } \underline{x}\in\Gamma, t\leqslant 0, \tag{1}$$

and

$$a(\underline{x},t)=0 \quad \text{for } \underline{x}\in S, t\leqslant 0, \tag{2}$$

where an overdot denotes differentiation with respect to time. The acoustic pressure field is then governed by the linearized wave equation

$$\nabla^2 u=\frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} \quad \text{for } \underline{x}\in\Gamma, \tag{3}$$

where c is the sound speed and ∇^2 is the Laplacian operator. Continuity of normal acceleration on the surface S gives

$$\nabla u\cdot\underline{n}=-\rho a(\underline{x},t) \quad \text{for } \underline{x}\in S, \tag{4}$$

where ρ is the mean fluid density. Equations (1)–(4) constitute a well-posed initial value problem for $u(\underline{x},t)$. In this article we present results only for the Neumann boundary condition in Eq. (4). However, other conditions may be accommodated using this framework as demonstrated in Appendix A.

III. THE MODAL FORMULATION FOR SPHERICAL RADIATION

Consider the particular case when the surface S is a sphere of radius R as shown in Fig. 1(b). Assume also that the surface excitation is axially symmetric about the z axis. The acoustic pressure and surface acceleration can then be expanded in terms of a transverse orthogonal basis. This gives

$$u(\underline{x},t)=u(r,\theta,t)=\sum_{n=0}^{\infty} u_n(r,t)P_n(\cos\theta), \tag{5}$$

and

$$a(\underline{x},t)=a(\theta,t)=\sum_{n=0}^{\infty} a_n(t)P_n(\cos\theta), \tag{6}$$

where r and θ are spherical radial and angular coordinates and $P_n(\ )$ is the Legendre polynomial of order n. Substitution of Eqs. (5) and (6) into Eqs. (1)–(4) and use of the orthogonality properties of the Legendre polynomial[13] yields the following initial value problem for each modal component of pressure,

$$\frac{\partial^2 u_n}{\partial r^2}+\frac{2}{r}\frac{\partial u_n}{\partial r}-n(n+1)\frac{u_n}{r^2}=\frac{1}{c^2}\frac{\partial^2 u_n}{\partial t^2} \quad \text{for } r\geqslant R, \quad t\geqslant 0, \tag{7}$$

where

$$\left.\frac{\partial u_n}{\partial r}\right|_{r=R}=-\rho a_n(t) \quad \text{for } t\geqslant 0. \tag{8}$$

The initial conditions are

$$u_n(r,t)=0 \quad \text{and} \quad \dot{u}_n(r,t)=0 \quad \text{for } t\leqslant 0. \tag{9}$$

IV. LAPLACE TRANSFORM OF THE MODAL SOLUTION

The solution of Eq. (7) subject to initial conditions (9) is obtained by transforming the equation into Laplace transform space. This gives

$$\frac{d^2\overline{u_n}}{dr^2}+\frac{2}{r}\frac{d\overline{u_n}}{dr}-\left(\frac{n(n+1)}{r^2}+\frac{s^2}{c^2}\right)\overline{u_n}=0, \tag{10}$$

where

$$\left.\frac{d\overline{u_n}}{dr}\right|_{r=R}=-\rho\overline{a_n}(s). \tag{11}$$

Here $\overline{u_n}(r,s)$ and $\overline{a_n}(s)$ denote the Laplace transforms of $u_n(r,t)$ and $a_n(t)$. The solution for $\overline{u_n}$, which excludes inwardly propagating disturbances, is given by

$$\overline{u_n}(r,s)=\frac{-\rho c\overline{a_n}(s)}{s}\frac{k_n(sr/c)}{k_n'(sR/c)}, \tag{12}$$

where $k_n(\ )$ and $k_n'(\ )$ are the modified spherical Bessel function of the third kind and its derivative. These can be expressed as finite series[13]

$$k_n(z)=\frac{\pi e^{-z}}{2z}\sum_{i=0}^{n}\gamma_i\left(\frac{1}{2z}\right)^i, \tag{13a}$$

and

$$k_n'(z)=\frac{-\pi e^{-z}}{2z}\sum_{i=0}^{n+1}\delta_i\left(\frac{1}{2z}\right)^i, \tag{13b}$$

where

$$\gamma_i=\frac{(n+i)!}{i!(n-i)!} \quad (0\leqslant i\leqslant n), \quad \delta_{n+1}=2(n+1)\gamma_n,$$
$$\delta_0=\gamma_0, \quad \text{and} \quad \delta_i=\gamma_i+2i\gamma_{i-1} \quad (1\leqslant i\leqslant n). \tag{14}$$

Substitution into Eq. (12) and some rearrangement of terms then yields

$$\overline{u_n}(r,s)=\rho c\left(\frac{R}{r}\right)e^{-sx}\left\{\frac{\overline{a_n}(s)(\nu_0 s^n+\nu_1 s^{n-1}+\cdots\nu_n)}{(\mu_0 s^{n+1}+\mu_1 s^n+\cdots\mu_{n+1})}\right\}$$
$$=\rho c\left(\frac{R}{r}\right)e^{-sx}\left\{\frac{\overline{a_n}(s)F_n(r,s)}{G_n(s)}\right\}, \tag{15}$$

where $x=(r-R)/c$. The coefficients in the polynomials $F_n(r,s)$ and $G_n(s)$ are given by $\nu_i=\gamma_i(c/2r)^i$ and $\mu_i=\delta_i(c/2R)^i$. The solution for the transient modal amplitude

$u_n(r,t)$ is then obtained by taking the inverse Laplace transform of the above expression. The roots of the denominator $G_n(s)$ in Eq. (15) are central to this procedure.

A. The roots of $G_n(s)$

The polynomial $G_n(s)$ has $n+1$ roots. When n is odd these occur in $m[=(n+1)/2]$ conjugated pairs, $a_j \pm ib_j$ $(j=1,2,...,m)$ where $i=\sqrt{-1}$. When n is even these occur as $m(=n/2)$ conjugated pairs, $a_j \pm ib_j$ $(j=1,2,...,m)$ plus a single real root a_0. No multiple roots are present in either case.

These roots have been obtained in the current instance by using a two-stage procedure whereby roots determined initially from an eigenvalue method[29] are ''polished'' by an iterative Newton–Raphson procedure to ensure high accuracy. Both stages are implemented using variable extended-precision arithmetic.[30] In general the Newton–Raphson procedure converges within two iterations or not at all for a given root z_j. Convergence is defined in the current instance by the criteria that the root is accurate to at least 16 significant figures. To ensure the desired accuracy (say P), the convergence criteria is calculated using the slope at each root such that $|G_n(z_j)| \leqslant P/2 \times |G_n'(z_j)|$. If this criteria was not met for all z_j, the precision was increased. However, the accurate representation of the coefficients in $G_n(z)$ was also a consideration in deciding the amount of precision required. It was found that, in the *absence* of any special coding procedures to avoid floating point errors, arithmetic precision of approximately $2\times(n-5)$ significant figures was required to accurately represent these terms for a multipole of order n $(n<100)$. This was the overriding consideration in determining the precision required.

Roots were calculated in this way for multipoles of order up to and including $n=75$. These are tabulated for reference in Table I for $n=0,1,...,25$. It should be noted that the values have been normalized with respect to R and c so the actual value of a root is given by $z_j=(c/R)\hat{z}_j$ where $\hat{z}_j$ is the tabulated value.

V. PARTICULAR SOLUTIONS

The inversion of Eq. (15) is presented for three specific forms of the modal surface acceleration: an impulse function, a step function, and a time-harmonic excitation starting from rest. A multi-modal solution is also presented for the exterior sound field due to the transient motion of a spherical piston in a spherical baffle.

A. The multipole impulse solution, $a_n(t)=\delta(t)$

The transient modal solution generated by a unit surface impulse, $a_n(t)=\delta(t)$, where $\delta(\ )$ is the Dirac delta function, is now presented. This will be termed the *multipole impulse solution* and will be denoted by $\hat{u}_n(r,t)$. It is obtained by setting $\overline{a_n}(s)=1.0$. The inversion of expression (15) is performed in the usual way by expanding the right-hand side of this equation as a series of partial fractions and inverting each in turn. After some manipulation this gives,

for n odd,

$$\hat{u}_n(r,t)=\rho c(R/r)\mathrm{H}(t')\Lambda(r,t), \tag{16a}$$

and for n even

$$\hat{u}_n(r,t)=\rho c(R/r)\mathrm{H}(t')[C(r)e^{a_0 t'}+\Lambda(r,t)], \tag{16b}$$

where

$$\Lambda(r,t)=\sum_{j=1}^{m} e^{a_j t'}[A_j(r)\sin(b_j t')+B_j(r)\cos(b_j t')], \tag{17}$$

$t'=t-(r-R)/c$, $m=(n+1)/2$ (n odd), $m=n/2$ (n even) and H( ) is the Heaviside step function in the usual notation. The real coefficients $A_j(r)$, $B_j(r)$, and $C(r)$ are polynomials in $1/r$, obtained from,

for n odd,

$$A_j(r)+iB_j(r)=[F_n(r,s)/b_j\Phi(s)]_{s=a_j+ib_j}, \tag{18a}$$

and for n even

$$A_j(r)+iB_j(r)=[F_n(r,s)/b_j(s-a_0)\Phi(s)]_{s=a_j+ib_j}, \tag{18b}$$

$$C(r)=[F_n(r,s)/\Psi(s)]_{s=a_0}, \tag{18c}$$

where

$$\Phi(s)=\prod_{\substack{k=1\\(k\neq j)}}^{m}((s-a_k)^2+b_k^2), \tag{19a}$$

and

$$\Psi(s)=\prod_{k=1}^{m}((s-a_k)^2+b_k^2). \tag{19b}$$

B. The multipole step solution, $a_n(t)=\mathrm{H}(t)$

The transient modal solution caused by a unit step acceleration at $t=0$, $a_n(t)=\mathrm{H}(t)$, will be termed the *transient step response* and denoted by $u_n^{\mathrm{H}}(r,t)$. After setting $\overline{a_n}(s)=1/s$, the following solution is obtained.

For n odd,

$$u_n^{\mathrm{H}}(r,t)=\rho c(R/r)\mathrm{H}(t')[X(r)+\Lambda(r,t)], \tag{20a}$$

and for n even

$$u_n^{\mathrm{H}}(r,t)=\rho c(R/r)\mathrm{H}(t')[X(r)+C(r)e^{a_0 t'}+\Lambda(r,t)]. \tag{20b}$$

The real coefficients $X(r)$, $A_j(r)$, $B_j(r)$, and $C(r)$ are polynomials in $(1/r)$ and are given by,

for n odd,

$$A_j(r)+iB_j(r)=[F_n(r,s)/b_j s\Phi(s)]_{s=a_j+ib_j}, \tag{21a}$$

$$X(r)=[F_n(r,s)/\Psi(s)]_{s=0}, \tag{21b}$$

and for n even

$$A_j(r)+iB_j(r)=[F_n(r,s)/b_j s(s-a_0)\Phi(s)]_{s=a_j+ib_j}, \tag{21c}$$

$$C(r)=[F_n(r,s)/s\Psi(s)]_{s=a_0}, \tag{21d}$$

$$X(r)=[F_n(r,s)/(s-a_0)\Psi(s)]_{s=0}. \tag{21e}$$

TABLE I. Table of the zeros of $G_n(s)$ for $0 \leqslant n \leqslant 25$ [root $z_j = (c/R)\hat{z}_j$ where $\hat{z}_j$ is the tabulated value].

n	Real	Imaginary
0	−1.000 0000 0000 0000 E+0	
1	−1.000 0000 0000 0000 E+0	±1.000 0000 0000 0000 E+0
2	−1.783 2434 2804 8749 E+0	
	−1.108 3782 8597 5626 E+0	±1.954 0933 9251 2700 E+0
3	−2.298 1335 4024 9599 E+0	±8.906 0065 6002 5654 E−1
	−1.201 8664 5975 0401 E+0	±2.903 9165 3244 7329 E+0
4	−3.051 5949 0491 1486 E+0	
	−2.691 6433 9723 8399 E+0	±1.779 3524 9328 8277 E+0
	−1.282 5591 5030 5858 E+0	±3.857 4502 7037 2908 E+0
5	−3.630 2924 1265 8535 E+0	±8.768 3052 2985 6771 E−1
	−3.015 9041 6262 5265 E+0	±2.674 4583 2516 9343 E+0
	−1.353 8034 2471 6200 E+0	±4.815 1765 8380 3848 E+0
6	−4.355 6134 6954 4028 E+0	
	−4.109 7280 0232 7458 E+0	±1.754 9674 6683 6428 E+0
	−3.294 5775 4567 6078 E+0	±3.576 5024 9982 2298 E+0
	−1.417 8877 1722 4451 E+0	±5.776 5832 8455 4173 E+0
7	−4.959 4880 1296 4765 E+0	±8.726 0763 2422 3758 E−1
	−4.523 5374 7940 7286 E+0	±2.637 5497 2867 4625 E+0
	−3.540 6157 9118 9423 E+0	±4.484 8715 3558 8382 E+0
	−1.476 3587 1643 8526 E+0	±6.741 0920 9079 9774 E+0
8	−5.669 8102 8003 2257 E+0	
	−5.482 7335 9119 9120 E+0	±1.746 5811 0966 3632 E+0
	−4.890 0991 8522 3068 E+0	±3.525 1789 1644 6869 E+0
	−3.761 9626 9316 7544 E+0	±5.398 8008 0694 0225 E+0
	−1.530 2993 9039 4140 E+0	±7.708 2105 1798 4931 E+0
9	−6.287 2298 6802 9775 E+0	±8.707 7344 2556 8537 E−1
	−5.947 6787 1661 4656 E+0	±2.623 7246 3970 5732 E+0
	−5.220 7168 4326 5393 E+0	±4.417 7439 2809 8047 E+0
	−3.963 8781 9865 6686 E+0	±6.317 5907 7977 0616 E+0
	−1.580 4963 7343 3490 E+0	±8.677 5418 7538 9750 E+0
10	−6.988 3277 2702 7391 E+0	
	−6.837 2561 3052 3973 E+0	±1.742 6744 9462 4877 E+0
	−6.368 0891 8334 1414 E+0	±3.504 6044 2830 7722 E+0
	−5.522 9078 0428 6889 E+0	±5.314 9227 6256 5554 E+0
	−4.150 0434 0914 7560 E+0	±7.240 6442 2522 4911 E+0
	−1.627 5396 0918 6468 E+0	±9.648 7693 6101 3351 E+0
11	−7.614 2189 7441 7910 E+0	±8.698 1385 5382 7800 E−1
	−7.335 6485 8061 0051 E+0	±2.616 9286 8078 8785 E+0
	−6.753 1459 4213 0212 E+0	±4.389 3142 6931 8021 E+0
	−5.801 9578 8319 6006 E+0	±6.216 3460 1832 7422 E+0
	−4.323 1450 0098 0251 E+0	±8.167 4603 3218 6014 E+0
	−1.671 8836 1866 5569 E+0	±1.062 1638 5694 9332 E+1
12	−8.309 0822 7346 4450 E+0	
	−8.182 3559 5238 3462 E+0	±1.740 5322 9658 1262 E+0
	−7.792 9394 2297 8456 E+0	±3.494 0405 6357 5514 E+0
	−7.109 3243 0501 2825 E+0	±5.277 7552 4157 6970 E+0
	−6.061 7428 6228 6940 E+0	±7.121 6538 8721 5912 E+0
	−4.485 2096 9552 8239 E+0	±9.097 6197 9819 2960 E+0
	−1.713 8866 2507 7853 E+0	±1.159 5943 1383 1414 E+1
13	−8.940 7749 4522 8151 E+0	±8.692 4948 6852 9681 E−1
	−8.704 4226 7020 7853 E+0	±2.613 0602 1891 4424 E+0
	−8.216 5721 5719 3049 E+0	±4.374 1836 7100 5271 E+0
	−7.441 3868 7131 8633 E+0	±6.169 7508 2368 7491 E+0
	−6.305 1992 0897 2076 E+0	±8.030 5152 4315 0631 E+0
	−4.637 8077 0659 1432 E+0	±1.003 0769 9842 5719 E+1
	−1.753 8364 4048 8806 E+0	±1.257 1513 7884 9302 E+1
14	−9.631 1451 9159 0926 E+0	
	−9.521 9903 4113 6378 E+0	±1.739 2290 9563 2290 E+0
	−9.188 6897 7381 8928 E+0	±3.487 8385 1871 1752 E+0
	−8.612 0325 5803 9124 E+0	±5.257 3705 9761 1117 E+0
	−7.752 9528 5788 0572 E+0	±7.065 0978 3612 6813 E+0
	−6.534 6110 7765 9618 E+0	±8.942 6323 2443 4038 E+0
	−4.782 1826 5279 3125 E+0	±1.096 6612 3099 5598 E+1
	−1.791 9681 4287 6793 E+0	±1.354 8210 1138 4696 E+1
15	−1.026 7060 3639 8684 E+1	±8.688 8962 9647 1922 E−1
	−1.006 1726 0884 0310 E+1	±2.610 6409 8492 4524 E+0
	−9.641 2617 3989 4723 E+0	±4.365 0669 8985 1717 E+0
	−8.983 4977 0133 3351 E+0	±6.143 5339 8972 0146 E+0
	−8.046 8456 2096 3261 E+0	±7.963 5898 1524 9862 E+0
	−6.751 7941 2335 9739 E+0	±9.857 7398 7605 5406 E+0
	−4.919 3378 7582 1906 E+0	±1.190 4892 0812 1376 E+1
	−1.828 4764 8623 7083 E+0	±1.452 5914 4320 6254 E+1
16	−1.095 4039 5504 4612 E+1	
	−1.085 8168 7554 8810 E+1	±1.738 3767 0251 6551 E+0
	−1.056 6631 4550 8984 E+1	±3.483 8686 2253 1245 E+0
	−1.006 6793 4083 5680 E+1	±5.244 8141 6971 5957 E+0
	−9.334 2324 6462 5338 E+0	±7.032 5675 2727 7090 E+0
	−8.325 3155 5590 6164 E+0	±8.865 0277 1686 6870 E+0
	−6.958 2180 3315 7319 E+0	±1.077 5602 3064 8507 E+1
	−5.050 0957 2500 9697 E+0	±1.284 5390 4186 3882 E+1
	−1.863 5248 2714 3683 E+0	±1.550 4527 1478 1576 E+1
17	−1.159 3165 9340 4287 E+1	±8.686 4616 8436 8076 E−1
	−1.141 1608 3703 8859 E+1	±2.609 0245 9972 9372 E+0
	−1.104 1782 4754 7462 E+1	±4.359 1145 1048 6985 E+0
	−1.046 8928 1354 9631 E+1	±6.127 0752 1153 7584 E+0
	−9.666 8440 6698 5269 E+0	±7.924 3475 3444 3790 E+0
	−8.590 1889 3393 3273 E+0	±9.769 2245 2781 4408 E+0
	−7.155 0909 6553 5080 E+0	±1.169 6010 3243 1153 E+1
	−5.175 1394 3403 7279 E+0	±1.378 7917 8735 0248 E+1
	−1.897 2516 8410 6712 E+0	±1.648 3963 2214 2634 E+1
18	−1.227 7495 1850 5371 E+1	
	−1.219 2021 6470 1196 E+1	±1.737 7884 2829 7625 E+0
	−1.193 2833 6937 8446 E+1	±3.481 1674 1311 1369 E+0
	−1.149 1156 4451 9311 E+1	±5.236 4744 0667 7450 E+0
	−1.085 0579 3034 9564 E+1	±7.011 8039 3774 0712 E+0
	−9.983 4521 9406 3059 E+0	±8.818 7446 5485 6362 E+0
	−8.842 9706 5683 6907 E+0	±1.067 6006 8738 2377 E+1
	−7.343 4193 9478 9007 E+0	±1.261 8777 6517 3627 E+1
	−5.295 0434 2309 3039 E+0	±1.473 2309 3641 7574 E+1
	−1.929 7756 4920 5957 E+0	±1.746 4149 4500 2798 E+1
19	−1.291 9146 0509 1870 E+1	±8.684 7381 0966 5672 E−1
	−1.275 6405 7302 6832 E+1	±2.607 8900 2877 9320 E+0
	−1.242 6122 3917 0677 E+1	±4.355 0000 1921 2713 E+0
	−1.191 7937 0509 9271 E+1	±6.115 9794 8795 2964 E+0
	−1.121 4118 6127 7730 E+1	±7.898 9312 6822 3903 E+0
	−1.028 5806 5686 6302 E+1	±9.715 6302 0341 9265 E+0
	−9.084 9172 2160 5599 E+0	±1.158 5215 1411 0589 E+1
	−7.524 0515 5541 5614 E+0	±1.354 3738 0387 9310 E+1
	−5.410 2956 8963 4106 E+0	±1.567 8420 1370 8806 E+1
	−1.961 1991 2801 7860 E+0	±1.844 5022 3511 6019 E+1
20	−1.360 1347 5726 0897 E+1	
	−1.352 4234 2949 8953 E+1	±1.737 3652 5530 7674 E+0
	−1.329 0869 6856 1004 E+1	±3.479 2433 3706 6860 E+0
	−1.289 4901 4630 4205 E+1	±5.230 6300 7025 9606 E+0
	−1.232 4718 6155 4657 E+1	±6.997 6213 1060 6737 E+0
	−1.156 1506 3018 4626 E+1	±8.788 3761 0502 4624 E+0
	−1.057 5370 5250 7374 E+1	±1.061 4879 5712 2257 E+1
	−9.317 0897 9331 4458 E+0	±1.249 6702 8817 2940 E+1
	−7.697 7096 4406 2652 E+0	±1.447 0742 6402 7235 E+1

TABLE I. (*Continued*.)

n	Real	Imaginary
	−5.521 3146 5994 4756 E+0	±1.662 6122 5224 9505 E+1
	−1.991 6112 3026 5466 E+0	±1.942 6526 4976 0915 E+1
21	−1.424 5035 2509 5977 E+1	±8.683 4733 4783 2387 E−1
	−1.409 7563 5251 4625 E+1	±2.607 0625 5040 0543 E+0
	−1.379 9062 5417 6754 E+1	±4.352 0313 8410 5355 E+0
	−1.334 1965 7300 5157 E+1	±6.108 1093 1141 4979 E+0
	−1.271 3647 4432 9065 E+1	±7.881 3670 7433 2156 E+0
	−1.189 4383 7795 5032 E+1	±9.680 0518 7654 3492 E+0
	−1.085 3381 8669 2819 E+1	±1.151 6373 9614 5975 E+1
	−9.540 3936 2141 5492 E+0	±1.341 0335 9014 4012 E+1
	−7.865 0141 2713 0227 E+0	±1.539 9657 7487 4755 E+1
	−5.628 4620 7950 8090 E+0	±1.757 5303 3092 6353 E+1
	−2.021 0900 3425 1912 E+0	±2.040 8613 1938 8935 E+1
22	−1.492 5490 7825 0745 E+1	
	−1.485 5246 5278 9887 E+1	±1.737 0506 3191 9017 E+0
	−1.464 2987 4175 5074 E+1	±3.477 8229 6067 6930 E+0
	−1.428 3962 2081 7595 E+1	±5.226 3653 7856 1890 E+0
	−1.376 9629 1451 3085 E+1	±6.987 4531 0666 9800 E+0
	−1.308 6522 4278 9280 E+1	±8.767 1691 9588 4732 E+0
	−1.221 4141 1274 3256 E+1	±1.057 3870 5060 3804 E+1
	−1.112 0898 1171 3988 E+1	±1.242 0001 1673 1512 E+1
	−9.755 6078 5124 1861 E+0	±1.432 5991 2375 9705 E+1
	−8.026 5023 8686 6059 E+0	±1.633 0362 8505 0095 E+1
	−5.732 0530 1705 5048 E+0	±1.852 5861 6046 2295 E+1
	−2.049 7043 8236 1665 E+0	±2.139 1239 4141 3745 E+1
23	−1.557 0856 4908 1945 E+1	±8.682 5178 7095 8789 E−1
	−1.543 6025 1419 2300 E+1	±2.606 4402 4354 9175 E+0
	−1.516 3652 7795 9159 E+1	±4.349 8163 4891 8186 E+0
	−1.474 8035 8024 5967 E+1	±6.102 3083 9350 7367 E+0
	−1.417 9832 5516 2790 E+1	±7.868 6529 5916 5230 E+0
	−1.344 4868 4078 8119 E+1	±9.654 9714 3028 4560 E+0
	−1.252 1967 3267 8536 E+1	±1.146 9744 8136 4126 E+1
	−1.137 8830 7816 1521 E+1	±1.332 5655 7982 4416 E+1
	−9.963 4084 3606 5207 E+0	±1.524 3556 1342 0046 E+1
	−8.182 6432 3077 5507 E+0	±1.726 2748 9634 1421 E+1
	−5.832 3637 2999 3578 E+0	±1.947 7707 0764 7581 E+1
	−2.077 5153 2046 2335 E+0	±2.237 4366 9420 9306 E+1
24	−1.624 9853 5202 6011 E+1	
	−1.618 5353 6824 7834 E+1	±1.736 8103 2799 0838 E+0
	−1.599 0674 9184 0214 E+1	±3.476 7438 6008 9849 E+0
	−1.566 2168 1134 1328 E+1	±5.223 1529 2712 3997 E+0
	−1.519 3353 9046 3363 E+1	±6.979 8907 6524 7421 E+0
	−1.457 4222 2963 9613 E+1	±8.751 6847 2456 7659 E+0
	−1.378 9990 7788 8729 E+1	±1.054 4712 8384 3300 E+1
	−1.281 8888 3193 8241 E+1	±1.236 7589 9474 5815 E+1
	−1.162 7971 7142 9776 E+1	±1.423 3239 1918 6804 E+1
	−1.016 4385 9995 5295 E+1	±1.616 2927 0669 8962 E+1
	−8.333 8483 2778 0100 E+0	±1.819 6717 2163 8913 E+1
	−5.929 6379 2121 6811 E+0	±2.043 0758 4996 7550 E+1
	−2.104 5772 6342 9114 E+0	±2.335 7961 6688 4198 E+1
25	−1.689 6625 6143 2097 E+1	±8.681 7784 6123 4850 E−1
	−1.677 2433 0413 4704 E+1	±2.605 9603 1648 3995 E+0
	−1.652 1935 9001 1347 E+1	±4.348 1182 2774 0866 E+0
	−1.614 0723 8137 4977 E+1	±6.097 9014 7165 5553 E+0
	−1.562 1673 9503 4242 E+1	±7.859 1203 5051 0501 E+0
	−1.495 4208 7838 4365 E+1	±9.636 5140 9894 9407 E+0
	−1.412 3016 8996 0779 E+1	±1.143 6330 3800 0604 E+1
	−1.310 5796 3163 7697 E+1	±1.326 7324 2045 1941 E+1
	−1.186 9013 6960 2609 E+1	±1.514 2659 1520 2252 E+1
	−1.035 9059 9404 7652 E+1	±1.708 4008 8424 1078 E+1
	−8.480 4813 2683 7337 E+0	±1.913 2177 6336 4088 E+1
	−6.024 0917 6938 8539 E+0	±2.138 4942 5455 9088 E+1
	−2.130 9389 4756 9443 E+0	±2.434 1993 0148 6750 E+1

C. The transient harmonic multipole, $a_n(t) = \mathrm{H}(t)\sin(\Omega t)$

The transient modal solution generated by a sinusoidal surface excitation of unit amplitude and frequency Ω starting from rest at time $t=0$ will be termed the *transient harmonic multipole solution* and denoted by $u_n^{\Omega}(r,t)$. It is obtained by setting $\overline{a_n}(s)=\Omega/(\Omega^2+s^2)$ and proceeding as before. This gives,

for n odd,

$$u_n^{\Omega}(r,t)=\rho c(R/r)\mathrm{H}(t')[\Upsilon(r,t)+\Lambda(r,t)], \tag{22a}$$

and for n even

$$u_n^{\Omega}(r,t)=\rho c(R/r)\mathrm{H}(t')[C(r)e^{a_0 t'}+\Upsilon(r,t)+\Lambda(r,t)], \tag{22b}$$

where

$$\Upsilon(r,t)=X(r)\sin(\Omega t')+Y(r)\cos(\Omega t'). \tag{23}$$

The real coefficients $A_j(r)$, $B_j(r)$, $X(r)$, $Y(r)$ and $C(r)$ are given by,

for n odd,

$$A_j(r)+iB_j(r)=[F_n(r,s)/b_j(s^2+\Omega^2)\Phi(s)]_{s=a_j+ib_j}, \tag{24a}$$

$$X(r)+iY(r)=[F_n(r,s)/\Omega\Psi(s)]_{s=i\Omega}, \tag{24b}$$

and for n even

$$A_j(r)+iB_j(r)=[F_n(r,s)/b_j(s^2+\Omega^2)\times(s-a_0)\Phi(s)]_{s=a_j+ib_j}, \tag{24c}$$

$$X(r)+iY(r)=[F_n(r,s)/\Omega(s-a_0)\Psi(s)]_{s=i\Omega}, \tag{24d}$$

$$C(r)=[F_n(r,s)/(s^2+\Omega^2)\Psi(s)]_{s=a_0}. \tag{24e}$$

D. The general transient multipole, $a_n(t)=f(t)$

The preceding solution is a particular case of the general solution due to an arbitrary transient surface excitation of the form $a_n(t)=f(t)$. This can be written in terms of the unit impulse solution by means of a convolution integral of the form

$$u_n(r,t)=\int_0^t f(\tau)\hat{u}_n(r,t-\tau)\,d\tau. \tag{25}$$

A discrete approximation to this relationship is given by

$$u_n(r,t)\simeq\sum_k f(t_k)\hat{u}_n(r,t-t_k)\Delta t, \tag{26}$$

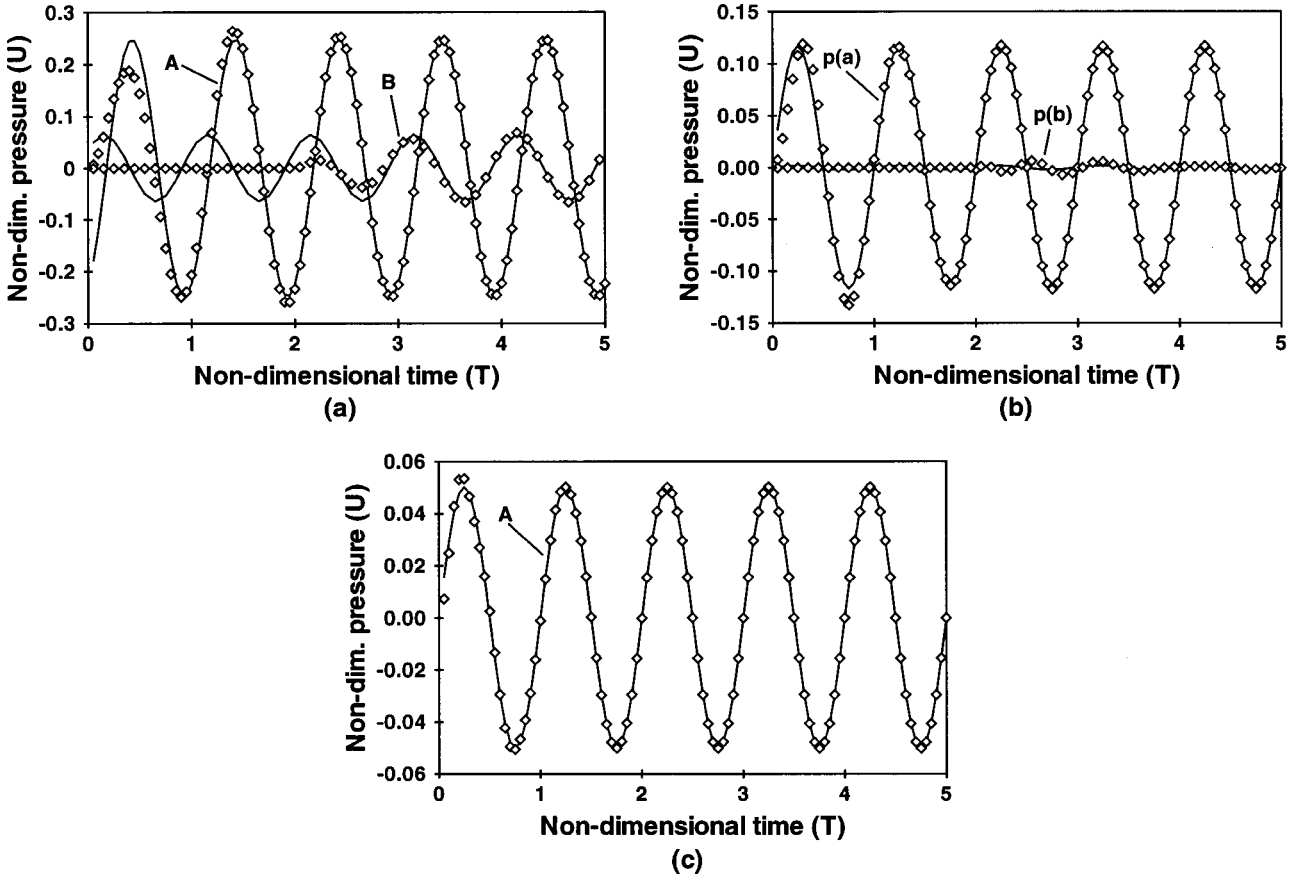


FIG. 2. Transient multipole solutions; ◇ exact, ——— steady. (a) $n=5$. (b) $n=10$. (c) $n=20$.

where $t_k(=k\Delta t)$ is a discrete point in the time history. The accuracy of this approximation will clearly be governed by the size of the time increment Δt, which can be reduced until convergence is achieved.

E. Transient radiation from a piston in a spherical baffle

More complex transient solutions can be formed by superposition of two or more of the previous multipole components. Consider, for example, the case of a spherical ''piston'' which occupies the region $0\leqslant\theta\leqslant\theta_2$ on the surface of a spherical baffle of radius R. The piston experiences a transient normal acceleration of the form

$$a(\theta,t)=g(\theta)f(t), \tag{27}$$

where

$$g(\theta)=\begin{cases} 1, & 0\leqslant\theta\leqslant\theta_1, \\ \dfrac{\cos\theta-\cos\theta_2}{\cos\theta_1-\cos\theta_2}, & \theta_1<\theta\leqslant\theta_2, \\ 0, & \theta_2<\theta\leqslant\pi. \end{cases} \tag{28}$$

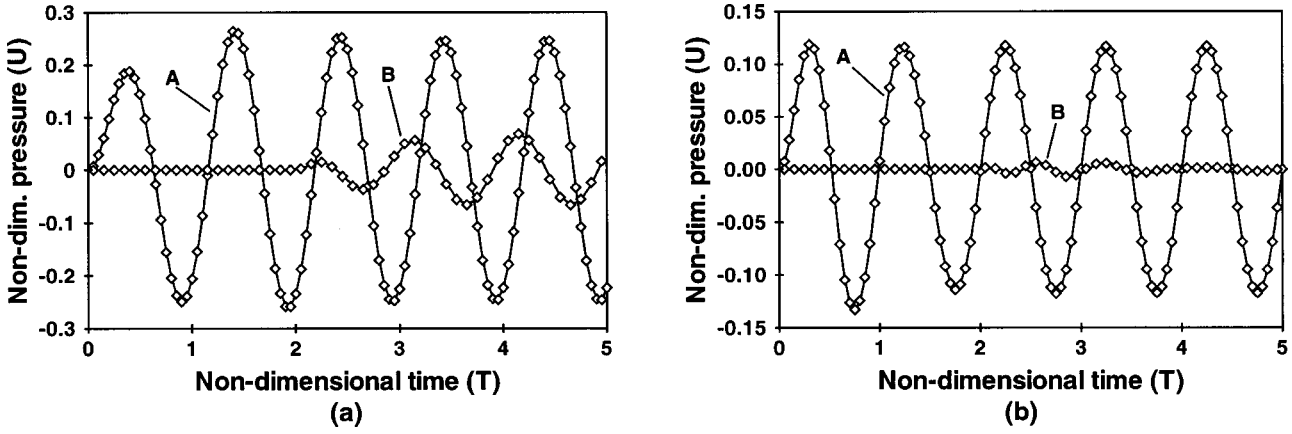


FIG. 3. Transient multipole solutions; ◇ exact, ——— FE. (a) $n=5$. (b) $n=10$.

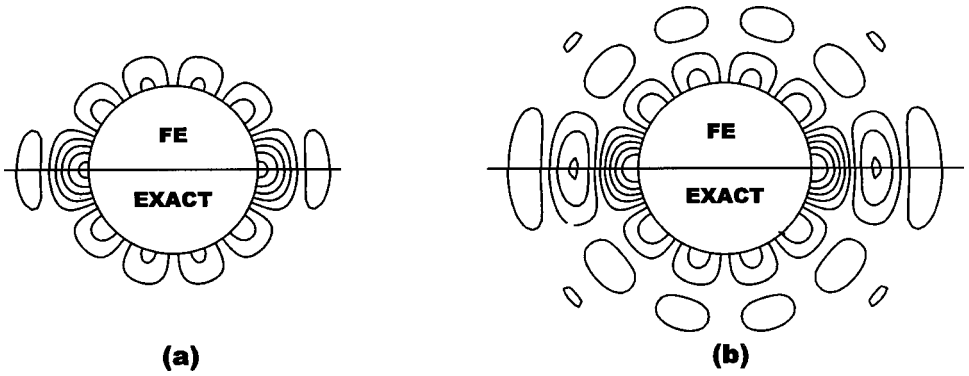


FIG. 4. Multipole order 5, comparison exact and FE solutions. (a) $T=1$. (b) $T=2$.

An expansion of $g(\theta)$ in terms of the transverse surface modes gives

$$a(\theta,t)=\sum_{n=0}^{\infty} A_n f(t) P_n(\cos\theta), \tag{29}$$

where the coefficients A_n are given by

$$A_0=\tfrac{1}{4}(2-x_1-x_2), \tag{30a}$$

$$A_1=\frac{1}{4}\left[3-3x_1^2+\frac{1}{x_1-x_2}(2x_1^3-3x_1^2x_2+x_2^3)\right], \tag{30b}$$

$$A_n=\frac{1}{2(x_1-x_2)(2n+3)}[P_n(x_1)-P_n(x_2)-P_{n+2}(x_1)+P_{n+2}(x_2)]+\frac{1}{2(x_1-x_2)(2n-1)}[P_n(x_1)-P_n(x_2)-P_{n-2}(x_1)+P_{n-2}(x_2)], \quad n\geqslant 2, \tag{30c}$$

where $x_1=\cos(\theta_1)$ and $x_2=\cos(\theta_2)$. The corresponding transient solution is then given by a summation of multipole terms of the form

$$u(r,\theta,t)=\sum_{n=0}^{\infty} A_n u_n(r,t) P_n(\cos\theta), \tag{31}$$

where $u_n(r,t)$ is given in the general case by Eq. (25) or in the case of harmonic excitation from rest by $u_n^{\Omega}(r,t)$ [see Eq. (22a) or (22b)]. The exact steady harmonic solution is given by

$$u(r,\theta,t)=-\operatorname{Im}\left\{e^{i\omega t}\sum_{n=0}^{\infty} A_n \frac{\rho h_n^{(2)}(kr)}{k h_n'^{(2)}(kR)} P_n(\cos\theta)\right\}, \tag{32}$$

where $h_n^{(2)}(\ )$ and $h_n'^{(2)}(\ )$ are the spherical Bessel function of the second kind and its derivative. This has been used extensively as a benchmark problem for time-harmonic formulations and for transient formulations at large times.[27,28] Both solutions—steady and transient—form challenging validation problems for numerical schemes. Although a discontinuity in the surface acceleration has been avoided in the current instance by smoothing the piston acceleration over the interval $\theta_1\leqslant\theta\leqslant\theta_2$, large gradients are still present in the vicinity of the surface and many terms are needed in Eq. (29) to give a smooth approximation to the surface acceleration. Césaro summation is frequently used to mitigate this effect in problems where real discontinuities occur on the surface—scattering of a step pressure pulse for example[18,20,22]—but has not been used in the current instance.

VI. RESULTS AND VERIFICATION

Solutions obtained by evaluating the closed form analytic expressions of the preceding section are now presented. They are compared to steady frequency-domain solutions to demonstrate their correctness and accuracy at large times. At

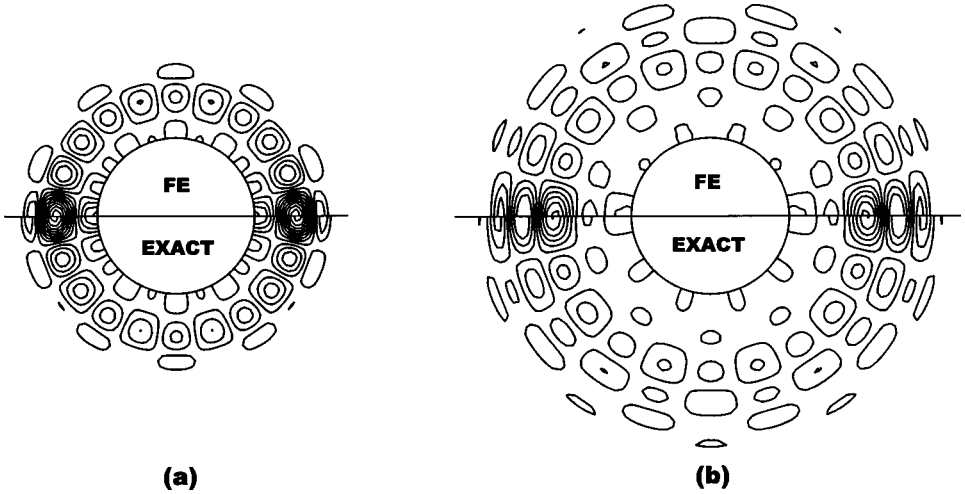


FIG. 5. Multipole order 10, comparison exact and FE solutions. (a) $T=1$. (b) $T=2$.

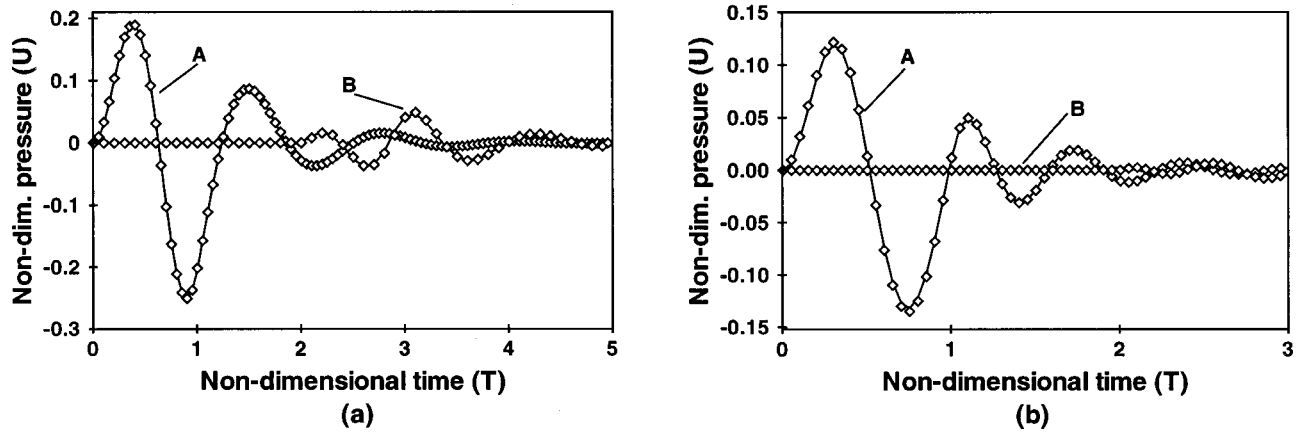


FIG. 6. Transient multipole solutions, response to a single sinusoidal pulse; ◇ exact, ——— FE. (a) $n=5$. (b) $n=10$.

finite elapsed times, their validity is established with reference to high-order numerical schemes for which high levels of confidence have already been established.

A. Multipole radiation

1. Transient harmonic excitation. Convergence to the steady solution at large times

Transient multipole solutions, $u_n^{\Omega}(r,t)$, are presented in Fig. 2 for multipoles of order $n=5$, 10, and 20. A sinusoidal excitation of frequency Ω commences at time $t=0$ where $\Omega R/c=2\pi$. Time histories are presented as plots of nondimensional dependent versus independent variables U $(=u/\rho c^2)$ and $T(=ct/R)$. These are shown at $r/R=1$ (point A) and at $r/R=3$ (point B) on the positive z axis. The response at B is not plotted for the case $n=20$ [Fig. 2(c)] due to the very large rate of radial attenuation at high multipole orders which renders the solution at B negligibly small to the scale shown. The exact steady solution is also shown. Clearly, the transient solutions converge quite rapidly to the steady case, the two responses becoming indistinguishable in each case after one or two cycles.

2. Transient harmonic excitation. Comparison with numerical results

Analytic time histories for multipoles of order 5 and 10 are plotted in Fig. 3 and compared to high-order transient finite/infinite element (FE/IE) solutions.[15] The latter are computed using infinite elements of order 6 and 11 which are known to accurately model multipole components of orders 5 and 10.[11] Excellent correspondence exists between the exact and numerical solutions. Contours of the exact and numerical solutions are plotted in Figs. 4 and 5 at $T=1$ and 2 over the region $R\leqslant r\leqslant 3R$. These indicate again a close correspondence between the exact and numerical solutions over the entire exterior field.

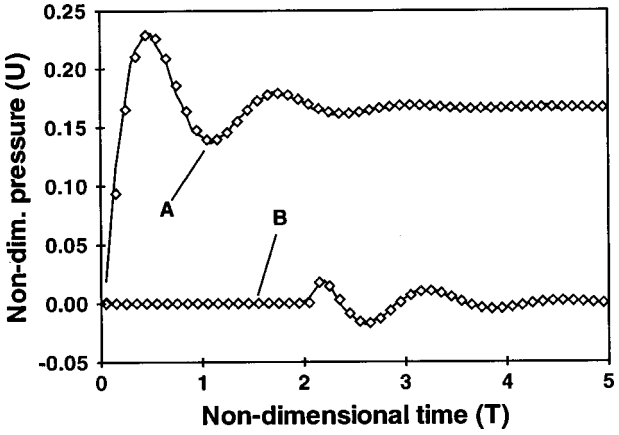


FIG. 7. Transient multipole solutions, response to a step input, $n=5$; ◇ exact, ——— FE.

3. Transient sinusoidal pulse. Comparison with numerical results

Time histories at points A and B are shown in Fig. 6 for a single sinusoidal pulse defined by a surface acceleration

$$a_n(t)=\begin{cases}\sin(\Omega t), & 0\leqslant t\leqslant 2\pi/\Omega,\\ 0, & \text{otherwise,}\end{cases} \tag{33}$$

where again $\Omega R/c=2\pi$. The solution can be obtained in this case by two different approaches. It can be regarded as the superposition of two transient harmonic solutions, in which case

$$u_n(r,t)=u_n^{\Omega}(r,t)-u_n^{\Omega}(r,t-2\pi/\Omega), \tag{34}$$

or as a convolution integral in terms of the impulse response solution $\hat{u}_n(r,t-\tau)$ which gives

$$u_n(r,t)=\int_0^{2\pi/\Omega}\sin(\Omega\tau)\hat{u}_n(r,t-\tau)\,d\tau. \tag{35}$$

Both approaches have been used. The resulting solutions are indistinguishable provided that a sufficiently large number of integration points is used when evaluating Eq. (35). The results shown in Fig. 6 are obtained by using Eq. (35) with 200 integration points on the interval $[0,2\pi/\Omega]$. Time histories are presented for multipoles of order 5 and 10. The correspondence between the exact and numerical FE/IE solutions is again close, confirming the effectiveness of the impulse response solution as a means of generating more general solutions for arbitrary temporal variations in the normal surface acceleration.

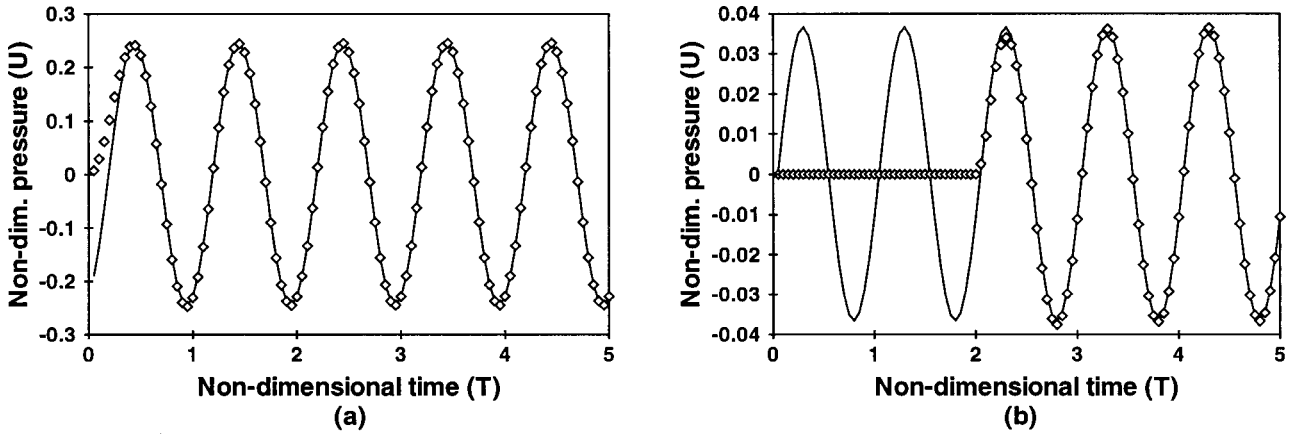


FIG. 8. Spherical piston radiation, results at $\theta=0$ degrees. (a) $r/R=1$. (b) $r/R=3$; ◇ series, ——— steady.

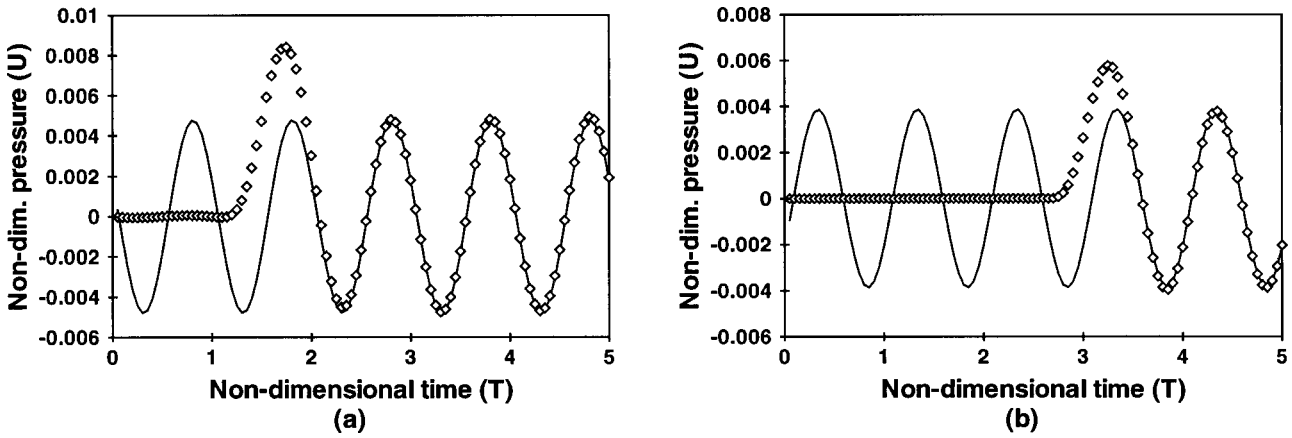


FIG. 9. Spherical piston radiation, results at $\theta=90$ degrees. (a) $r/R=1$. (b) $r/R=3$; ◇ series, ——— steady.

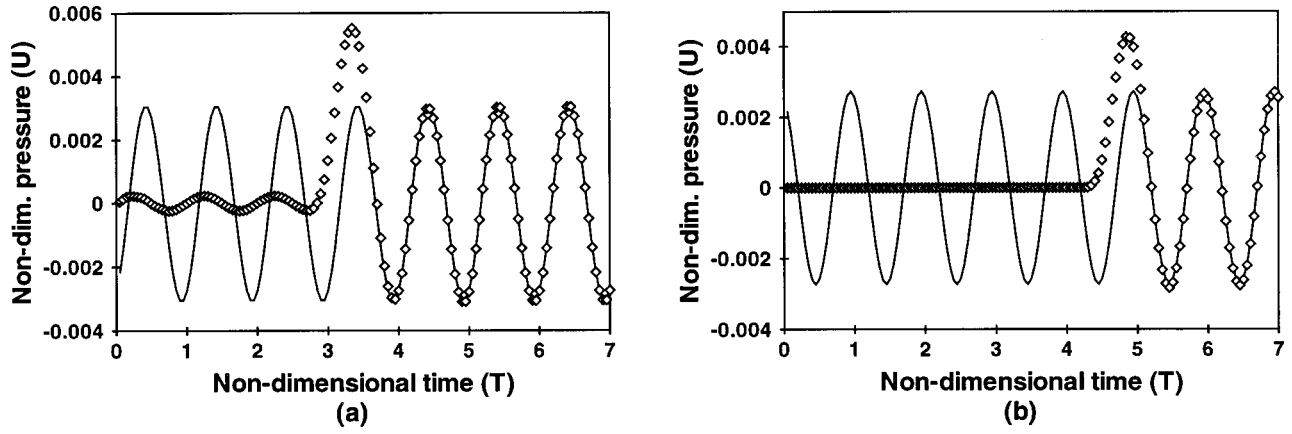


FIG. 10. Spherical piston radiation, results at $\theta=180$ degrees. (a) $r/R=1$. (b) $r/R=3$; ◇ series, ——— steady.

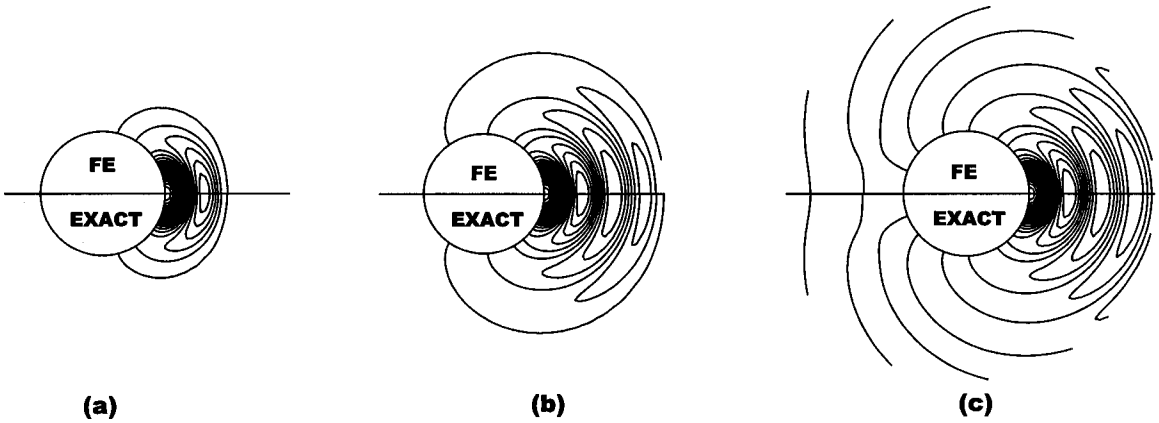


FIG. 11. Spherical piston radiation, comparison of FE and series solutions. (a) $T=1$. (b) $T=2$. (c) $T=4$.

4. Transient step response. Comparison with numerical results

Time histories at points A and B are shown in Fig. 7 for the transient step response, $u_n^{\mathrm{H}}(r,t)$. The results for a multipole of order 5 are shown, where the numerical solution used infinite elements of order 6. The correspondence between the two solutions is again close.

B. Radiation from a piston in a spherical baffle

Solutions are presented for the case of a spherical piston for which $\theta_1=15$ degrees and $\theta_2=30$ degrees. The piston experiences a sinusoidal excitation of frequency Ω, where $\Omega R/c=2\pi$, commencing at time $t=0$. Both the transient and steady series solutions [Eqs. (31) and (32)] are truncated after 25 terms. Pressure histories at $r/R=1$ and $r/R=3$ are shown in Figs. 8–10 for $\theta=0$, 90, and 180 degrees. The transient solution converges to the steady result at all points after one or two cyles.

An apparent inconsistency in the transient solution lies in the appearance of small spurious oscillations at the ''back'' of the baffle ($\theta=180$ degrees). These clearly precede the arrival of the first disturbance from the piston at $T=\pi-\theta_2$. Although physically unacceptable in the problem as originally specified, this behavior is consistent with the initial decomposition of the surface acceleration into a finite sum of multipole components. The effect can be reduced by increasing the number of terms in the surface expansion. When 30 terms are used, for example, the oscillations are virtually eliminated to the scale shown. However, this phenomenon causes no difficulty in benchmark solutions provided that the same truncated form of the surface acceleration is used as input to the numerical scheme being tested.

Contour plots of a FE/IE solution obtained using infinite elements of order 10 are shown in Fig. 11 and compared to the current series transient solution at $T=1$, 2, and 4. In both cases 25 terms are taken in the expansion of the surface acceleration. Although these solutions are obtained by quite different means, the correspondence is extremely close. Both solutions are to a large extent validated by each other in this comparison.

VII. CONCLUSIONS

(1) A family of exact transient solutions for axisymmetric spherical radiation has been derived as a verification tool for assessing the accuracy of numerical methods.
(2) Specific multipole solutions have been derived and presented for the sinusoidal and step excitations starting from rest and for more general inputs using an impulse response function. Solutions are presented in a form which permits their reconstruction for multipole components in the range [0,25].
(3) Multipole transient solutions have been validated by comparison with steady solutions—at large times—and with high-order numerical schemes.
(4) A novel benchmark solution has been derived for the transient sound field generated by the motion of a spherical piston in a rigid spherical baffle. This is constructed by superposition of multipole transient components and is shown to be in close agreement both with the steady solution at large times and with high-order transient infinite element computations at finite elapsed times.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the support of The Royal Society of New Zealand, Marsden Fund, Contract/grant No. UOC-907 ''Numerical solutions of unbounded wave problems.''

APPENDIX A: INCLUSION OF OTHER BOUNDARY CONDITIONS

One may construct a more general boundary condition of the form

$$\alpha\nabla u\cdot\underline{n}+\beta u=a(\underline{x},t),\quad \text{for } \underline{x}\in S. \tag{A1}$$

If $\beta=0$ a Neumann condition is described as before and if $\alpha=0$, a Dirichlet condition results where $a(\underline{x},t)$ is now a prescribed acoustic pressure. If both terms are nonzero, then a mixed or Robin condition is imposed. This may be considered as a simple form of a transient acoustic impedance condition in which the impedance is independent of frequency. The procedure for the implementation of the Dirichlet condition will now be described, however, the general nature of

the Robin condition makes a general solution impracticable since it will depend upon specific values of α and β.

After the prescribed acoustic pressure, $a(\underline{x},t)$, is decomposed into modal components the solution for $\overline{u_n}(r,s)$ subject to the Dirichlet condition is given by

$$\overline{u_n}(r,s)=\overline{a_n}(s)k_n(sr/c)/k_n(sR/c). \tag{A2}$$

Substitution of Eqs. (13a) and (15) yields

$$\overline{u_n}(r,s)=e^{-sx}\overline{a_n}(s)F_n(r,s)/H_n(s), \tag{A3}$$

where x and $F_n(r,s)$ are as per the Neumann case and $H_n(s)$ is a polynomial of the form $H_n(s)=\lambda_0 s^n+\lambda_1 s^{n-1}+\ldots+\lambda_n$ with coefficients $\lambda_i=\gamma_i(c/2R)^i$ [see Eq. (15) for the definition of γ_i].

The solution of Eq. (A3) is obtained in the same way as the Neumann problem. The inverse Laplace transform is obtained by performing a partial fraction expansion, the roots of $H_n(S)$ being central to this. The roots are generated using the same procedure as for $G_n(S)$ and may be obtained from the corresponding author.

[1] T. L. Geers, ''Benchmark Problems,'' in *IUTAM Symposium on Computational Methods for Unbounded Domains*, edited by T. L. Geers (Kluwer Academic, Dordrecht, 1998), pp. 1–9.

[2] A. Bayliss and E. Turkel, ''Radiation Boundary Conditions for Wave Like Equations,'' Commun. Pure Appl. Math. **33**, 702–725 (1980).

[3] A. Bayliss, M. Gunzberger, and E. Turkel, ''Boundary Conditions for the Numerical Solution of Elliptical Equations in Exterior Regions,'' SIAM (Soc. Ind. Appl. Math.) J. Appl. Math. **42**, 430–450 (1982).

[4] D. Givoli, ''A Spatially Exact Nonreflecting Boundary Condition for Time Dependent Problems,'' Comput. Methods Appl. Mech. Eng. **95**, 97–113 (1992).

[5] R. J. Astley, ''FE Mode-Matching Schemes for the Exterior Helmholtz Problem and their Relationship to the FE-DtN Approach,'' Commun. Numer. Methods Eng. **12**, 257–267 (1996).

[6] P. Bettess, *Infinite Elements* (Penshaw, Sunderland, UK, 1992).

[7] C. H. Wilcox, ''A Generalization of the Theorems of Rellich and Atkinson,'' Proc. Am. Math. Soc. **7**, 271–276 (1956).

[8] D. S. Burnett, ''A Three-dimensional Acoustic Infinite Element Based on a Prolate Spheroidal Multipole Expansion,'' J. Acoust. Soc. Am. **96**, 2798–2816 (1994).

[9] D. S. Burnett and R. L. Holford, ''An Ellipsoidal Acoustic Infinite Element,'' Comput. Methods Appl. Mech. Eng. **164**, 49–76 (1998).

[10] R. J. Astley, ''Mapped Spheroidal Elements for Unbounded Wave Problems,'' Int. J. Numer. Methods Eng. **41**, 1235–1254 (1998).

[11] R. J. Astley, ''Infinite Elements for Wave Problems: A Review of Current Formulations and an Assessment of Accuracy,'' Int. J. Numer. Methods Eng. **49**, 951–976 (2000).

[12] P. M. Morse and K. M. Ingard, *Theoretical Acoustics* (McGraw–Hill, New York, 1968).

[13] M. Abramowicz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1970).

[14] L. G. Olsen and K-J. Bathe, ''An Infinite Element for the Analysis of Fluid-Structure Interaction,'' Eng. Comput. **2**, 319–329 (1985).

[15] R. J. Astley, J-P. Coyette, and L. Cremers, ''Three-dimensional Wave-Envelope Elements of Variable Order for Acoustic Radiation and Scattering. Part 2. Formulation in the Time Domain,'' J. Acoust. Soc. Am. **103**, 64–72 (1998).

[16] L. L. Thompson and R. Huan, ''Computation of Transient Radiation in Semi-infinite Regions Based on Exact Nonreflecting Boundary Conditions and Mixed Time Integration,'' J. Acoust. Soc. Am. **106**, 3095–3108 (1999).

[17] L. F. Kallivokas, J. Bielak, and R. C. MacCamy, ''A Simple Impedance-Infinite Element for the Finite Element Solution of the Three-Dimensional Wave Equation in Unbounded Domains,'' Comput. Methods Appl. Mech. Eng. **147**, 235–262 (1997).

[18] P. Zhang and T. L. Geers, ''Excitation of a Fluid Filled Submerged Spherical Shell by a Transient Acoustic Wave,'' J. Acoust. Soc. Am. **93**, 696–705 (1993).

[19] T. L. Geers and T. Ju, ''A Computer Program for a Canonical Program in Underwater Shock,'' Shock and Vibration **1**, 331–337 (1994).

[20] H. Huang and G. C. Gurnaurd, ''Transient, Diffraction of a Plane Step Wave by a Hard Sphere: Neoclassical Solution,'' J. Acoust. Soc. Am. **104**, 3236–3244 (1998).

[21] H. Huang, ''Transient Interaction of Plane Acoustic Waves with a Spherical Elastic Shell,'' J. Acoust. Soc. Am. **45**, 661–670 (1969).

[22] H. Huang and H. U. Mair, ''Neoclassical Solution of Transient Interaction of Plane Acoustic Waves with a Spherical Elastic Shell,'' Shock and Vibration **3**, 85–98 (1996).

[23] H. Huang, ''Transient Response of Two Fluid-coupled Spherical Elastic Shells to an Incident Pressure Pulse,'' J. Acoust. Soc. Am. **65**, 881–887 (1979).

[24] R. T. Menton and E. B. Magrab, ''Far-field Radiation of Internally Generated Transient Acoustic Pulses Transmitted Through a Spherical Shell,'' J. Acoust. Soc. Am. **55**, 237–246 (1974).

[25] H-W. Chen and P. Stepanishen, ''Acoustic Transient Radiation from Fluid-loaded Shells of Revolution Using Time-dependent *in vacuo* Eigenvector Expansions,'' J. Acoust. Soc. Am. **95**, 601–616 (1994).

[26] P. Stepanishen and H-W. Chen, ''Acoustic Time-dependent Loading on Elastic Shells of Revolution Using the Internal Source Density and Singular Value Decomposition Method,'' J. Acoust. Soc. Am. **99**, 1913–1923 (1996).

[27] P. M. Pinsky and N. N. Abboud, ''Transient Finite Element Analysis of the Exterior Structural Acoustics Problem,'' ASME J. Vibr. Acoust. **112**, 245–256 (1990).

[28] R. Huan and L. L. Thompson, ''Accurate Radiation Boundary Conditions for the Time Dependent Wave Equation on Unbounded Domains,'' Int. J. Numer. Methods Eng. **47**, 1569–1603 (2000).

[29] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in FORTRAN* (Cambridge U.P., New York, 1992), 2nd ed.

[30] D. M. Smith, ''Multiple Precision Complex Arithmetic and Functions,'' ACM Trans. Math. Softw. **24**, 359–367 (1998).

Energy losses in an acoustical resonator

Yurii A. Ilinskii,[a] Bart Lipkens,[b] and Evgenia A. Zabolotskaya[a]
MacroSonix Corp., 1570 East Parham Road, Richmond, Virginia 23228

(Received 6 February 2000; accepted for publication 23 January 2001)

A one-dimensional model has recently been developed for the analysis of nonlinear standing waves in an acoustical resonator. This model is modified to include energy losses in the boundary layer along the resonator wall. An investigation of the influence of the boundary layer on the acoustical field in the resonator and on the energy dissipation in the resonator is conducted. The effect of the boundary layer is taken into account by introducing an additional term into the continuity equation to describe the flow from the boundary layer to the volume. A linear approximation is used in the development of the boundary layer model. In addition to the viscous attenuation in the boundary layer, the effect of acoustically generated turbulence is modeled by an eddy viscosity formulation. Calculatons of energy losses and a quality factor of a resonator are included into the numerical code. Results are presented for resonators of three different geometries: a cylinder, a horn cone, and a bulb-type resonator. A comparison of measured and predicted dissipation shows good agreement.
 [DOI: 10.1121/1.1359798]

PACS numbers: 43.25.Gf [MFH]

I. INTRODUCTION

The resonant macrosonic synthesis technology developed at MacroSonix allows the creation of high-amplitude, shock-free standing waves in acoustical resonators.[1] Power is delivered to the resonator by a linear motor that shakes the entire cavity. We have observed peak pressures that exceed the ambient pressure by factor of 3 to 4. The standing waves are strongly nonlinear and phenomena such as harmonic generation, resonance frequency shift, and hysteresis are observed. Previously, a one-dimensional model has been developed to analyze the high-amplitude standing waves in acoustical resonators.[2] The only attenuation taken into account in this model is that caused by viscosity in the volume of the resonator. It is known that most energy losses occur in the acoustic boundary layer along the resonator wall. When an accurate computation of the energy dissipation is needed, it is necessary to take into account the energy dissipation in the boundary layer. It may be considered an oversimplification to analyze the boundary layer motion within a one-dimensional model because the fluid motion in a resonator with a boundary layer is no longer one-dimensional. However, if the transverse component of the particle velocity is small, a one-dimensional treatment of the standing waves is still valid.[3]

Here, the effects of the boundary layer on nonlinear standing waves as well as energy dissipation are considered within the one-dimensional model. Energy losses generated by the turbulence are included. Three steps should be distinguished in this investigation. First, we assume that the boundary layer does not affect the acoustical field in the resonator. Therefore, all the acoustical characteristics, i.e., pressure and velocity, are calculated with the one-dimensional model[2] that includes volume absorption only. Then, the model is completed by adding a linear equation for the boundary layer. The linear equation uses the results of the volume model to calculate the velocity profile in the boundary layer as well as the energy dissipation and the quality factor of the resonator.

Second, the influence of the boundary layer on the acoustical field in the resonator is taken into account. Chester[3] introduced a concept that takes into account the influence of the boundary layer on the mass conservation equation. Following this concept, we include an additional term into the continuity equation that describes the mass flow from the boundary layer to the volume. A new nonlinear model equation is then derived in the time domain to describe the high-amplitude standing waves in the resonator with the boundary layer. The time-domain equation is replaced by a set of spectral equations for the harmonic amplitudes. All acoustical characteristics are calculated with the new model equation. The energy losses in the boundary layer are calculated within the linear approximation of the model.

Finally, rough estimations are done to take into account the energy dissipation caused by the acoustically generated turbulence in the boundary layer. A simple eddy viscosity model is applied to include the losses associated with the turbulence. In this model the eddy viscosity is related to the acoustic Reynolds number based on the boundary layer thickness.

The new model equation is solved numerically for three types of resonators, a cylinder, a horn cone, and a bulb resonator. The distortion of the pressure and velocity waveforms are calculated and demonstrated for different points along the resonator axis. The distributions of the harmonic amplitudes and phases of the pressure and velocity waves along the resonator are shown for the first three harmonics. The dependence of the energy losses in the boundary layer and the quality factor on resonator shape is also presented.

[a]Also at: Department of Mechanical Engineering, University of Texas at Austin, Austin, TX 78712-1063.
[b]Also at: Mechanical Engineering, Virginia Commonwealth University, 601 West Main, P.O. Box 843015, Richmond, VA 23284-3015.

II. ENERGY LOSSES IN THE BOUNDARY LAYER

In this section the following assumptions are used: (1) the gas motion in the boundary layer does not affect the properties of the standing waves in the resonator; and (2) a linear equation is sufficient to describe the boundary layer motion.

A. Basic equations and solutions

The absorption due to tube wall boundary layer effects is well known. Blackstock[4] presents a discussion of the absorption of sound by viscous shear forces and heat transfer at walls of a tube. In this section we summarize some of the results. The momentum equation for the boundary layer of a viscous fluid in the linear approximation is

$$\frac{\partial u_z}{\partial t} = \nu_0 \frac{\partial^2 u_z}{\partial y^2} - \frac{1}{\rho_0}\frac{\partial p}{\partial z}, \tag{1}$$

where u_z is the particle velocity component along the resonator wall, p is total pressure, z is the coordinate along the resonator wall, and y is the coordinate normal to the resonator wall, $\nu_0 = \eta_0/\rho_0$ is the kinematic viscosity coefficient, ρ_0 is the equilibrium density of the gas, and η_0 is the dynamic viscosity. The solution of Eq. (1) is given by

$$u_z = \frac{1}{2}\sum_n u_n e^{in\omega t} + \text{c.c.}, \tag{2}$$

where u_n are the complex amplitudes of the harmonic components of the velocity wave, and c.c. are their complex conjugates. The amplitudes of the harmonic velocity components are given by

$$u_n = u_n^{(0)}(1 - e^{-\alpha y}), \tag{3}$$

where α is equal to

$$\alpha = \sqrt{\frac{in\omega\rho_0}{\eta_0}}, \tag{4}$$

and $u_n^{(0)}$ are the mainstream harmonic velocities that are parallel to the wall. They represent the sources for the boundary layer.

It is necessary to clarify that in this derivation we assume that the difference in the velocity component along the resonator axis and along the resonator wall is small. The same assumption applies for the coordinates along the wall and the resonator axis.

A similar derivation can be done for the heat transfer at the walls,[4] and the result is a thermal boundary layer with a temperature profile. It is possible to derive a quasi-plane-wave equation that includes the combined effect of viscosity and thermal conduction. Including the heat transfer at the wall amounts to replacing the factor $\sqrt{\eta_0}$ by $\sqrt{\eta_0}[1+(\gamma-1)/\sqrt{\text{Pr}}]$, where γ is the ratio of specific heats and Pr is the Prandtl number. In this paper we focus our discussion on the effect of viscosity.

B. Energy dissipation

Once the velocity in the boundary layer is known, the energy dissipation can be calculated. For a one-dimensional fluid motion, the energy dissipation per unit time and per unit volume equals[5]

$$q = \eta_0\left(\frac{\partial u_z}{\partial y}\right)^2. \tag{5}$$

After substituting Eq. (2) into Eq. (5), we obtain

$$q = \frac{\eta_0}{2}\sum_n \left(\frac{\partial u_n}{\partial y}\frac{\partial u_n^*}{\partial y}\right), \tag{6}$$

where u_n^* is a complex conjugate to u_n.

To calculate the energy that is dissipated in the boundary layer, Eq. (6) should be integrated over the volume. The equation for the energy losses can be presented in the form

$$W = \frac{\sqrt{2\eta_0\rho_0}}{4}\int_0^l \sum_n |\tilde{u}_n|^2\sqrt{n\omega}\,2\pi r(1+r_1^2)^{3/2}dx, \tag{7}$$

where $r_1 = dr/dx$, l is the length of the resonator, and $\tilde{u}_n$ is the axial component of velocity obtained with the one-dimensional model. To derive Eq. (7), the following equations were applied:

$$dS = 2\pi r\,dz = 2\pi r\frac{dx}{\cos\theta}, \tag{8}$$

$$u_n^{(0)} = \frac{\tilde{u}_n}{\cos\theta}, \tag{9}$$

$$\cos\theta = \frac{1}{\sqrt{1+r_1^2}}. \tag{10}$$

Here, x is the coordinate along the axis, r is the radius of the resonator, and θ is the angle between the wall and the axis. Equation (7) is the expression for the energy that is dissipated in the boundary layer due to viscosity.

As mentioned in the assumptions, we do not directly include the effect of thermal conductivity on the energy dissipation. Swift[6] included the effect of viscosity and thermal conductivity on energy dissipation for a cylindrical resonator. His calculations show that the energy dissipation is proportional to the factor $\sqrt{\eta_0}[1+(\gamma-1)/\sqrt{\text{Pr}}]$. We can draw two conclusions regarding the effect of thermal conductivity on energy dissipation for shaped resonators. First, we can approximate the effect of thermal conductivity by increasing the value of viscosity to account for the effect of thermal conductivity. Second, our application of interest is that of acoustic compressors with refrigerants as working gases. For most refrigerants, the value of γ is lower than that for air; it is in the range of 1.1–1.2, an indication that the effect of thermal conductivity is relatively small in comparison with that of viscosity. In numerical calculations, we adjust the value of viscosity so that a good agreement is obtained between calculated and measured dissipation.

C. Preparation for numerical calculations

As the energy losses are calculated numerically, we introduce the following dimensionless variables:[2]

$$X=\frac{x}{l},\quad T=\omega_0\Omega t,\quad R=\frac{r}{l},\quad \Omega=\frac{\omega}{\omega_0}. \tag{11}$$

Here, $\omega_0=\pi c_0/l$ is the resonance frequency of a cylindrical resonator of length l. Previously we used the velocity potential φ (or Φ in dimensionless form) and the function V. They are related to the particle velocity u as

$$u=\frac{\partial\varphi}{\partial x}=l\omega_0\frac{\partial\Phi}{\partial X}, \tag{12}$$

and

$$V=R^2\frac{\partial\Phi}{\partial X}. \tag{13}$$

The spectral component amplitudes of the velocity wave are expressed through the dimensionless harmonic component amplitudes of the function V as

$$\tilde{u}_n=2l\omega_0\frac{V_n}{R^2}. \tag{14}$$

So, the energy losses in the boundary layer written in dimensionless variables are

$$W=\Gamma\sqrt{\tilde{\eta}_0}\int_0^1\sum_n |V_n|^2\sqrt{n\Omega}\frac{(1+R_1^2)^{3/2}}{R^3}dX. \tag{15}$$

Here, $\tilde{\eta}_0$ is a normalized attenuation coefficient

$$\tilde{\eta}_0=\frac{\eta_0}{\rho_0 c_0 l}, \tag{16}$$

Γ has dimension of power and is given by

$$\Gamma=2\pi\rho_0 l^4\omega_0^2\sqrt{2c_0 l\omega_0}, \tag{17}$$

and

$$R_1=\frac{dR}{dX}. \tag{18}$$

The calculation of the energy losses is reduced to solving the following integral:

$$L=\sqrt{\tilde{\eta}_0}\int_0^1\sum_n |V_n|^2\sqrt{n\Omega}\frac{(1+R_1^2)^{3/2}}{R^3}dX. \tag{19}$$

Instead of directly calculating the integral L, we solve the differential equation

$$\frac{dL}{dX}=\sqrt{\tilde{\eta}_0}\sum_n |V_n|^2\sqrt{n\Omega}\frac{(1+R_1^2)^{3/2}}{R^3}. \tag{20}$$

The solution of Eq. (20) yields the function L that is proportional to energy dissipation.

To calculate the dissipated energy in the boundary layer, we add Eq. (20) to the set of $(4N+1)$ differential equations we solved earlier. Now we have a set of $(4N+2)$ equations, and the energy losses in the boundary layer are one of the variables that we solve in this set of differential equations.

D. Quality factor of the resonator

The quality factor Q of the resonator is

$$Q=2\pi\frac{E}{WT_p}. \tag{21}$$

Here, E is the total energy stored in a resonator, and T_p is the period of the resonator oscillation.

Since the averaged kinetic energy is the same as the averaged internal energy, we calculate the total energy stored in the resonator as the doubled kinetic energy

$$E=\frac{1}{T_p}\int\int\rho u^2\,dv\,dt, \tag{22}$$

where dv is a volume element, and dt is a time element.

In dimensionless variables we have

$$E=\pi^3\rho_0 c_0^2 l^3\int\int\frac{\rho}{\rho_0}\frac{V^2}{R^2}dX\,dT=\pi^3\rho_0 c_0^2 l^3\tilde{S}, \tag{23}$$

where

$$\tilde{S}=\int\frac{\rho}{\rho_0}\frac{V^2}{R^2}dX. \tag{24}$$

As shown in a previous paper,[2] ρ/ρ_0 takes the form

$$\frac{\rho}{\rho_0}=\left[1-(\gamma-1)\pi^2\left(\Omega\frac{\partial\Phi}{\partial T}+\frac{V^2}{2R^4}+AX-\frac{G}{\pi^3R^2}\frac{\partial V}{\partial X}\right)\right]^{1/(\gamma-1)}, \tag{25}$$

where $A=a/l\omega_0^2$ is a dimensionless acceleration (a is the acceleration of the resonator), $G=\pi\delta\omega_0/c_0^2$ is a dimensionless volume attenuation coefficient [$\delta=(\zeta+4\eta/3)/\rho_0$].

The integral $\tilde{S}$ is obtained as the solution of the differential equation

$$\frac{d\tilde{S}}{dX}=\frac{\rho}{\rho_0}\frac{V^2}{R^2}. \tag{26}$$

One more equation is added to the set of $(4N+2)$ equations, and the $(4N+3)$th variable gives the desired integral $\tilde{S}$. Finally, the quality factor of the resonator is

$$Q=\frac{\Omega}{2}\sqrt{\frac{\pi}{2}}\frac{\tilde{S}}{L}. \tag{27}$$

III. THE EFFECT OF THE BOUNDARY LAYER ON THE ACOUSTIC FIELD IN THE RESONATOR

The boundary layer introduces not only energy losses but changes the characteristics of the acoustic field inside the resonator. These changes can be taken into account by adding a term in the continuity equation to describes the mass flow from the boundary layer to the volume part of the resonator. This is a concept first introduced by Chester.[3] Using the new continuity equation, we modify the nonlinear model equation in the time domain. Then, the solution scheme that has been previously[2] is applied to this analysis. The time-domain model equation is replaced by a set of equations in

the frequency domain that is complemented by two equations for the energy losses and the quality factor. These equations are solved numerically.

A. Modification of the previous derivation

Here, the previous derivation is modified in order to include the effect of the boundary layer.

The derivation starts with Eq. (14) from Ref. 2, and the terms are rearranged as follows:

$$\frac{\partial\varphi}{\partial t}+\frac{1}{2}\left(\frac{\partial\varphi}{\partial x}\right)^2-\frac{\delta}{r^2}\frac{\partial}{\partial x}\left(r^2\frac{\partial\varphi}{\partial x}\right)+a(t)x-\frac{c_0^2}{\gamma-1}$$

$$\equiv H=-\frac{\gamma p_0}{(\gamma-1)\rho_0^{\gamma}}\rho^{(\gamma-1)}=-\frac{c^2}{\gamma-1}. \tag{28}$$

Equations (16), (17), and (18) of Ref. 2 have been applied to obtain Eq. (28).

By differentiating the function H with respect to time t, we obtain the model equation for nonlinear standing waves

$$\frac{\partial H}{\partial t}=-\frac{\partial\varphi}{\partial x}\frac{\partial H}{\partial x}-\frac{(\gamma-1)}{r^2}\frac{\partial}{\partial x}\left(r^2\frac{\partial\varphi}{\partial x}\right)H. \tag{29}$$

Equation (29) is rewritten here in a form that is not only more compact but also more convenient for our further investigation. By substituting the expression for H into Eq. (29), one will recover Eq. (28) of Ref. 2 for the velocity potential φ.

B. Flow from the edge of the boundary layer into the volume of the resonator

Next, the mass flow from the edge of the boundary layer into the volume is estimated and included into the continuity equation for the volume.

The mass flow through a unit element of the boundary layer edge into the volume is given by

$$f_n=\rho u_{yn}2\pi r\frac{1}{\cos\theta}. \tag{30}$$

Here, f_n is a spectral component of the flow, and u_{yn} is a spectral component of the velocity component normal to the resonator wall.

To calculate the normal velocity component u_y, we combine the continuity equation for the boundary layer with the continuity equation for the volume. The continuity equation for the boundary layer is

$$\frac{\partial\rho}{\partial t}+\frac{\partial}{\partial y}(\rho u_y)+\frac{1}{r}\frac{\partial}{\partial z}(\rho r u_z)=0, \tag{31}$$

and the continuity equation in the volume is

$$\frac{\partial\rho}{\partial t}+\frac{1}{r}\frac{\partial}{\partial z}(\rho r u_z^{(0)})=0, \tag{32}$$

where $u_z^{(0)}$ is the velocity component parallel to the resonator wall. Since the pressure is the same in the boundary layer and in the volume, the density is also the same, and we obtain from Eqs. (31) and (32)

$$\frac{\partial}{\partial y}(\rho u_y)+\frac{1}{r}\frac{\partial}{\partial z}[\rho r(u_z-u_z^{(0)})]=0. \tag{33}$$

Integration yields

$$u_y=-\frac{1}{\rho r}\int_0^{\infty}\frac{\partial}{\partial z}[\rho r(u_z-u_z^{(0)})]dy. \tag{34}$$

Since u_z is small and the difference between ρ and ρ_0 is also small, ρ can be replaced by ρ_0. This approach is consistent with the use of Eq. (1), the solution of which is used to integrate Eq. (34) over y

$$u_{yn}=-\frac{1}{\alpha r}\frac{\partial}{\partial z}(r u_n^{(0)}). \tag{35}$$

Next, we neglect the difference between the velocity component along the wall and along the axis of the resonator and we replace z by x. We then obtain

$$u_{yn}=-\frac{1}{\alpha r}\frac{\partial}{\partial x}(r\tilde{u}_n). \tag{36}$$

Substitution of u_{yn} into Eq. (30) yields

$$f_n=-F_n\rho, \tag{37}$$

where

$$F_n=-\pi\sqrt{\frac{2\eta}{n\omega\rho_0}}(1-i)\frac{\partial}{\partial x}(r\tilde{u}_n). \tag{38}$$

This flow from the boundary layer to the volume is taken into account in the continuity equation for the volume. The continuity equation is then used to derive the model equation that describes the acoustic field in the resonator with an acoustic boundary layer.

C. Derivation of the new model equations

The mass conservation equation becomes

$$\frac{\partial\rho}{\partial t}+\frac{1}{r^2}\frac{\partial}{\partial x}(r^2\rho u)+\frac{F\rho}{\pi r^2}=0, \tag{39}$$

where the function F has spectral components F_n. The inclusion of the boundary layer flow into the mass conservation equation changes the model equation (29), which is now given by

$$\frac{\partial H}{\partial t}=-\frac{\partial\varphi}{\partial x}\frac{\partial H}{\partial x}-\frac{(\gamma-1)}{r^2}\frac{\partial}{\partial x}\left(r^2\frac{\partial\varphi}{\partial x}\right)H$$

$$-(\gamma-1)\frac{F}{\pi r^2}H. \tag{40}$$

By expressing H as a function of the velocity potential φ [Eq. (28)], we can rewrite Eq. (40) and obtain

$$\frac{c_0^2}{r^2}\frac{\partial}{\partial x}\left(r^2\frac{\partial\varphi}{\partial x}\right)-\frac{\partial^2\varphi}{\partial t^2}+\frac{\delta}{r^2}\frac{\partial^2}{\partial t\partial x}\left(r^2\frac{\partial\varphi}{\partial x}\right)$$
$$=\frac{da}{dt}x+a(t)\frac{\partial\varphi}{\partial x}+\frac{\gamma-1}{r^2}a(t)x\frac{\partial}{\partial x}\left(r^2\frac{\partial\varphi}{\partial x}\right)$$
$$+2\frac{\partial^2\varphi}{\partial x\partial t}\frac{\partial\varphi}{\partial x}+\frac{\gamma-1}{r^2}\frac{\partial\varphi}{\partial t}\frac{\partial}{\partial x}\left(r^2\frac{\partial\varphi}{\partial x}\right)+\frac{1}{3}\frac{\partial}{\partial x}\left(\frac{\partial\varphi}{\partial x}\right)^3$$
$$+\frac{\gamma-1}{2r^2}\left(\frac{\partial\varphi}{\partial x}\right)^2\frac{\partial}{\partial x}\left(r^2\frac{\partial\varphi}{\partial x}\right)-\frac{c_0^2}{\pi r^2}F. \tag{41}$$

Equation (41) is the modified model equation that describes the acoustic field inside the resonator with a boundary layer. Since we used a linear approximation to calculate the power that is dissipated in the boundary layer, we retained only the linear term to describe the flow F from the boundary layer.

The following dimensionless variables are introduced:

$$\Phi=\frac{\varphi}{\omega_0 l^2},\quad V=R^2\frac{\partial\Phi}{\partial X}. \tag{42}$$

The new function V allows one to reduce the order of Eq. (41). We are interested in the periodic solutions

$$V=\sum_{k=-N}^{N}V_k e^{ikT}, \tag{43}$$

$$\Phi=\sum_{k=-N}^{N}\Phi_k e^{ikT}. \tag{44}$$

Equation (41) in new dimensionless variables reduces to ordinary differential equations for the complex amplitudes. For the kth harmonic component the equations are

$$\frac{1}{\pi^2}\frac{dV_k}{dX}+k^2\Omega^2R^2\Phi_k+\frac{ik\Omega G}{\pi^3}\frac{dV_k}{dX}-\frac{(1-i)b}{\sqrt{k\Omega}R}\frac{dV_k}{dX}$$
$$=ik\Omega R^2XA_k+\frac{ik\Omega}{R^2}[V^2]_k+\sum_{l=-N+k}^{N}\left\{A_{k-l}V_l\right.$$
$$\left.+(\gamma-1)XA_{k-l}\frac{dV_l}{dX}+i(k-l)(\gamma-1)\Omega\Phi_{k-l}\frac{dV_l}{dX}\right\}$$
$$+\sum_{l=-N}^{N}\left\{\frac{\gamma+1}{2R^4}[V^2]_{k-l}\frac{dV_l}{dX}-\frac{2}{R^5}\frac{dR}{dX}[V^2]_{k-l}V_l\right\}$$
$$-\frac{(1-i)b}{\sqrt{k\Omega}R^2}\frac{dR}{dX}V_k, \tag{45}$$

$$\frac{d\Phi_k}{dX}=\frac{V_k}{R^2}. \tag{46}$$

Notice that $V_{-k}=V_k^*$ where V_k^* are complex conjugates V_k, and $V_0=V_{\rm dc}$ is the dc component of the function V. The same notation applies for the velocity potential. N is the number of harmonics included in the calculation. Both functions $V_{\rm dc}$ and $\Phi_{\rm dc}$ are real. The acceleration is assumed to be periodic

$$A=\sum_{k=-N}^{N}A_k e^{ikT}. \tag{47}$$

The factor b is defined as

$$b=\frac{1}{\pi^2}\sqrt{\frac{2\tilde{\eta}_0}{\pi}}. \tag{48}$$

To describe the acoustic field in the resonator with a boundary layer, we have to integrate $(4N+1)$ ordinary differential equations, i.e., the real and imaginary parts of N complex amplitudes for each of the two functions V and Φ as well as the real function $V_{\rm dc}$. The function $\Phi_{\rm dc}$ is not involved in Eqs. (45) and (46). To estimate the energy losses in the boundary layer and the quality factor of the resonator, we add Eqs. (20) and (26) to the set of $(4N+1)$ equations and solve the set of $(4N+3)$ equations.

Equations (45) and (46), with the additional term that takes into account the influence of the boundary layer on the dynamics of the processes in the resonator, have been used to calculate numerically the losses in the resonator. The numerical results have been compared with experiments performed at MacroSonix Corp. and showed good agreement for weak excitation, when energy losses increase as the square of the acoustical amplitude. It is important to note that energy losses have been calculated using not only Eqs. (45) and (46) together with Eq. (15) but also by estimating work performed by an external force to excite oscillations in the resonator. This second approach, which we now describe, was used to check the calculations based on Eqs. (45) and (46).

We calculate the work done by an external force $F_{\rm ex}$. The power required to excite oscillations by moving an entire resonator with velocity v_0 is equal to $v_0F_{\rm ex}$. The force $F_{\rm ex}$ can be calculated using the equation for the gas pressure in the resonator as well as the derivative of the full momentum M_g of gas in the resonator. Both ways yield the same result. The average power of excitation per period is

$$\langle v_0F_{\rm ex}\rangle=\langle v_0\dot{M}_g\rangle=-\langle\dot{v}_0M_g\rangle$$
$$=-\left\langle a\int\rho u\,dV\right\rangle=-\left\langle\int\rho au\,dV\right\rangle, \tag{49}$$

where dV is a volume element. The last expression is the average power associated with an inertial force $-\rho a(t)$. It is easier to evaluate this quantity than to integrate pressure over the resonator wall. Energy losses calculated numerically with Eqs. (45), (46), and (15) were compared with Eq. (49) and found to agree within a few percent. This fact indicates that Eqs. (45) and (46) with Eq. (15) are sufficient to describe dynamics of a resonator with boundary layer and the associated energy losses.

IV. TURBULENCE DISSIPATION

With increasing amplitude of excitation, energy losses increase faster than the square of the acoustical amplitude as discussed below in connection with the measurements shown in Fig. 15. Harmonic generation can account for only 10%–20% of ''excess'' losses that exceed linear losses. This is consistent with theoretical predictions of additional attenuation due to harmonics, since energy losses in the boundary

layer depend on frequency as $\sqrt{\omega}$, and the amplitudes of the harmonics are small because of the special shape of the resonator. The second harmonic does not exceed 20% of the fundamental wave amplitude and contribution of the second harmonic to the excess losses is less than 10%. The higher harmonics decrease very rapidly and do not contribute significantly to energy losses. The conclusion is that for high-amplitude wave excitation in a resonator with a shape that suppresses harmonic generation, significant excess losses are mostly not associated with harmonic generation but with increasing effective viscosity and thermal conductivity due to turbulence.

Noticeable deviation of our numerical results from experiments starts when the Reynolds number Re in the boundary layer is of the order of hundreds. This coincides with experimental investigations performed by Merkli and Thomann,[7] which show that the transition to turbulence in the boundary layer of oscillating pipe flow occurs at $\mathrm{Re}=400$. As the Reynolds number in the volume is greater than in the boundary layer, it is natural to anticipate that turbulence starts in the volume earlier than in the boundary layer. But, energy losses in the boundary layer are considerably greater than in the volume. For weak excitation the ratio between boundary layer losses and volume losses is of the order of one wavelength or resonator length divided by the boundary layer thickness. We assume that the contribution of turbulence in the volume to the losses in the resonator can be ignored in this preliminary investigation. The similarity of Reynolds numbers observed in experiment when turbulence occurs in the boundary layer of oscillating pipe flow[7] and the appearance of excess losses in resonators by MacroSonix Corp. indicates the suitability of our approach.

In this section we estimate energy losses due to turbulence in the boundary layer using a simple eddy viscosity model. The eddy viscosity model that we apply here does not describe the dynamics of the turbulence. Experiment shows this model is good enough to estimate integral losses associated with the turbulence. The model equation for the eddy viscosity is taken as a simple amplitude-dependent function that has parameters to adjust calculated energy losses to measurements. The viscosity coefficient is presented in the form

$$\eta=\eta_0+\eta_e, \tag{50}$$

where η_0 is the dynamic viscosity coefficient and η_e is the eddy viscosity coefficient.

The transition to turbulence is governed by the local Reynolds number based on the boundary layer thickness δ

$$\mathrm{Re}=\frac{\sqrt{2}Mc_0\delta}{\nu_0}, \tag{51}$$

where M is a Mach number in the boundary layer. The boundary layer thickness δ is equal to

$$\delta=\sqrt{\frac{2\nu}{\omega}}. \tag{52}$$

The kinematic viscosity coefficient $\nu=\eta/\rho_0$ in Eq. (52) includes the total viscosity, i.e., the sum of dynamic viscosity and eddy viscosity.

According to Merkli and Thomann, the transition to turbulence occurs at

$$\mathrm{Re}=\mathrm{Re}_0=400. \tag{53}$$

Merkli and Thomann also noted that initially turbulence does not exist during the entire period of the wave, i.e., it consists of bursts of turbulence within each cycle.

We assume that the eddy viscosity coefficient η_e is

$$\eta_e=\zeta\eta_0\left(\sqrt{1+\left(\frac{\mathrm{Re}}{\mathrm{Re}_0}\right)^2}-1\right). \tag{54}$$

Here, ζ is a parameter that characterizes the ratio between the dynamic viscosity and the eddy viscosity. The value of ζ is determined by matching calculated and measured dissipation. Equation (54) has a very simple interpretation: it states that the eddy viscosity coefficient is dependent on the amplitude of the Reynolds number. At low Reynolds numbers, $\mathrm{Re}\ll\mathrm{Re}_0$, the eddy coefficient η_e becomes very small. At high Reynolds numbers, $\mathrm{Re}\gg\mathrm{Re}_0$, η_e is proportional to the Reynolds number, i.e., $\eta_e\approx\zeta\eta_0(\mathrm{Re}/\mathrm{Re}_0-1)$.

Equation (50) can be written as

$$\tilde{\nu}=1+\zeta(\sqrt{1+\beta M^2\tilde{\nu}}-1), \tag{55}$$

where $\tilde{\nu}=\nu/\nu_0$, $\beta=4\,\mathrm{Re}_1/\pi\mathrm{Re}_0^2\Omega$, and $\mathrm{Re}_1=\rho_0c_0l/\eta_0\equiv1/\tilde{\eta}_0$. The value $\tilde{\nu}$ is given by

$$\tilde{\nu}=\kappa+\sqrt{\kappa^2+2\zeta-1}, \tag{56}$$

where κ is defined as

$$\kappa=1-\zeta+\tfrac{1}{2}\zeta^2\beta M^2. \tag{57}$$

In order to take into account the eddy viscosity factor, we need to multiply the dimensionless dynamic viscosity $\tilde{\eta}$ in Eqs. (19), (27), and (48) by the factor $\tilde{\nu}$ given by Eq. (56).

V. RESULTS OF NUMERICAL CALCULATIONS

The fifth-order Runge–Kutta scheme with adaptive step-size control[8] is used to numerically integrate the set of $(4N+3)$ equations. The Runge–Kutta routine is applied to the two-point boundary value problem. The solution of Eqs. (45) and (46) allows one to calculate the pressure and velocity. Energy losses and quality factor Q are obtained by integrating Eqs. (20) and (26).

During each calculation a frequency sweep is performed, where the frequency is varied from slightly below the fundamental resonance frequency of the resonator to slightly above the resonance frequency. The following acoustical characteristics are calculated: amplitude and phase of each spectral component of the pressure wave, amplitude and phase distributions along the resonator for the harmonic components of the pressure and particle velocity waves, waveforms for pressure and velocity at ten points along the resonator axis, and the dependence of energy losses and quality factor Q on frequency. Calculations are performed for different acceleration amplitudes. For each calculation ten harmonics are taken into account. In all equations the coefficient of viscosity $\tilde{\eta}$ is replaced by $\tilde{\eta}\tilde{\nu}$ to take into account eddy viscosity. Numerical results are presented for resonators of three geometries: a cylinder, a horn cone, and a

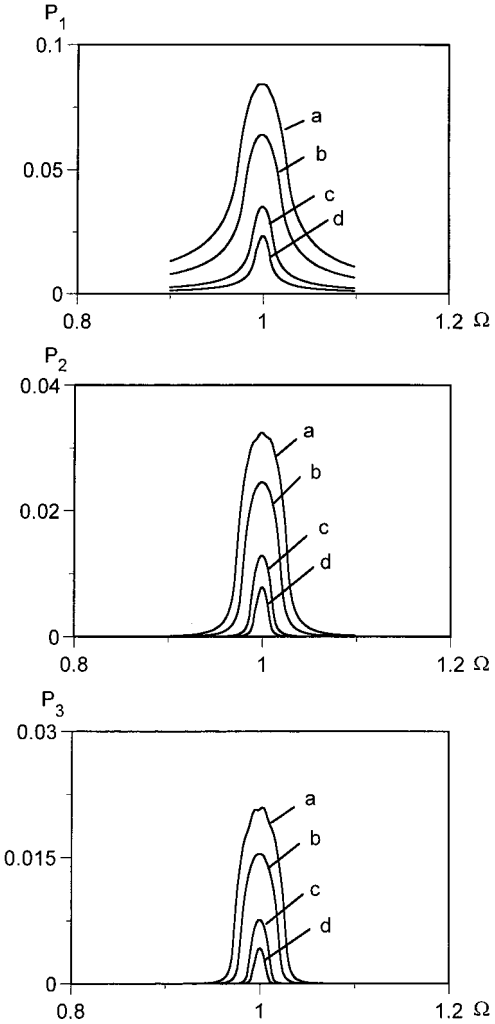


FIG. 1. Dependence of the first three harmonic amplitudes of the pressure wave on the frequency in the cylindrical resonator for different acceleration amplitudes: $A=5\times10^{-4}$ (curves *a*), $A=3\times10^{-4}$ (curves *b*), $A=1\times10^{-4}$ (curves *c*), and $A=0.5\times10^{-4}$ (curves *d*) at $X=0$. Here, pressure amplitudes P_1, P_2, and P_3 are normalized by ambient pressure.

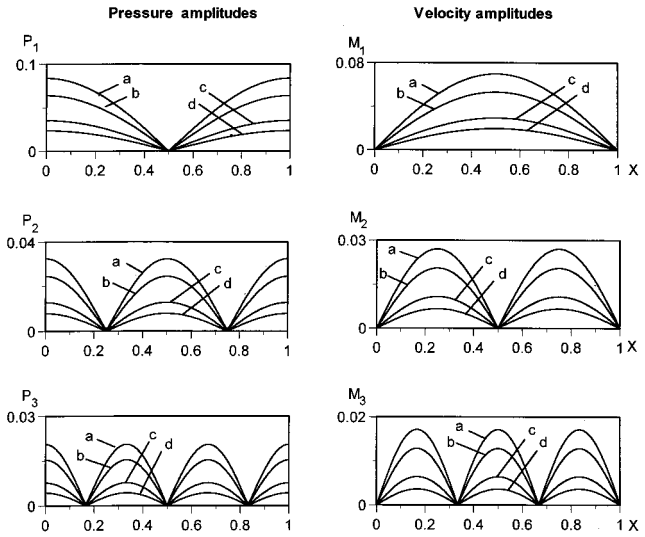


FIG. 2. The first three harmonic amplitude distributions along the cylindrical resonator for pressure wave (left column) and particle velocity wave (right column) for different acceleration amplitudes: $A=5\times10^{-4}$ (curves *a*), $A=3\times10^{-4}$ (curves *b*), $A=1\times10^{-4}$ (curves *c*), and $A=0.5\times10^{-4}$ (curves *d*) at frequency $\Omega=1$. Here, M_1, M_2, M_3 are acoustic Mach numbers that correspond to harmonic amplitudes of the velocity wave normalized by the sound speed, and P_1, P_2, and P_3 are corresponding sound pressure normalized by ambient pressure.

bulb-type resonator. The cylinder is an example of a consonant resonator, i.e., a resonator with an equidistant spectrum of its modes. The horn cone and the bulb-type resonator are dissonant, they have a nonequidistant mode spectrum. The nonlinear distortion is different in all three resonators.

A. Cylinder

The dependence of the first three harmonic amplitudes of the pressure wave on frequency at $X=0$ are shown in Fig. 1 for different amplitudes of resonator acceleration: curve *a* corresponds to acceleration $A=5\times10^{-4}$, curve *b* to $A=3\times10^{-4}$, curve *c* to $A=1\times10^{-4}$, and curve *d* to $A=0.5\times10^{-4}$. From this graph we observe that for a cylindrical resonator, the resonance frequency does not depend on the amplitude of resonator oscillation. With increasing amplitude the resonance peak broadens.

The first three harmonic amplitude distributions of pressure and velocity along the resonator axis are plotted in Fig. 2 for different acceleration amplitudes. Curves *a*, *b*, *c*, and *d* correspond to $A=5\times10^{-4}$, $A=3\times10^{-4}$, $A=1\times10^{-4}$, and $A=0.5\times10^{-4}$, respectively. The left column presents the results for the pressure wave, and the right for the velocity wave. As expected, all curves are symmetric with respect to $X=0.5$. The pressure and velocity nodes and antinodes are at the same location for all acceleration amplitudes.

The maximum values of pressure and velocity amplitudes increase with increasing acceleration amplitude.

Pressure (left column) and particle velocity (right column) waveforms are displayed in Fig. 3 at different points along the resonator axis for a high level of acceleration, $A=5\times10^{-4}$. Figure 3 reveals the strong nonlinear distortion of the wave in the cylindrical resonator. Shock fronts occur in the pressure wave as well as in the particle velocity wave. The shape of the velocity wave is different than that of the pressure wave. At $X=0$ the pressure waveform has the classical saw-tooth shape. In the middle of the resonator, $X=0.5$, a saw-tooth shape is observed but at twice the frequency of the fundamental. This can be explained by the fact that the amplitudes of the fundamental component and all odd harmonics are zero at this location. Maximum pressure amplitude is about 0.10. In the middle of the resonator the shape of the velocity wave is that of a square wave, while near the end walls the shape is closer to that of an impulse function. Maximum velocity amplitude is about 0.1.

Energy losses (upper plot) and the quality factor (lower plot) for different acceleration amplitudes are shown in Fig. 4. The energy loss curves show the resonance behavior and the energy losses reach their maximum value at $\Omega=1$. Energy losses increase with increasing levels of acceleration. The quality factor shows a similar but opposite behavior. It reaches a minimum value close to the resonance frequency. The quality factor decreases with increasing level of acceleration.

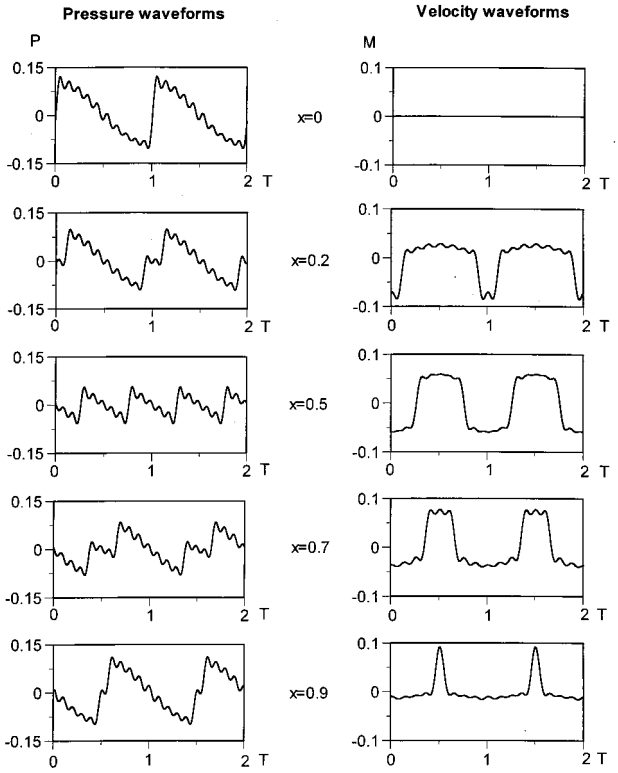


FIG. 3. The wave shapes of the pressure wave (left column) and the velocity wave (right column) at the different points along the cylindrical resonator axis. Here, $A=5\times10^{-4}$, $\Omega=1$. Here, M is Mach number, and P is the dimensionless sound pressure.

B. Horn cone

The characteristics of standing waves in the horn-cone resonator are different from those in the cylinder. The resonance curves for the fundamental pressure wave amplitude are presented in Fig. 5 for different amplitudes of acceleration. Curves *a*, *b*, *c*, and *d* correspond to accelerations $A=5\times10^{-4}$, $A=3\times10^{-4}$, $A=1\times10^{-4}$, and $A=0.5\times10^{-4}$, respectively. With increasing amplitude of acceleration the resonance frequency shifts to higher frequencies. The horn-cone resonator is an example of hardening resonance behavior. For high acceleration amplitude the resonance curves (curves *a* and *b*) are characterized by hysteresis. To obtain the resonance curve *a* or *b* the numerical code is run twice. First, we calculated from low frequency up to the frequency where the function $P_1(\Omega)$ becomes multivalued (the upper branch of curve *a*). Second, we swept down in frequency from high frequencies up to the frequency that corresponds to multivalued behavior.

The first three harmonic amplitude distributions of pressure (left column) and velocity (right column) along the resonator are shown in Fig. 6. The harmonic amplitudes of the pressure wave are no longer symmetric along the resonator axis but have the same general behavior, i.e., half wavelength for the fundamental, full wavelength for the second harmonic, and one and a half wavelengths for the third harmonic. The fundamental pressure amplitude reaches a value of 1.25 at $X=0$ for the highest level of acceleration. In general the same observation applies to the velocity distributions. However, for high acceleration levels, the second harmonic distribution is different, i.e., an extra node and antinode occur. It is interesting to note that for the highest acceleration level the third harmonic amplitude of the pressure wave is of the same order as the second one, and the third harmonic amplitude of the velocity wave is greater than the second one. The nodes in the second and third harmonic distributions shift to higher values of X. The occurrence of an extra node and antinode in the velocity distribution of the second harmonic can be explained by the resonance harden-

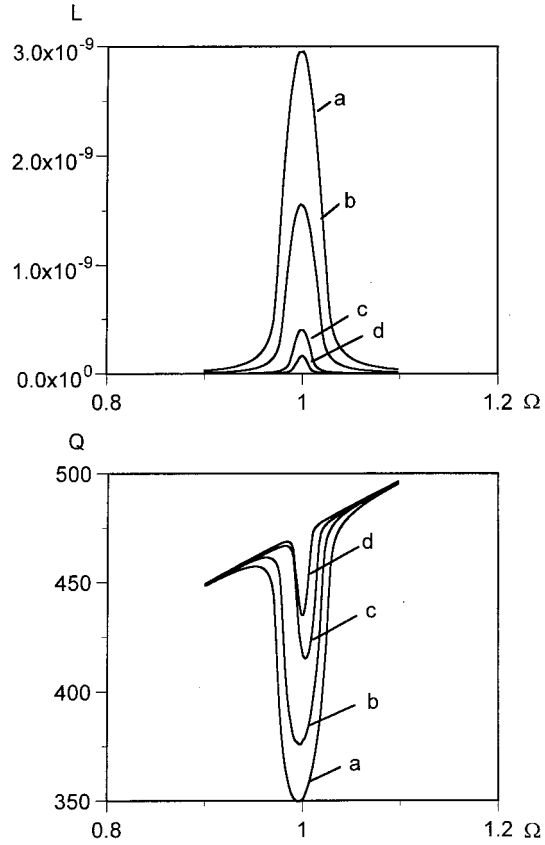


FIG. 4. Dependence of energy losses (upper plot) and the quality factor (lower plot) on the frequency in the cylindrical resonator for different acceleration amplitudes: $A=5\times10^{-4}$ (curves *a*), $A=3\times10^{-4}$ (curves *b*), $A=1\times10^{-4}$ (curves *c*), and $A=0.5\times10^{-4}$ (curves *d*).

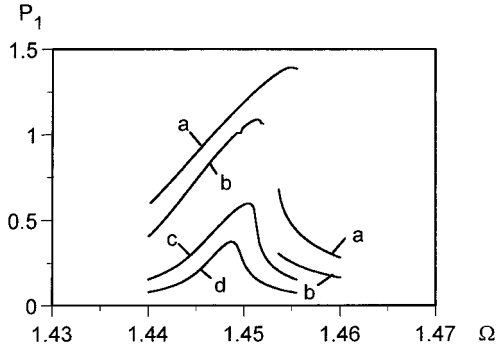


FIG. 5. Frequency response of the fundamental pressure wave in the horn cone for different acceleration amplitudes: $A=5\times10^{-4}$ (curve *a*), $A=3\times10^{-4}$ (curve *b*), $A=1\times10^{-4}$ (curve *c*), and $A=0.5\times10^{-4}$ (curve *d*) at $X=0$. P_1 is dimensionless amplitude of the fundamental wave.

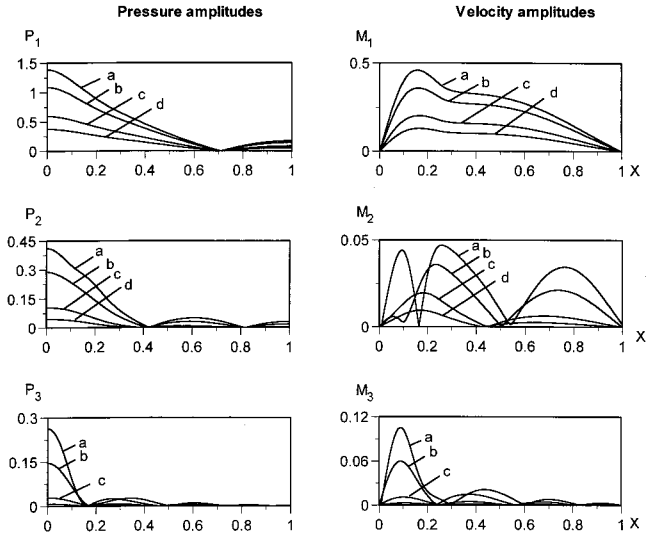


FIG. 6. The first three harmonic amplitude distributions along the horn cone for the pressure wave (left column) and the particle velocity wave (right column) for different acceleration amplitudes: $A=5\times10^{-4}$ (curves *a*), $A=3\times10^{-4}$ (curves *b*), $A=1\times10^{-4}$ (curves *c*), and $A=0.5\times10^{-4}$ (curves *d*) at resonance frequencies for each amplitude of acceleration. P_n and M_n are defined as in Fig. 2.

ing behavior of this resonator. With increasing amplitude the resonance frequency moves to a higher frequency, and the second harmonic frequency is getting closer to the frequency of the third mode. The influence of the third mode on the second harmonic generation leads to additional maximum and minimum in the second harmonic amplitude distribution.

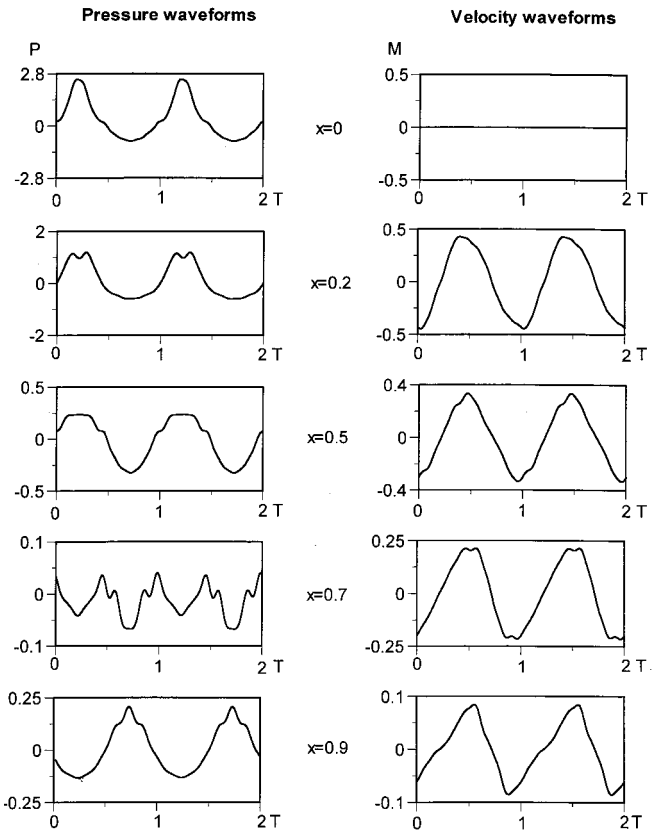


FIG. 7. The wave shapes of the pressure wave (left column) and the velocity wave (right column) at different points along the horn-cone resonator. Here, $A=5\times10^{-4}$, $\Omega=1.454$.

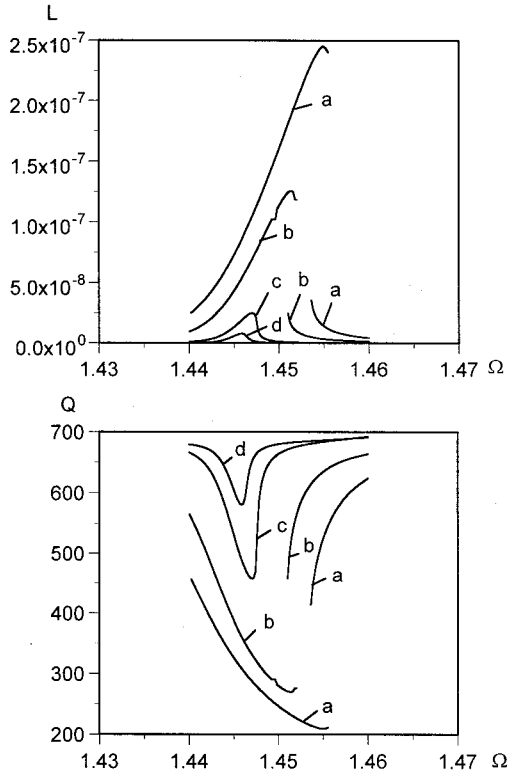


FIG. 8. Dependence of energy losses (upper plot) and the quality factor (lower plot) on the frequency in the horn cone for different acceleration amplitudes: $A=5\times10^{-4}$ (curves *a*), $A=3\times10^{-4}$ (curves *b*), $A=1\times10^{-4}$ (curves *c*), and $A=0.5\times10^{-4}$ (curves *d*).

Waveforms for pressure and velocity are presented in Fig. 7. No shocks are observed in the pressure or velocity wave. It is also clear from Fig. 7 that for identical acceleration the peak pressure is much higher than that in the cylinder. This phenomenon can be explained by the fact that the mode spectrum of the horn cone is nonequidistant. Harmonic generation is suppressed and harmonic phases are changed so that shock formation is prevented. At $X=0$ the pressure waveform for the horn cone is a *U*-waveform, i.e., sharper peaks and broader valleys. The velocity waveform has the general appearance of a triangular waveform. Maximum pressure amplitude is about 2.5 and maximum velocity amplitude is about 0.5.

The variation of the energy losses and quality factor as a functions of frequency is displayed in Fig. 8. The energy losses and behavior of Q is similar to that of the pressure. Hysteresis and resonance hardening are observed when the amplitude of oscillation is high.

C. Bulb-type resonator

In Fig. 9 we show the resonance curves for the fundamental pressure amplitude for different amplitudes of acceleration. Hysteresis is observed at even a relatively low amplitude of acceleration (curves *c*). The left and right branches of curve *c* that correspond to $A=1\times10^{-4}$ overlap. Curves *a*

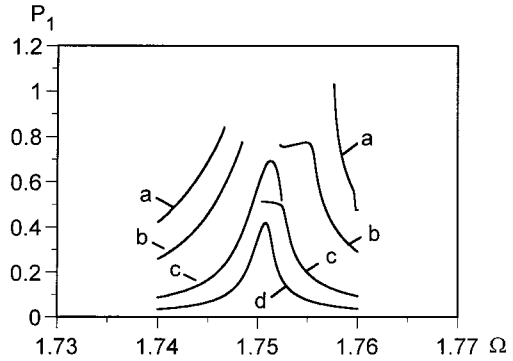


FIG. 9. Frequency response of the fundamental pressure wave in the bulb-type resonator for different acceleration amplitudes: $A=5\times10^{-4}$ (curves *a*), $A=3\times10^{-4}$ (curves *b*), $A=1\times10^{-4}$ (curves *c*), and $A=0.4\times10^{-4}$ (curve *d*) at $X=0$.

and *b* display a complicated resonance behavior. This type of behavior is typical for a system when a higher harmonic is coincident with a resonator mode. The resonance frequency of this resonator is higher than the resonance frequency of the cylinder and the horn cone.

The first three harmonic amplitude distributions along the resonator are presented in Fig. 10 for the pressure wave (left column) and the velocity wave (right column). Three amplitudes of acceleration are presented: $A=5\times10^{-4}$ (curves *a*), $A=3\times10^{-4}$ (curves *b*), and $A=1\times10^{-4}$ (curves *c*). The third harmonic is higher than the second one in the pressure and velocity wave. The third harmonic here is very close to the fifth mode, as can be seen in Fig. 11. Therefore, five sections of a half wavelength are seen for the third harmonic distribution. It also explains the relatively high levels of third harmonic generation. An explanation for the lower second harmonic is that it is generated in two ways which compete with each other. One mechanism is the contribution through the second mode, while the second mechanism is the

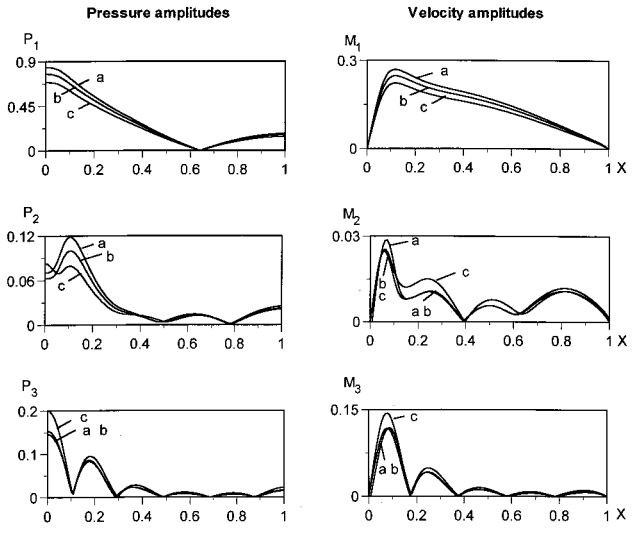


FIG. 10. The first three harmonic amplitude distributions along the bulb-type resonator for the pressure wave (left column) and the velocity wave (right column) for different acceleration amplitudes: $A=5\times10^{-4}$ (curves *a*), $A=3\times10^{-4}$ (curves *b*), $A=1\times10^{-4}$ (curves *c*) at the resonance frequencies.

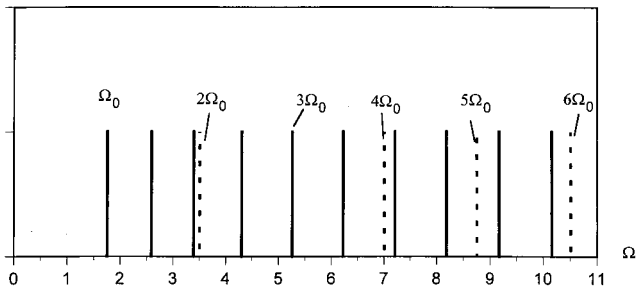


FIG. 11. Spectrum of the bulb-type resonator (solid lines) and harmonic frequencies (dashed lines).

contribution through the third mode. Since the second harmonic is very close to the third mode, a significant part is contributed through the third mode. These two mechanisms determine the second harmonic, but each contributes with a different phase, so that they tend to cancel each other.

In Fig. 11 the modal spectrum and the harmonics for the bulb resonator are shown. The solid lines denote the first ten modal frequencies for this resonator. The dashed lines denote the harmonics of the fundamental frequency. As mentioned before, the third harmonic is nearly identical to the fifth modal frequency, and the second harmonic is situated between the third and fourth mode.

The pressure and velocity waveforms are shown in Fig. 12 for different points along the resonator axis. There are no

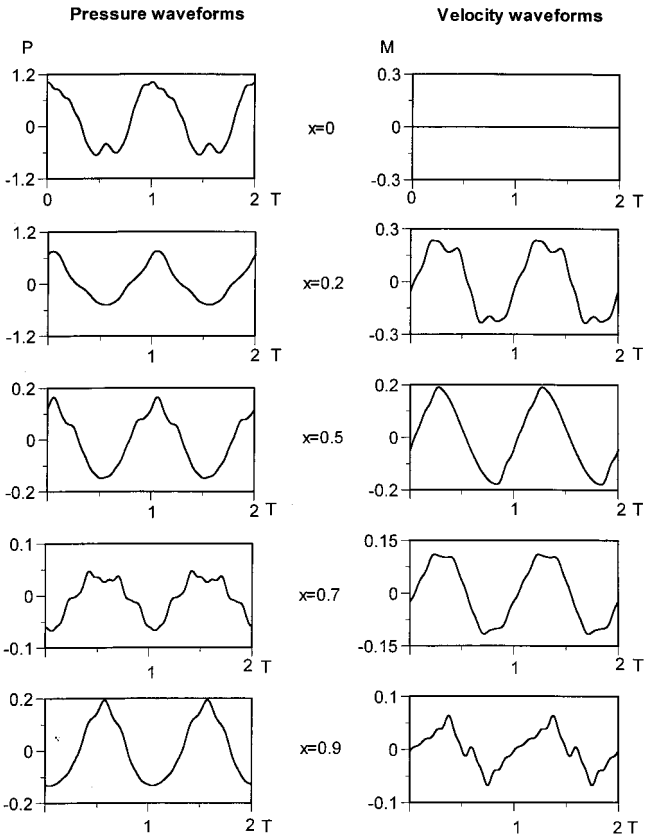


FIG. 12. The shapes of the pressure wave (left column) and the velocity wave (right column) at different points along the bulb-type resonator axis. Here, $A=5\times10^{-4}$, $\Omega=1.7466$.

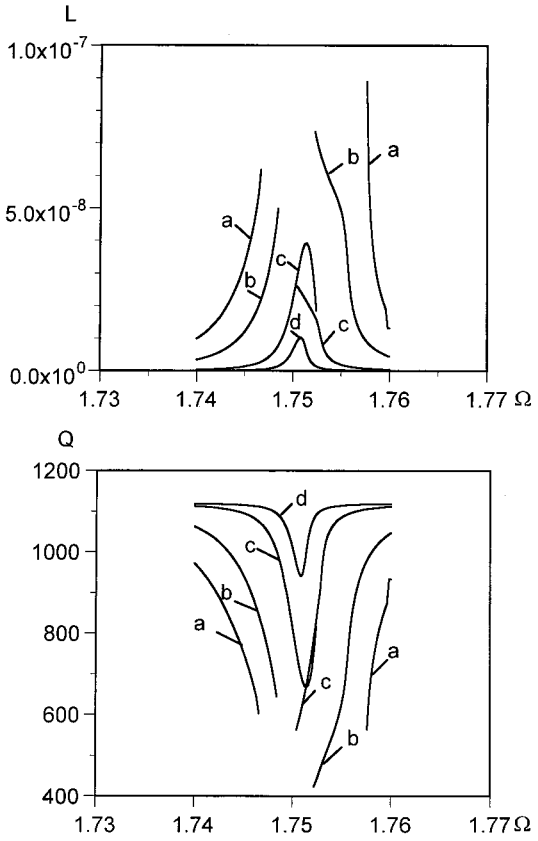


FIG. 13. Dependence of energy losses (upper plot) and the quality factor (bottom plot) on the frequency in the bulb-type resonator for different acceleration amplitudes: $A=5\times10^{-4}$ (curves *a*), $A=3\times10^{-4}$ (curves *b*), $A=1\times10^{-4}$ (curves *c*), $A=0.4\times10^{-4}$ (curves *d*).

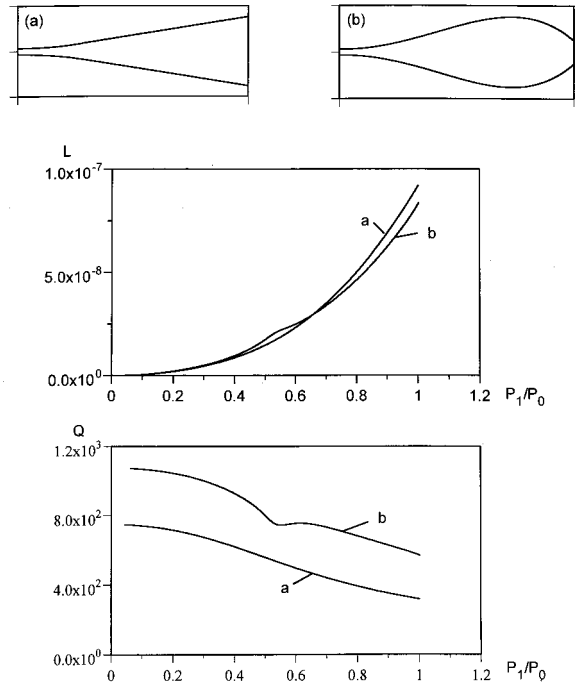


FIG. 14. Dependence of energy losses and the quality factor on the amplitude of the fundamental wave for two types of resonators: for the horn-cone (curves *a*) and for the bulb-type resonator (curves *b*). Shapes of both resonators are shown on the top: (a) is the horn-cone, and (b) is the bulb-type resonator.

shock fronts in the pressure (left column) or velocity (right column) waveforms.

The energy losses L and quality factor of the resonator are plotted in Fig. 13. Again, we observe hysteresis for curves *a*, *b*, and *c*. From comparison of the quality factors of the three resonators, it is clear that the bulb resonator has the highest Q for a given level of acceleration.

The dependence of the energy losses L and the quality factor on the amplitude of the fundamental pressure amplitude is shown for two types of resonators, a horn cone (curve *a*) and a bulb (curve *b*), in Fig. 14. The shape of the resonators is shown as well. We note that the energy losses are about the same, while there is a significant difference in Q. This example is an illustration of the fact that resonator geometry can significantly alter the characteristics of the standing wave.

D. Comparison with experiment

The theoretical predictions for the energy losses in the boundary layer show good agreement with experimental data. Two resonators are chosen to demonstrate this agreement, the horn-cone and the bulb resonator. For the parameter Re_0 we use the same value of 400 as proposed by Merkli and Thomann.[7] In order to determine the value of the parameter ζ we matched the predicted energy loss curve in the horn-cone resonator with the measured values of the energy losses. As seen in Fig. 15(a), a value of $\zeta=1.75$ provided a good agreement. This value of ζ was then kept constant for all other calculations. As shown in Fig. 15(b), the comparison between the calculated and measured results for the bulb-type resonator is very good. The geometry of the resonators is shown at the top of Fig. 15. For this calculation we

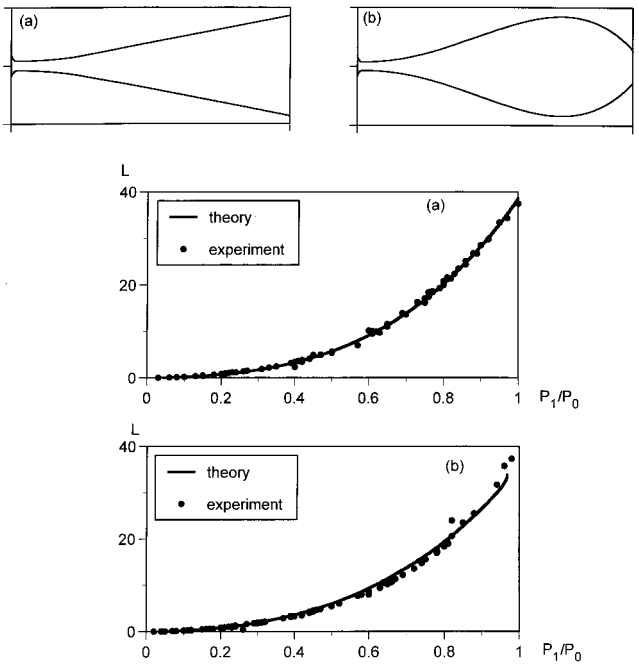


FIG. 15. Comparison of experimental data with the theoretical predictions for two resonators: the horn cone with the flare (a) and the bulb-type resonator with the flare (b): energy losses dependence on the amplitude of the fundamental pressure wave for both resonators, equilibrium pressure is $P_0=40$ psi (absolute)$=2.72$ bar, $T_0=130\,^\circ\mathrm{F}=327$ K.

swept from low to high acceleration level and at each level the calculations were done at the resonance frequency of each resonator.

VI. CONCLUSION

Recently, we have developed a one-dimensional mathematical model and a numerical scheme to analyze nonlinear standing waves in axisymmetric resonators.[2] Here, this model is modified to include the effect of the acoustic boundary layer along the resonator wall. The boundary layer introduces energy losses and influences the acoustic field in the resonator. This influence is included into the model equation by introducing an additional term into the continuity equation to describe the mass flow from the boundary layer into the resonator volume. Calculations of the energy losses and the quality factor of the resonator are included into the numerical scheme. The energy dissipation associated with turbulence is taken into account by a simple eddy viscosity model.

Results are shown for three resonator geometries, a cylinder, a horn-cone, and a bulb resonator, and for several levels of resonator acceleration. Frequency response curves for the fundamental pressure component and distributions of the harmonic amplitudes of the pressure and velocity wave are shown. The pressure and velocity waveforms calculated at several positions along the resonator axis are presented. The dependence of the energy losses and the quality factor on frequency is shown. The pressure waveform in the cylindrical resonator exhibits the typical saw-tooth waveform with very limited pressures, while the waveforms for the horn-cone and bulb resonator are not shocked. Much higher pressures are attained for the same levels of acceleration.

Finally, the energy losses predicted with this numerical code are in good agreement with the experimental data. A comparison of numerical and experimental results of energy losses is shown for two resonators, a horn-cone resonator and bulb-type resonator.

[1] C. Lawrenson, B. Lipkens, T. S. Lucas, D. K. Perkins, and T. W. Van Doren, ''Measurements of macrosonic standing waves in oscillating closed cavities,'' J. Acoust. Soc. Am. **104**, 623–636 (1998).

[2] Yu. A. Ilinskii, B. Lipkens, T. S. Lucas, T. W. Van Doren, and E. A. Zabolotskaya, ''Nonlinear standing waves in an acoustical resonator,'' J. Acoust. Soc. Am. **104**, 2664–2674 (1998).

[3] W. Chester, ''Resonant oscillations in closed tubes,'' J. Fluid Mech. **18**, 44–64 (1964).

[4] D. T. Blackstock, *Fundamentals of Physical Acoustics* (Wiley, New York, 2000).

[5] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics* (Pergamon, Oxford, New York, 1987).

[6] G. W. Swift, ''Thermoacoustic engines,'' J. Acoust. Soc. Am. **84**, 1145–1180 (1988).

[7] P. Merkli and H. Thomann, ''Transition to turbulence in oscillating pipe flow,'' J. Fluid Mech. **68**, 567–576 (1975).

[8] W. H. Press, S. A. Teukovsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in* FORTRAN (Cambridge University Press, Cambridge, England, 1992).

Atmospheric scattering for varying degrees of saturation and turbulent intermittency

David E. Norris[a)]
Graduate Program in Acoustics, The Pennsylvania State University, University Park, Pennsylvania 16802

D. Keith Wilson
Battlefield Environment Division, U.S. Army Research Laboratory, AMSRL-IS-EP, 2800 Powder Mill Road, Adelphi, Maryland 20783-1197

Dennis W. Thomson
Department of Meteorology, The Pennsylvania State University, University Park, Pennsylvania 16802

(Received 18 August 1999; revised 5 February 2001; accepted 16 February 2001)

Atmospheric turbulence is inherently inhomogeneous and intermittent. Short periods of high activity are embedded in longer periods of relative calm. Local spatial and temporal changes in sound speed associated with this intermittency increase the likelihood of measuring large values of scattered acoustic signals. Previous work successfully predicted the probability density functions (pdf's) of fully saturated, scattered signals measured within an acoustic shadow zone [Wilson *et al.*, J. Acoust. Soc. Am. **99**, 3393–3400 (1996)]. The more general case of incompletely saturated scattering is considered in this paper; using the Rice–Nagakami distribution a theory is developed. The predicted intensity pdf has two free parameters: one to describe the degree of intermittency and a second for the degree of saturation. For validation purposes, outdoor propagation measurements were made over a flat, hard ground at ranges of 146–283 m and at frequencies of 50–540 Hz. The saturation parameter was determined from the acoustic data and also estimated from the turbulence conditions. The degree of saturation increased with frequency, and measured intensity pdf's were found to be in excellent agreement with the theory.
[DOI: 10.1121/1.1364488]

PACS numbers: 43.28.Fp, 43.28.Vd, 43.60.Cg [LCS]

I. INTRODUCTION

Scattering refers to the spreading of sound in a variety of directions from discontinuities in the propagation medium. Obstacles, rough surfaces, and regions of turbulence all can produce scattering.[1] In the absence of physical obstructions, the main source of atmospheric scattering is turbulent inhomogeneities, which are known to occur at the boundary of thermal eddies or in regions of wind shear.

The properties of scattered acoustic signals depend strongly on the integral of the signal's multiple propagation paths. As scattering increases, the number of paths influencing a received signal increases. At full saturation, the integral properties of different paths are uncorrelated and, thus, any further increase in scattering produces no change in the statistics of a received signal. Fully saturated scattering typically occurs deep within an acoustic shadow zone. When some of the propagation paths are sufficiently close so as to be correlated, incomplete saturation is observed.

Established scattering theory is based upon the use of ensemble-averaged descriptors of the turbulence as the sole determinant of the scattering statistics.[2] The probability density function (pdf) of received intensity for fully saturated scattering is then described with an exponential distribution. Wilson *et al.*[3] modified this formulation by including a significant property of turbulence that is neglected in the classical approach. Turbulence is intermittent: periods of high activity are surrounded by periods of relative calm. An illustration of this property is given in the upper panels of Fig. 1, where snapshots of the turbulent field graphically depict the intermittent nature of the turbulence strength. Early examples of intermittency are observed in the SODAR-derived time series of ϵ by Thomson *et al.*,[4] where they use ϵ, the dissipation rate of turbulent kinetic energy, to quantify the strength of the turbulence.

By defining the turbulence descriptors locally, over a given scattering volume, Wilson *et al.* derived an expression for the intensity pdf that agreed well with measurements made deep within a shadow zone. Intermittency was found to increase the occurrence of large intensity deviations. Again referring to Fig. 1, the lower panels illustrate the effect turbulent intermittency would have on scatttered acoustic signals. The more intermittent the turbulence, the more likely the occurrence of large acoustic deviations, as seen in the scatter diagrams of the complex acoustic signal.

In this work, the statistics of the received intensity are studied for the case of incompletely saturated scattering. In Sec. II A, saturation is described in the context of scattering regimes. The work for fully saturated scattering is reviewed in Sec. II B and is then extended to the incompletely saturated case in Sec. II C. Section II D addresses estimation of the degree of saturation. The experiments are described in

[a)]Present address: BBN Technologies, 1300 N. 17th St., Arlington, VA 22209.

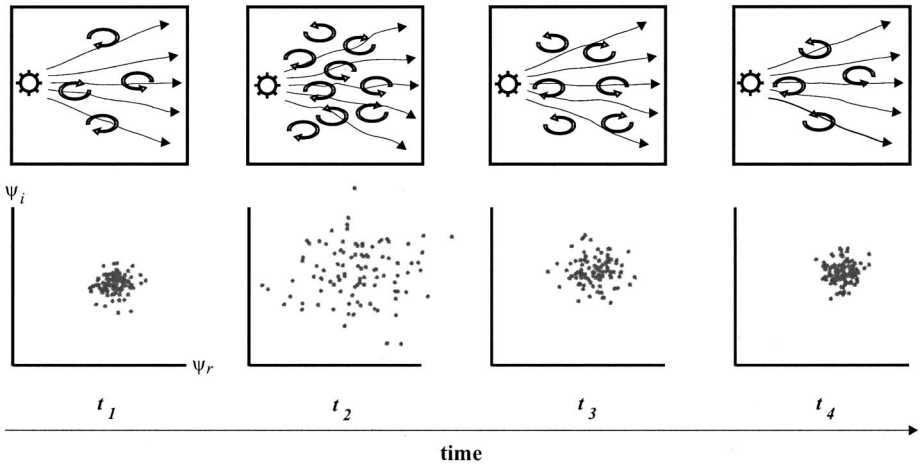


FIG. 1. Conceptual illustration of the concept of turbulent intermittency. Upper panels depict snapshots of the turbulence strength within a scattering volume. The lower panels are scatter diagrams of the propagated complex signal ($\psi_r + i\psi_i$). Intermittency is seen in the burst of strong turbulent activity (t_2), surrounded by periods of relative calm (t_1, t_3, t_4). The intermittency causes dramatic changes in the propagated signal statistics over time.

Sec. III, where comparisons are made between the measurements and predictions. Finally, Sec. IV discusses asymptotic limits, the scintillation index, phase variance, and applicability of the model.

II. THEORY

A. Scattering regimes

For isotropic, homogeneous turbulence, Flatté *et al.*[5] classified propagation by scattering into different regimes based upon two parameters: the diffraction parameter, Λ, and the strength parameter, Φ. The diffraction parameter is inversely proportional to the size of the turbulent inhomogeneities relative to the path's Fresnel zone radius, and the strength parameter characterizes the phase fluctuations of the received wave. Figure 2 shows the Λ–Φ space and regimes. (A similar classification is found in de Wolf.[6]) When Λ and Φ are both close to zero, the wave is not scattered and geometrical acoustics identifies the refractive path. For the case where $\Lambda > 0$ and $\Phi \gtrsim 1$, weak scattering from the turbulent inhomogeneities results in a diffracted path. Using Rytov's method, the signal can be characterized as a small perturbation from the invariant refracted path. For $\Phi \gtrsim 1$, the turbulence inhomogeneities are strong enough that multiple paths appear. At full saturation, the scattered paths become uncorrelated.

B. Full saturation

For fully saturated scattering, the classical approach begins by expressing the complex acoustical pressure in the form

$$\mathbf{p} = \psi_r + i\psi_i = A \exp(i\phi), \tag{1}$$

where A is the pressure amplitude and ϕ the phase. The real (ψ_r) and imaginary (ψ_i) part of $\mathbf{p}$ are assumed to have identical, zero-mean Gaussian pdf's with variance $\langle \psi^2 \rangle$. The resulting amplitude has a Rayleigh distribution in the form

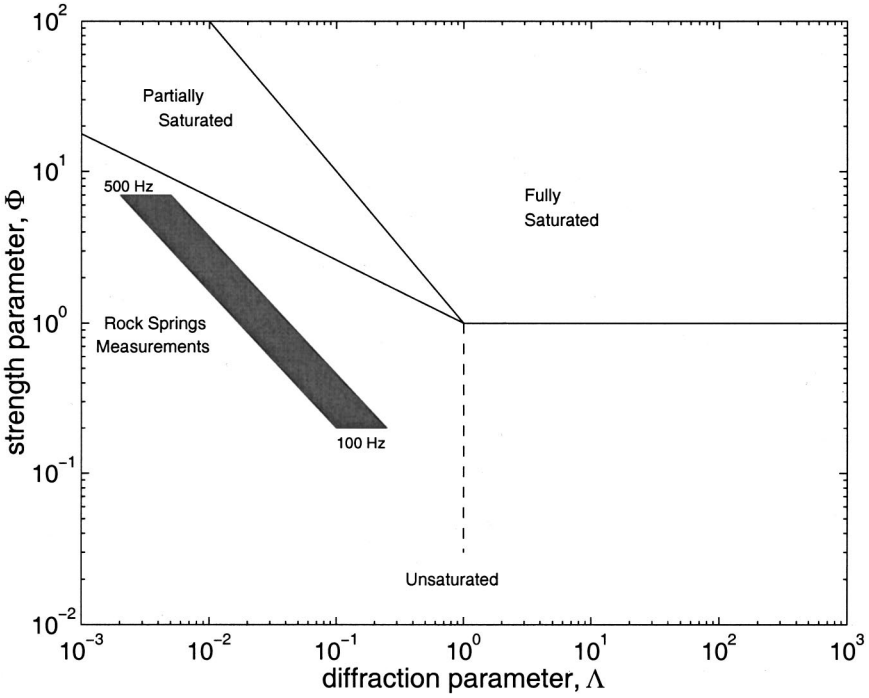


FIG. 2. Regions in Λ–Φ space [based upon Flatté *et al.* (1979), Ref. 5]. The shaded area indicates the approximate domain of the Rock Springs measurements, using the calculations in Sec. III D.

$$P(A)=\frac{A}{\langle\psi^2\rangle}\exp\left(-\frac{A^2}{2\langle\psi^2\rangle}\right). \tag{2}$$

The pdf for intensity, $I=A^2/2\rho_0 c_0$, can be derived from the probability identity[7]

$$P(I)=\frac{n_v P(A)}{|dI/dA|}, \tag{3}$$

where n_v is the valuation of the inverse function $p(I)$. In this case, $n_v=1$ and

$$P(I)=\frac{1}{I_0}\exp\left(-\frac{I}{I_0}\right), \tag{4}$$

where I_0 is used for the mean intensity ($I_0=\langle I\rangle=\langle\psi^2\rangle/\rho_0 c_0$). The mean intensity is related to ensemble-averaged statistics for the turbulence, as originally formulated by Kolmogorov.[8]

In a later work, Kolmogorov[9] presented his refined hypotheses which explicitly included the effects of intermittency by considering local rather than ensemble-averaged statistics. Obukhov[10] stated this concept by defining the strength of the turbulence as a local property of the measurement volume. Functionally, this can be written in terms of the dissipation rate of turbulent kinetic energy. The average dissipation rate within a sphere of volume V is

$$\tilde{\epsilon}=\frac{1}{V}\int_V \epsilon\, dV. \tag{5}$$

The tilde here and in all following formulations indicates a quantity defined locally by averaging over a given scattering volume V.[3]

By considering Kolmogorov's refined hypothesis, as well as the work of Gurvich *et al.*,[11] who studied intermittency in radio propagation, Wilson *et al.*[3] derived a revised intensity pdf for full saturation. Their theoretical development is based upon using a locally defined mean intensity, $\tilde{I}_0$, in the form

$$\tilde{I}_0=\int_0^\infty p(I|\tilde{\epsilon})I\, dI. \tag{6}$$

Here, $p(I|\tilde{\epsilon})$ is the probability density function for intensity conditioned upon the local value of turbulent kinetic energy dissipation, $\tilde{\epsilon}$. The theory also uses the log-normal $\tilde{\epsilon}$ distribution

$$P(\tilde{\epsilon})d\tilde{\epsilon}=\frac{1}{\sqrt{2\pi}\sigma_\epsilon}\exp\left[-\frac{(\ln\tilde{\epsilon}-m_\epsilon)^2}{2\sigma_\epsilon^2}\right]\frac{d\tilde{\epsilon}}{\tilde{\epsilon}}, \tag{7}$$

where $m_\epsilon=\langle\ln\tilde{\epsilon}\rangle$ and $\sigma_\epsilon^2=\langle(\ln\tilde{\epsilon}-m_\epsilon)^2\rangle$. This distribution is well documented in the atmospheric literature.[12–14] The final expression for the probability density function of intensity is[3]

$$P(I)=\frac{1}{\sqrt{2\pi}\sigma_\epsilon}\int_0^\infty\frac{1}{\tilde{I}_0}\exp\left(-\frac{1}{\tilde{I}_0}\right)\times\exp\left[-\frac{(\ln(\tilde{I}_0/\langle I\rangle)+\sigma_\epsilon^2/2)^2}{2\sigma_\epsilon^2}\right]\frac{d\tilde{I}_0}{\tilde{I}_0}. \tag{8}$$

One free parameter, σ_ϵ, characterizes the strength of the intermittency. It can be found for a given data set by imposing the condition that the experimental and theoretical pdfs have the same second moment. It can also be evaluated from the propagation environment, in which case estimates of the turbulence's integral length scale and scattering volume must be made.

Wilson *et al.*[3] concluded that intermittency increases the probability of observing large scattered intensities relative to the mean value. They showed that the intensity pdf in the presence of intermittency departs from the exponential distribution: as the intermittency parameter σ_ϵ increases, the pdf tail increases (see Fig. 4 of Ref. 3). In a later work, the additional influence of intermittency in the turbulent temperature field was considered.[15] The results indicated that scattering from a combination of temperature and wind inhomogeneities produces essentially the same pdf as will the wind inhomogeneities alone.

C. Incomplete saturation

We now consider the case where the requirement for full saturation is relaxed. To analyze scattering for incomplete saturation, the Rice–Nagakami distribution is used. The Rice–Nagakami distribution was developed, in part, by Rice's studies of the statistics of a stable signal in the presence of noise.[16] Flatté *et al.*[5] and de Wolf,[6] among others, mentioned the applicability of the Rice–Nagakami distribution in the context of incompletely saturated acoustic and electromagnetic signals, respectively. Birkemeier *et al.*[17] used the Rice–Nagakami pdf to study microwave scattering in the atmosphere. None of these references, however, considered the effects of intermittency.

We begin the theoretical development by assuming the pressure amplitude, A, has a Rice–Nagakami distribution in the form

$$P(A)=\frac{A}{\langle\psi^2\rangle}\exp\left(-\frac{A^2+A_u^2}{2\langle\psi^2\rangle}\right)\mathcal{I}_0\left(\frac{AA_u}{\langle\psi^2\rangle}\right), \tag{9}$$

where $\mathcal{I}_0$ is the modified Bessel function of the first kind. Here, instead of zero mean as in the fully saturated case, the components of the complex acoustic pressure, ψ_r and ψ_i, now each have mean A_u. From Eq. (3), the resulting intensity pdf is

$$P(I)=\frac{1}{I_0}\exp\left(-\frac{I+I_u}{I_0}\right)\mathcal{I}_0\left(\frac{\sqrt{II_u}}{I_0/2}\right), \tag{10}$$

where $I_u=A_u^2/2\rho_0 c_0$ is the intensity of the unperturbed signal, and I_0 is the mean intensity associated with the scattering. In this formulation, the mean intensity of the full signal $\langle I\rangle=I_0+I_u$.

To account for intermittency, we follow the same theoretical arguments as Wilson *et al.*[3] and extend the pdf to the local case (i.e., defined locally over a given scattering volume V), giving

$$P(I|\tilde{\epsilon})=\frac{1}{\tilde{I}_0}\exp\left(-\frac{I+I_u}{\tilde{I}_0}\right)\mathcal{I}_0\left(\frac{\sqrt{II_u}}{\tilde{I}_0/2}\right). \tag{11}$$

By definition, the intensity pdf can be written

$$P(I)=\int_0^\infty P(I|\tilde{\epsilon})P(\tilde{\epsilon})d\tilde{\epsilon}. \tag{12}$$

Wilson *et al.*[3] have shown that $\tilde{I}_0$ can be used as a surrogate for $\tilde{\epsilon}$. Therefore, Eqs. (12) and (7) can be rewritten in the form

$$P(I)=\int_0^\infty P(I|\tilde{I}_0)P(\tilde{I}_0)d\tilde{I}_0, \tag{13}$$

and

$$P(\tilde{I}_0)d\tilde{I}_0=\frac{1}{\sqrt{2\pi\sigma}}\exp\left(-\frac{(\ln\tilde{I}_0-m)^2}{2\sigma^2}\right)\frac{d\tilde{I}_0}{\tilde{I}_0}, \tag{14}$$

where

$$\sigma=\langle(\ln\tilde{I}_0-m)^2\rangle=(2/3)\sigma_\epsilon. \tag{15}$$

The variable m is defined such that the mean of $P(I)$ is $\langle I\rangle$, which, from Eq. (A13), gives

$$m=\ln(\langle I\rangle-I_u)-\sigma^2/2. \tag{16}$$

Using Eq. (16) in (14) and substituting the result into (13), the final expression for the intensity pdf is found to be

$$P(I)=\frac{1}{\sqrt{2\pi\sigma}}\int_0^\infty\frac{1}{\tilde{I}_0}\exp\left(-\frac{I+I_u}{\tilde{I}_0}\right)\mathcal{I}_0\left(\frac{\sqrt{II_u}}{\tilde{I}_0/2}\right)\times\exp\left[-\frac{(\ln(\tilde{I}_0/\langle I\rangle-I_u)+\sigma^2/2)^2}{2\sigma^2}\right]\frac{d\tilde{I}_0}{\tilde{I}_0}, \tag{17}$$

with the condition that the mean intensity $\langle I\rangle$ is greater than the unperturbed intensity I_u. The integral in Eq. (17) cannot be solved explicitly, but it is straightforward to evaluate numerically.

The two model parameters are the intermittency parameter, σ, and the saturation parameter, S. In the Appendix, the saturation parameter is shown to take the form

$$S=1-I_u/\langle u\rangle. \tag{18}$$

The saturation parameter, which characterizes the degree of saturation, ranges from 0 to 1. For the fully saturated case ($S=1$), the intensity of the unperturbed signal becomes zero and the intensity pdf reduces to the result of Wilson *et al.*[3]

Both σ and S can be determined for a given data set by imposing the condition that the experimental and theoretical pdfs have the same second moment (variance) and third moment (skewness). Equations for the moments are developed in the Appendix. Calculated moments from the measurements can be plugged into these equations and S and σ solved for using the Levenberg–Marquardt method.[18] Figures 3 and 4 provide a graphical depiction of the dependence of the normalized variance [Eq. (A23)] and normalized skewness [Eq. (A24)] on the saturation and intermittency parameter. For weak scattering ($S<0.2$), the intermittency has little effect on the moments. Thus, moment-matched values of σ in this region are likely to be unreliable.

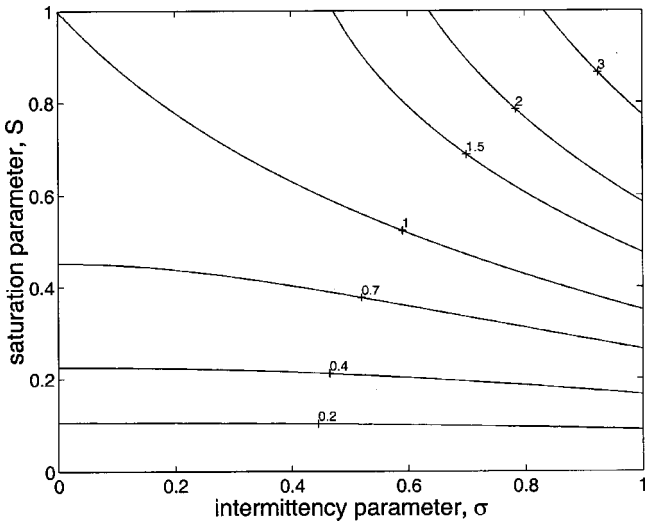


FIG. 3. Normalized variance (second moment) of intensity, m_2', as a function of the saturation parameter S and intermittency parameter σ.

D. Saturation parameter

When the diffraction parameter Λ is much greater than 1, Flatté *et al.* suggest that amplitude statistics can be described with a Rice–Nagakami pdf, where

$$\frac{A_u}{\langle A^2\rangle^{1/2}}=\exp\left(-\frac{\Phi^2}{2}\right). \tag{19}$$

Nothing that $A_u^2/\langle A^2\rangle=I_u/\langle I\rangle$, we have from Eq. (18) and Flatté *et al.*'s Eq. (8.3.9)

$$S=1-\exp(-\Phi^2). \tag{20}$$

When the source–receiver range, R, is much greater than the integral length scale, $\mathcal{L}$, the strength parameter Φ can be given by[5]

$$\Phi^2=2k^2\langle\mu^2\rangle R\mathcal{L}, \tag{21}$$

where $\langle\mu^2\rangle$ is the variance of the refractive-index fluctuations. The variance $\langle\mu^2\rangle$ can be approximated from the velocity fluctuations, u, and nominal sound speed, c_0, using

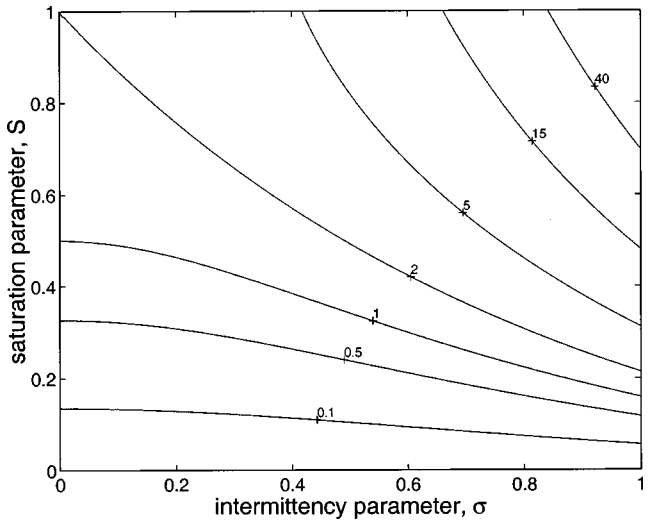


FIG. 4. Same as Fig. 3, but for normalized skewness (third moment) of intensity, m_3'.

TABLE I. Summary of the test runs and meteorological conditions for the 1995 and 1996 Rock Springs field experiments. See the text for explanation of variables.

Run	Date	Time (local)	Range (m)	Recording (min)	$\overline{\mu_p}$ (m/s)	$\Delta\overline{T}$ (K)	$\langle\mu^2\rangle$	Refractive conditions
RS95-A	18 Mar 95	1206	183	50	2.9	1.1	1.5E−5	downward
RS95-B	18 Mar 95	1307	274	50	1.8	0.7	1.6E−5	downward
RS96-A	1 Nov 96	0948	146	20	−6.2	<0.1	1.4E−5	upward
RS96-B	1 Nov 96	0948	196	20	6.2	<0.1	1.4E−5	downward

$$\langle\mu^2\rangle=\frac{\langle u^2\rangle}{c_0^2}. \tag{22}$$

Traditional approaches to determining turbulent scales are usually restricted to the inertial subrange.[19] Calculated lengths scales are typically on the order of 1 m and do not account for the large-scale dynamics of the turbulence. In the context of scattering parameters, Wilson[20] has derived turbulent scale relationships, based on the isotropic von Kármán turbulence model, that include both small- and large-scale atmospheric effects. Specifically, the equation for the integral length scale is

$$\mathcal{L}=1.3z\,\frac{1+0.015(z_i/z)(-z_i/\kappa L_{\mathrm{MO}})^{2/3}}{1+0.12(-z_i/\kappa L_{\mathrm{MO}})^{2/3}}, \tag{23}$$

where z is the height above ground, z_i the atmospheric boundary-layer inversion height, κ von Kármán's constant, and L_{MO} the Monin–Obukhov length.

III. EXPERIMENT

A. Setup

Two field experiments were conducted at Penn State University's micrometeorological field site near Rock Springs, PA. For both experiments, the test site was flat and clear of significant cover, and the ground surface was hard, compact earth.

The first test, referred to as RS95, was conducted on 15 March 1995. A source at a height of 1.5 m was programed to radiate continuous tones at 90, 200, 350, 500, and 660 Hz. The receiver consisted of a microphone positioned at a height of 1.0 m on a portable stand. A DAT recorder was used to make 50-min recordings at ranges of 183 (RS95-A) and 274 m (RS95-B). Winds were light and the sky clear. Recorded concurrently with the acoustic data were wind velocities measured with a sonic anemometer atop a 10-m tower at the source. Hourly averaged temperature data were also collected at 2 and 10 m.

A second test, RS96, was conducted on 1 November 1996. During these measurements, the receiver consisted of a microphone array mounted 6 m high on a tower. A nominally downwind source was positioned at a range of 146 m (RS96-A), and an upwind source at 196 m (RS96-B). The upwind source radiated continuous tones at 50, 110, 190, 260, 325, 395, 460, and 525 Hz. The downwind source radiated tones at 70, 130, 205, 275, 340, 410, 470, and 540 Hz. A 16-channel DAT recorder was used to make 60-min recordings. Winds were moderate to strong and the sky overcast. Velocity data were recorded with a sonic anemometer located near the receiver. As during RS95, hourly averaged temperature data were collected at 2 and 10 m. For a detailed description of the RS96 experiment, the reader is referred to Ref. 21.

A summary of the test parameters and meteorological conditions is given in Table I for both the RS95 and RS96 experiments. Wind and temperature data are hourly averages, where $\overline{u_p}$ is the horizontal wind speed at 10-m height projected along the propagation path, and $\Delta\overline{T}$ is the temperature at 2 m minus that at 10 m.

B. Data collection and analysis

The RS95 acoustic data were collected on digital audiotape at a sample rate of 5000 Hz. Due to hardware limitations, the digital signal was converted back to analog and resampled at 8000 Hz on a Sun Sparc 5 workstation. Any introduced aliasing and quantization errors were negligible since the sampling frequency was a factor of 8 greater than the highest source frequency and the sampling had 16-bit resolution. Once transferred to the workstation, the time data were segmented into 1-s intervals. The intervals were windowed with a Hann function, zeros padded, and transformed into the frequency domain using a 8192-point fast Fourier transform (FFT). The resulting FFT bin width was 0.9766 Hz. A peak peaking algorithm was used to recover the signal magnitude at the frequency bin nearest the source tone peaks. Noise levels were estimated from the signal variance at a frequency near each source tone, and signal-to-noise ratios (SNR) were calculated. The mean SNR was greater than 15 dB for all frequencies except 660 Hz, where it dropped considerably due to the low signal strength. As a result, 660-Hz data were not available for analysis. Approximately 3000 data points over the 50-min data sets were available to compute the intensity pdf. The normalized intensity was calculated by squaring the signal magnitude and scaling so that the mean was 2.

For the RS96 data, a Hilbert transform technique was used on the first 20 min of the record to recover the time series at each source frequency. Specifically, the entire acoustic time series, recorded at a sample rate of 6000 Hz, was first zero padded and transformed into the frequency domain using a 2^{23}-point two-sided FFT. The resulting spectrum, $S(f)$, had a FFT bin width of 7.15×10^{-4} Hz. Signal-to-noise ratios were calculated directly from the power spectral density. Due to microphone wind noise, data at the lowest frequencies, 50 and 70 Hz, had low SNR values and were not analyzed further. At the remaining source frequencies, filtered spectra, $S_w(f)$, were computed by applying a

TABLE II. Moments and model parameters. See the text for explanation of variables.

Run	Frequency (Hz)	Moments		Matched parameters	
		m_2'	m_3'	σ	S
RS95-A	90	0.0	0.1	0.84	0.01
	200	0.1	0.0	0.63	0.03
	350	0.3	0.2	0.51	0.15
	500	0.7	1.3	0.46	0.40
RS95-B	90	0.1	0.0	0.00	0.05
	200	0.2	0.1	0.00	0.10
	350	0.7	1.5	0.73	0.31
	500	2.0	16.1	0.78	0.79
RS96-A	130	0.2	0.0	0.28	0.08
	205	0.2	0.1	0.00	0.11
	275	0.4	0.3	0.42	0.21
	340	0.7	0.9	0.42	0.37
	410	0.9	1.7	0.44	0.51
	470	1.2	3.4	0.44	0.75
	540	1.4	5.4	0.40	1.00
RS96-B	110	0.2	0.5	0.98	0.10
	190	0.8	0.5	0.00	0.40
	260	0.3	0.3	0.57	0.15
	325	0.3	0.2	0.25	0.16
	395	1.5	4.8	0.39	0.11
	460	1.8	7.7	0.46	1.00
	525	1.2	3.9	0.53	0.68

Hann window to the double-sided spectral peaks.[18] The width of the Hann window was roughly twice the peak width (0.2 Hz at 50 Hz to 4 Hz at 540 Hz). The Hilbert transform of the filtered spectra, $S_w^h(f)$, was computed as

$$S_w^h(f) = \mathrm{sgn}(f)\exp(i\pi/2)S_w(f), \tag{24}$$

where $\mathrm{sgn}(f)$ is the sign of the source frequency f.[7] The intensity time series, $I(t)$, for a given source frequency was then computed from

$$I(t) = |\mathcal{F}^{-1}(S_w(f)) + i\mathcal{F}^{-1}(S_w^h(f))|^2, \tag{25}$$

where $\mathcal{F}^{-1}$ is the inverse FFT operator. The resulting time step was approximately 0.04 s, giving over 28 000 data points for evaluation of the intensity pdfs.

C. Comparisons

The parameters σ and S are determined from the measured second and third moments of the data. The measurements are equated to Eqs. (A23) and (A24), as described in Sec. II C. The results are summarized in Table II for runs RS95-A, -B, and runs RS96-A, -B. The normalized variance (second moment) and skewness (third moment) of intensity are m_2' and m_3', respectively. The mean intensity $\langle I\rangle = 2$.

Data from RS95-B are considered for detailed comparison. The degree of saturation, as indicated by S, increased with frequency, from almost unsaturated propagation at 90 Hz to nearly fully saturated propagation at 500 Hz. Comparisons between the calculated and predicted intensity pdfs are shown in Figs. 5–8. Agreement is excellent. The pdf magnitudes at the peak and tail are the significant features for comparison. The peak value is associated with the correlated

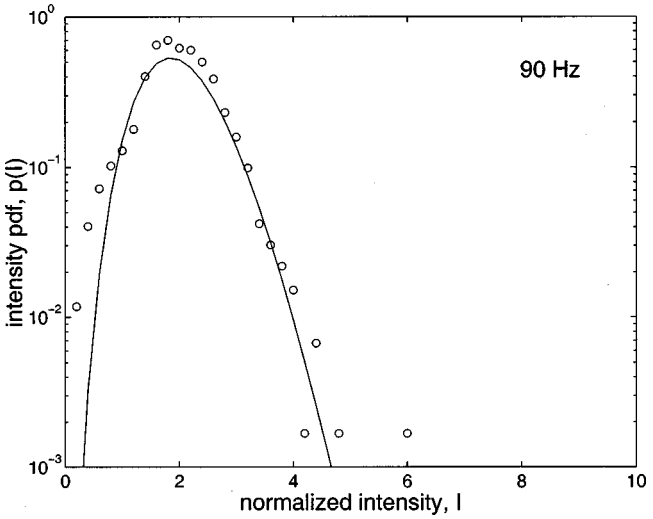


FIG. 5. Experimentally determined intensity pdf (circles) and model prediction (solid line) for a frequency of 90 Hz. The intensity is normalized so that $\langle I\rangle = 2.0$, $S = 0.05$, and $\sigma = 0.0$.

propagation paths. It ranges from zero, for the fully saturated case ($S=1$), up to the mean value of $\langle I\rangle = 2$ for the unsaturated case ($S=0$). The magnitude of the tail changes in response to the intermittency: the stronger the intermittency, the greater the probability of higher intensities. Both of these features are predicted well by the model for all four frequencies considered. Note that for the two higher frequencies, 350 and 500 Hz, the measured pdf varies about the predictions for values above about twice the mean. This variability is observed for both RS95 (1-s time step) and RS96 (0.04-s time step), suggesting that the dominant variability in intensity has time scales of the order 1 s or greater.

D. Estimation of saturation parameter

Our method for estimating the saturation parameter is based on Eqs. (20)–(23). Values for the variance $\langle u^2\rangle$, the heights z and z_i, and stability parameter L_{MO} are required.

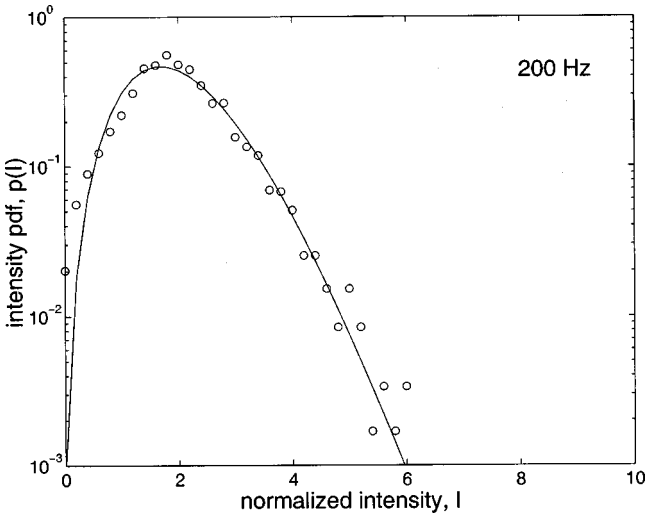


FIG. 6. Same as Fig. 5, but for a frequency of 200 Hz, $S = 0.10$, and $\sigma = 0.0$.

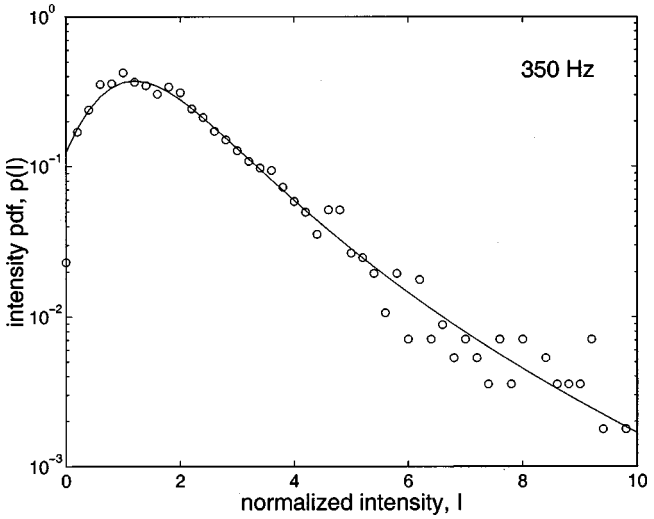


FIG. 7. Same as Fig. 5, but for a frequency of 350 Hz, $S=0.31$, and $\sigma=0.73$.

The along-path wind velocity variance $\langle u^2\rangle$ is computed from the sonic anemometer data, and the reference sound speed c_0 is found from the mean temperature and wind. Equation (22) is then used to find the refractive index variance $\langle \mu^2\rangle$. Results are shown in Table I.

The height z for RS95 is set to the source–receiver height of 1.0 m. For RS96, the determination of z is more ambiguous since the source is at 1 m and the receiver at 6 m. The mean value of 3.5 m is chosen as a representative value.

The inversion height z_i and Monin–Obukhov length L_{MO} are driven by the atmospheric forcing in the boundary layer.[22] During the 50-min collection period of RS95, the weather conditions were sunny with low wind. The values $z_I=1000$ m and $L_{\mathrm{MO}}=-10$ m are used as representative values for these conditions.[23] From Eq. (23), the resulting value for $\mathcal{L}=108$ m. For RS96, the detailed atmospheric measurements can be used to determine z_I and L_{MO} by fitting temperature and velocity data to mean profiles and energy spectra.[21] The calculations give $L_{\mathrm{MO}}=-1320$ m and $z_I=693$ m. The corresponding value of $\mathcal{L}$ is 15 m.

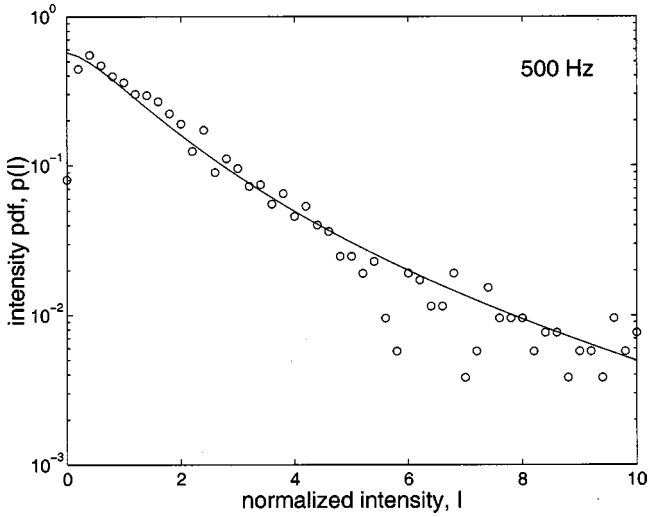


FIG. 8. Same as Fig. 5, but for a frequency of 500 Hz, $S=0.79$, and $\sigma=0.78$.

The saturation parameter S is estimated from the meteorological parameters using Eqs. (21) and (20). The results are shown in Figs. 9 and 10 for RS95 and RS96, respectively, along with the moment-matched values. All moment-matched data monotonically increase with frequency with the exception of RS95-B, where the strong presence of a ground-reflected ray path may be affecting the received signal. The RS96 estimates generally agree to within a few orders of magnitude.[24] Estimates for RS95 are much greater than observed. This mismatch may be due, in part, to an overestimation of the strength of the convective boundary layer, which would lead to an overestimate of $\mathcal{L}$ and S.

E. Discussion

1. Asymptotic limits

In the asymptotic limit of the ratio of the real amplitude's mean to variance, $A_u/\langle\psi^2\rangle \ll 1$, the modified Bessel function $\mathcal{I}_0$ goes to 1, and the Rice–Nagakami distribution, Eq. (9), simplifies to the Rayleigh distribution, Eq. (2). For the diffraction parameter Λ much greater than 1, we refer to Eq. (20) and find that the asymptotic limit is reached when the strength parameter $\Phi \gg 1$. This places the scattering within the fully saturated region of Flatté *et al.*, where the Rayleigh distribution applies.

In the opposite asymptotic limit, $A_u/\langle\psi^2\rangle \gg 1$, the modified Bessel function can be approximated as $\mathcal{I}_0(x)=1/\sqrt{2\pi x}\exp(x)$,[25] and the resulting Rice–Nagakami distribution [Eq. (9)] becomes

$$P(A)=\sqrt{\frac{A}{A_u}}\,\frac{1}{\sqrt{2\pi\langle\psi^2\rangle}}\exp\left(\frac{-(A-A_u)^2}{2\langle\psi^2\rangle}\right). \tag{26}$$

This function is observed to be Gaussian in shape with center A_u and variance $\langle\psi^2\rangle$.[26] For $\Lambda \gg 1$, the asymptotic limit is reached for $\Phi \ll 1$, placing the scattering within the unsaturated region. Here, Flatté *et al.*[5] used the log-normal distribution. This is consistent with the asymptotic limit, since the log-normal distribution can be approximated by the Gaussian distribution when the log mean is much greater than the log variance.

2. Scintillation index

A common parameter used in the context of scattering is the scintillation index, defined as[27]

$$\sigma_{\mathrm{scint}}=\frac{\langle I^2\rangle-\langle I\rangle^2}{\langle I\rangle^2}. \tag{27}$$

Using Eq. (A18), it can be expressed in terms of the saturation and intermittency parameters

$$\sigma_{\mathrm{scint}}=2S^2\exp\sigma^2-3S^2+2S. \tag{28}$$

Contours of σ_{scint} are given in Fig. 11 for various values of σ and S. For no intermittency ($\sigma=0$), σ_{scint} is one at full saturation ($S=1$). As the intermittency increases, σ_{scint} approaches 1 for smaller and smaller values of S. In addition, supersaturation, in which $\sigma_{\mathrm{scint}}>1$, can occur as a conse-

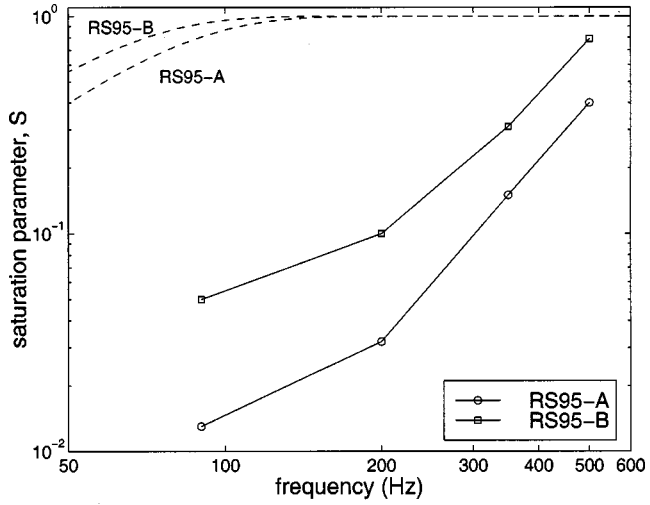


FIG. 9. Measured (solid line) and estimated (dashed line) saturation parameters for RS95.

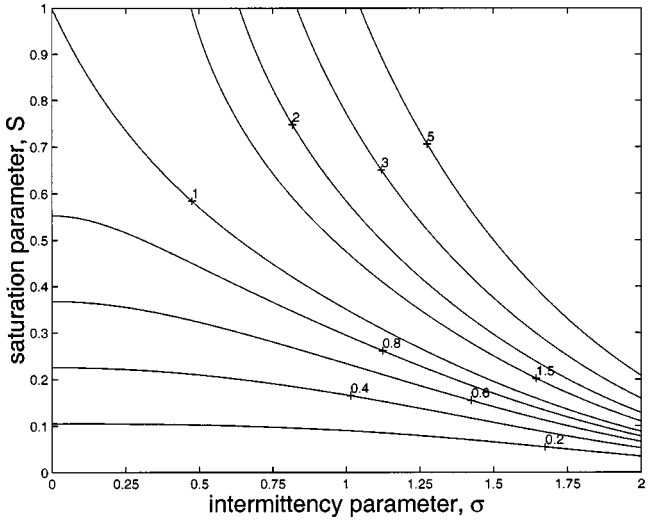


FIG. 11. Contours of scintillation index as function of intermittency and saturation parameter.

quence of intermittency. At $\sigma=1$, supersaturation is predicted for $S>0.4$.

An increase in intermittency results in higher signal variability and a higher variance in intensity. The scintillation index in turn increases due to its formulation in terms of second-order statistics. Because of this statistical dependence, this study suggests that, for cases of significant intermittency, the scintillation index is of limited use as a saturation descriptor.

3. Phase variance

The saturation parameter estimates of Sec. III D are in poor agreement with measurements. These estimates are based in part on the integral length scale, $\mathcal{L}$, as computed from Eq. (23). The validity of $\mathcal{L}$ can be tested by comparing measured phase variance with predictions.

In Fig. 2, the shaded region indicates the scattering domain for the Rock Spring measurements. All data fall within the geometrical optics region, where the phase variance is $\Phi^{2.5}$ $\mathcal{L}$ is related to Φ from Eq. (21).

Figure 12 gives the phase variance comparisons for the 1996 experiment. The agreement is good for RS96-A and excellent for RS96-B. Comparison were also made for the RS95 data with similar results. The agreement between these predictions and data supports the validity of Eq. (23). Therefore, the weak saturation parameter predictions of Sec. III D may result from the questionable validity of Eq. (20).

4. Applicability

The original derivation of the Rice–Nagakami distributions was in the context of a deterministic signal in the presence of noise. For our application, it is worthwhile to consider an interpretation of the acoustic signal in a similar context. The deterministic signal can be interpreted as that part of the scattered signal associated with a stable, refractive path. Noise can then be associated with the perturbations about it. The displaced mean, A_u, indicates the relative strength of the refractive path with respect to the perturba-

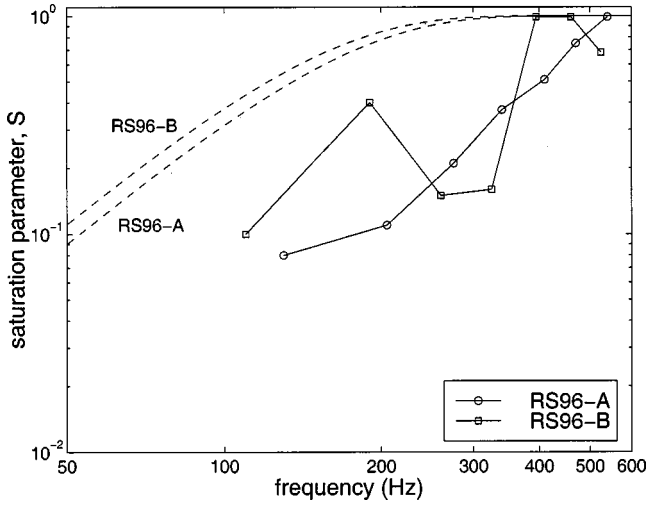


FIG. 10. Measured (solid line) and estimated (dashed line) saturation parameters for RS96.

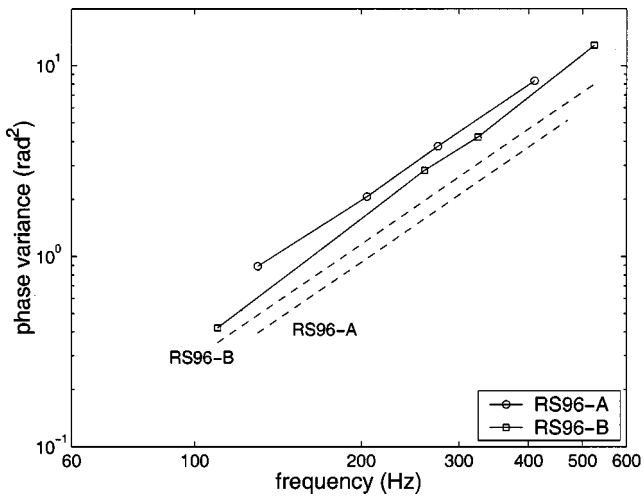


FIG. 12. Measured (solid line) and estimated (dashed line) phase variances for RS95.

tions. In the limit of full saturation, A_u becomes zero, as the paths become uncorrelated and the perturbations dominate the scattering.

Flatté *et al.* restrict their discussion of the Rice–Nagakami distribution to the region $\Phi \approx 1$ and $\Lambda \gg 1$. The calculation of these scattering parameters is strongly dependent on the turbulent length scales. If the turbulence is restricted to the inertial subrange, the scattering for RS95 and RS96 falls within the above range. However, for the more realistic length scales that include the large-scale turbulence [Eqs. (27) and (38) of Ref. 28], the scattering region falls in the range $\Phi \approx 1$ and $\Lambda \ll 1$. The excellent agreement between the predicted and measured intensity pdf's suggests that the Rice–Nagakami distribution as a descriptor for the scattering statistics can be extended for $\Lambda \ll 1$.

The extension of the theory to estimate the saturation parameter (Figs. 9 and 10) also assumes $\Lambda \ll 1$. The inconclusive comparisons with the moment-derived values suggest that this approach, in particular Eq. (20), may not be applicable, at least for this particular data set.

IV. CONCLUSION

Turbulent intermittency causes temporal and spatial variability in the scattering cross section and results in a departure from the exponential distribution for fully saturated intensity. Using a Rice–Nagakami distribution, we can account for correlated propagation paths for incompletely saturated signals. This distribution is used to derive an expression for the intensity pdf that has two free parameters; one quantifies the degree of intermittency and the other, the saturation.

The theoretical pdf was compared to measurements over flat terrain at ranges under 300 m. Free parameters were determined by matching the second and third moments of the data and model. Agreement was found to be excellent.

The saturation parameter was estimated from the scattering strength parameter, which was found, in part, from scattering length scales that include the influence of large-scale turbulence. Predictions of the saturation parameter from micrometeorological data were not reliable.

ACKNOWLEDGMENTS

This research was supported by Army Research Office Grant No. DAAH04-93-G-0153, administered by Walter E. Bach. For support during the field experiments, we would like to thank David Swanson, Richard Thompson, and Edward Maniet.

APPENDIX: MEAN AND MOMENTS FOR THE INTENSITY

Starting with Eq. (13), which relates the pdf of intensity $P(I)$ to the conditional probability $P(I|\tilde{I}_0)$, it follows that the nth moment of intensity can be expressed as

$$\langle I^n \rangle = \int_0^\infty P(I) I^n \, dI = \int_0^\infty f_n(\tilde{I}_0) P(\tilde{I}_0) \frac{d\tilde{I}_0}{\tilde{I}_0}, \tag{A1}$$

where we introduce the function

$$f_n(\tilde{I}_0) = \int_0^\infty \tilde{I}_0 I^n P(I|\tilde{I}_0) dI. \tag{A2}$$

When the Rice–Nagakami pdf is used for $P(I|\tilde{I}_0)$, we have

$$f_n(\tilde{I}_0) = \int_0^\infty I^n \exp\left(-\frac{I+I_u}{\tilde{I}_0} \right) \mathcal{I}\left(\frac{\sqrt{II_u}}{\tilde{I}_0/2} \right) dI. \tag{A3}$$

Using the identity

$$\mathcal{I}(z) = \frac{1}{2\pi} \int_0^{2\pi} \exp(z \cos\theta) d\theta, \tag{A4}$$

and making a transformation from polar to rectangular coordinates, $f_n(\tilde{I}_0)$ can be expressed as a double integral in the form

$$f_n(\tilde{I}_0) = \frac{\tilde{I}_0^{n+1}}{\pi} \int_{-\infty}^\infty \int_{-\infty}^\infty (x^2 - 2\sqrt{I_u/\tilde{I}_0} x + I_u/\tilde{I}_0 + y^2)^n \times \exp(-x^2 - y^2) dx \, dy. \tag{A5}$$

It is straightforward to carry out the integration for various values of n. The first three expressions are

$$f_1(\tilde{I}_0) = \tilde{I}_0 I_u + \tilde{I}_0^2, \tag{A6}$$

$$f_2(\tilde{I}_0) = \tilde{I}_0 I_u^2 + 4\tilde{I}_0^2 I_u + 2\tilde{I}_0^3, \tag{A7}$$

and

$$f_3(\tilde{I}_0) = \tilde{I}_0 I_u^3 + 9\tilde{I}_0^2 I_u^2 + 18\tilde{I}_0^3 I_u + 6\tilde{I}_0^4. \tag{A8}$$

The resulting first three moments from Eq. (A1) are

$$\langle I \rangle = I_u + \langle \tilde{I}_0 \rangle, \tag{A9}$$

$$\langle I^2 \rangle = I_u^2 + 4I_u \langle \tilde{I}_0 \rangle + 2\langle \tilde{I}_0^2 \rangle, \tag{A10}$$

$$\langle I^3 \rangle = I_u^3 + 9I_u^2 \langle \tilde{I}_0 \rangle + 18 I_u \langle \tilde{I}_0^2 \rangle + 6\langle \tilde{I}_0^3 \rangle. \tag{A11}$$

It is more convenient to write these results in terms of the intermittency parameter, σ. We note that the nth moment for the log-normal random variable $\tilde{I}_0$ can be written in the form[29]

$$\langle \tilde{I}_0^n \rangle = \exp(nm + n^2\sigma^2/2), \tag{A12}$$

where $m = \langle \ln(\tilde{I}_0) \rangle$. Substituting Eq. (A12) evaluated at $n = 1$ into (A9), the mean becomes

$$\langle I \rangle = I_u + \exp[m + \sigma^2/2]. \tag{A13}$$

For the higher-order moments, it is more convenient to write Eq. (A12) in the equivalent form

$$\langle \tilde{I}_0^n \rangle = \langle \tilde{I}_0 \rangle^n \exp[n(n-1)\sigma^2/2]. \tag{A14}$$

Substituting this equation at $n = 2$ and 3 into Eqs. (A10) and (A11), the higher-order moments become

$$\langle I^2 \rangle = I_u^2 + 4I_u \langle \tilde{I}_0 \rangle + 2\langle \tilde{I}_0 \rangle^2 \eta, \tag{A15}$$

$$\langle I^3 \rangle = I_u^3 + 9I_u^2 \langle \tilde{I}_0 \rangle + 18 I_u \langle \tilde{I}_0 \rangle^2 \eta + 6\langle \tilde{I}_0 \rangle^3 \eta^3, \tag{A16}$$

in which $\eta = \exp(\sigma^2)$.

We now introduce the saturation parameter, $S=\langle \tilde{I}_0\rangle/\langle I\rangle$. When Eq. (A9) is normalized by $\langle I\rangle$, the intensity of the unperturbed signal, I_u, is expressible as

$$I_u=\langle I\rangle(1-S). \tag{A17}$$

The normalized nth moment is defined as $m_n=\langle I^n\rangle/\langle I\rangle^n$. Substituting Eq. (A17) into Eqs. (A15) and (A16) and rearranging, the second and third normalized moments are

$$m_2=2S^2\eta-3S^2+2S+1, \tag{A18}$$

$$m_3=6S^3\eta^3-18S^3\eta+18S^2\eta+8S^3-15S^2+6S+1. \tag{A19}$$

Let us indicate the nth moment about the mean using the following notation:

$$m_n'=\frac{\langle (I-\langle I\rangle)^n\rangle}{\langle I\rangle^n}. \tag{A20}$$

The second moment m_2' is referred to as the normalized variance and the third moment m_3' as the normalized skewness. It can be shown directly from the definitions that

$$m_2'=m_2-1, \tag{A21}$$

$$m_3'=m_3-3m_2+2. \tag{A22}$$

Substituting with Eqs. (A18) and (A19)

$$m_2'=2S^2\eta-3S^2+2S, \tag{A23}$$

$$m_3'=6S^3\eta^3+12S^2\eta-18S^3\eta+8S^3-6S^2. \tag{A24}$$

In Eqs. (A23) and (A24), we have expressions for the normalized variance and skewness of intensity in terms of the saturation and intermittency parameters. Their application is described in Sec. II C.

[1] A. Pierce, *Acoustics: An Introduction to its Physical Principles and Applications* (Acoustical Society of America, Woodbury, NY, 1989).
[2] V. I. Tatarskii, *The Effects of the Turbulent Atmosphere on Wave Propagation* (Keter, Jerusalem, 1971).
[3] D. K. Wilson, J. C. Wyngaard, and D. I. Havelock, ''The effect of turbulent intermittency on scattering into an acoustic shadow zone,'' J. Acoust. Soc. Am. **99**, 3393–3400 (1996).
[4] D. W. Thomson, R. L. Coulter, and Z. Warhaft, ''Simultaneous measurements of turbulence in the lower atmosphere using Sodar and aircraft,'' J. Appl. Meteorol. **17**, 723–734 (1978).
[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] D. A. de Wolf, ''Propagation regimes for turbulent atmospheres,'' Radio Sci. **10**, 53–57 (1975).
[7] J. S. Bendat and A. G. Piersol, *Random Data: Analysis and Measurement Procedures*, 2nd ed. (Wiley, New York, 1986).
[8] A. N. Kolmogorov, ''Local structure of turbulence in incompressible fluid,'' C. R. Acad. Sci. Paris **30**, 299 (1941).
[9] A. N. Kolmogorov, ''A refinement of previous hypotheses concerning the local structure of turbulence in a viscous incompressible fluid at high reynolds number,'' J. Fluid Mech. **13**, 82–85 (1962).
[10] A. M. Obukhov, ''Some specific features of atmospheric turbulence,'' J. Fluid Mech. **13**, 77–81 (1962).
[11] A. S. Gurvich and V. P. Kukharets, ''The influence of intermittence of atmospheric turbulence on the scattering of radio waves,'' Sov. J. Commun. Technol. Electron. **30**, 52–58 (1986) (English translation).
[12] A. S. Gurvich and T. Makarova, ''Fluctuations in the rate of turbulent kinetic energy dissipation as related to local wind speed spectra,'' Izv. Atmos. Ocean. Phys. **24**, 743–746 (1988) (English translation).
[13] V. A. Bezverkhniy, A. S. Gurvich, T. Makarova, and M. Z. Kholmyanskiy, ''Variation of local temperature and wind speed spectra in the atmospheric surface layer,'' Izv. Atmos. Ocean. Phys. **24**, 519–527 (1988) (English translation).
[14] V. P. Kukharets, ''Investigation of variability of the structure characteristic of temperature in the atmospheric boundary layer,'' Izv. Atmos. Ocean. Phys. **24**, 578–582 (1988) (English translation).
[15] D. K. Wilson, ''Scattering of acoustic waves by intermittent temperature and velocity fluctuations,'' J. Acoust. Soc. Am. **101**, 2980–2982 (1997).
[16] S. O. Rice, ''Mathematical analysis of random noise,'' in *Selected Papers on Noise and Stochastic Processes*, edited by N. Wax (Dover, New York, 1954), pp. 133–294.
[17] W. P. Birkemeier, D. H. Sargeant, J. P. Aasterud, H. S. Merrill, Jr., D. W. Thomson, I. H. Gerks, C. M. Beamer, and G. T. Bergemann, ''A study of the lower atmosphere using scattering of microwaves,'' Technical Report No. 2, Dept. of Meteorol. and Dept. of Elec. Engr., University of Wisconsin, 1965.
[18] W. H. Press, S. A. Teukolsky, and W. T. V. B. P. Flannery, *Numerical Recipes in FORTRAN*, 2nd ed. (Cambridge University Press, Cambridge, England, 1992).
[19] D. E. Norris, L. Kristensen, J. Mann, D. W. Thomson, and D. C. Swanson, ''Characterization of magnitude and phase fluctuations in the presence of atmospheric turbulence,'' J. Acoust. Soc. Am. (submitted).
[20] D. K. Wilson, ''A turbulence spectral model for sound propagation in the atmosphere that incorporates shear and buoyancy forcings,'' J. Acoust. Soc. Am. **108**, 2021–2038 (2000).
[21] D. E. Norris, ''Atmospheric Acoustic Propagation: Characterization of Magnitude and Phase Variability,'' Ph.D. thesis, Graduate Program in Acoustics, The Pennsylvania State University, 1998.
[22] J. C. Kaimal and J. J. Finnigan, *Atmospheric Boundary Layer Flows: Their Structure and Measurement* (Oxford University Press, New York, 1994).
[23] No weather fronts passed over the test site during data collection. Therefore, z_i and L_{MO} are predominantly driven by the solar forcing and can be assumed not to have changed significantly.
[24] As pointed out by an anonymous reviewer, the theory does not account for finite ground impedance effects, such as surface waves, which could have influenced the measurements.
[25] M. Abramowitz and I. A. Stegun, editors, *Handbook of Mathematical Functions* (Dover, New York, 1965).
[26] D. C. Swanson, *Signal Processing for Intelligent Sensor Systems* (Marcel-Dekker, New York, 2000).
[27] Y. A. Kravtsov, ''Propagation of electromagnetic waves through a turbulent atmosphere,'' Rep. Prog. Phys. **55**, 39–112 (1992).
[28] D. K. Wilson, ''On the application of turbulence spectral/correlation models to sound propagation in the atmosphere,'' in Proceedings of the 8th International Symposium on Long Range Sound Propagation, edited by D. W. Thomson, K. E. Gilbert, and D. K. Wilson, 1998.
[29] J. Aitchison and J. A. C. Brown, *The Lognormal Distribution* (Cambridge University Press, London, 1973).

Sound propagation in a turbulent atmosphere near the ground: An approach based on the spectral representation of refractive-index fluctuations

Erik M. Salomons[a)]
TNO Institute of Applied Physics, P.O. Box 155, 2600 AD Delft, The Netherlands

Vladimir E. Ostashev
NOAA/Environmental Technology Laboratory, 325 Broadway, Boulder, Colorado 80305 and Physics Department, New Mexico State University, Las Cruces, New Mexico 88003

Steven F. Clifford and Richard J. Lataitis
NOAA/Environmental Technology Laboratory, 325 Broadway, Boulder, Colorado 80305

(Received 31 August 1999; revised 23 November 2000; accepted 18 January 2001)

A new, rigorous approach is presented for the computation of the fluctuating field of a monopole source in a nonrefracting, turbulent atmosphere above a ground surface. The time-averaged sound pressure level is considered, as well as statistical distributions of the sound pressure level. The computation is based on the Rytov solution of the wave equation for a turbulent medium, evaluated for the half-space above the ground surface. The solution takes into account the ground reflection of scattered waves, which has been neglected in previous work on this subject. The present approach is based on a Fourier–Stieltjes representation of refractive-index fluctuations, and makes use of a turbulent image atmosphere to account for the ground reflection of scattered waves. This approach is rigorous only for a rigid ground surface, but it is shown that it also yields a good approximation for a finite-impedance ground surface. The accuracy of the solution is demonstrated by comparison with results of numerical computations with the parabolic equation method for a turbulent atmosphere. The assumption of a nonrefracting atmosphere implies that direct application of the solution is limited to propagation over relatively small distances. However, this study can also be considered as a basis for a generalized solution for a downward refracting atmosphere, which can be applied for larger propagation distances.
[DOI: 10.1121/1.1356021]

PACS numbers: 43.28.Fp, 43.20.Bi, 43.20.Fn [LCS]

I. INTRODUCTION

A remarkable phenomenon in outdoor sound propagation near the ground is the occurrence of large fluctuations in the sound pressure field received from a source of constant strength. These fluctuations are induced by atmospheric turbulence. Experimental studies[1–4] indicate that the fluctuations are typically of the order of 10 dB, for propagation distances of about 100 m or more (see also Refs. 5–9).

Theoretical studies of wave propagation in unbounded turbulent media have been presented by Tatarskii,[10,11] Chernov,[12] Ishimaru[13] and Ostashev.[14] Ingard and Maling[1] have indicated that the fluctuations are considerably larger when the source or the receiver is near a ground surface. This is a consequence of the interference between the direct waves and the waves reflected by the ground. Small phase fluctuations in the direct and reflected waves may cause large fluctuations in the total field, in particular near interference minima.

Daigle *et al.*[15–17] have reported several experimental and theoretical studies of sound propagation through atmospheric turbulence near the ground. These studies have focused in particular on the time-averaged sound pressure level, which is the quantity commonly used for outdoor noise control. For the computation of this quantity, Daigle *et al.* presented a heuristic approach, which replaces a system with a ground surface by an approximately equivalent unbounded system. This allowed the application of results from the literature on wave propagation in unbounded turbulent media. The heuristic approach of Daigle *et al.* was generalized by L'Espérance *et al.*[18] to downward refracting, turbulent atmospheres.

Inspired by the work of Daigle *et al.*, Clifford and Lataitis[19] presented a more rigorous approach to the computation of the time-averaged sound pressure level, and obtained a result that is energy conserving. This approach was based on a rigorous evaluation of the Rytov integral along the direct ray and the reflected ray, and the solution contains no empirical parameters. Inspection of Eqs. (7) and (8) of Clifford and Lataitis, however, reveals that the ground reflection of scattered waves was neglected.[20] It will be shown below that this omission is equivalent to a truncation of the scattering volume, i.e., the integration volume of the Rytov solution. While this is a reasonable approximation for situations in which the source and/or the receiver are not very close to the ground surface, the question arises whether the scattering volume can be extended to include the ground reflection of scattered waves. In this article it will be shown that this is possible, at least for a rigid ground surface.

A turbulent atmosphere with both temperature and wind speed fluctuations will be assumed. The mathematical repre-

[a)]Electronic mail: salomons@tpd.tno.nl

sentation of such an atmosphere is based on the parabolic equation for a moving random medium.[14,21] It is assumed that the time-averaged atmosphere is nonrefracting, so that direct application of the results is limited to propagation over relatively small distances. However, the results can also be considered as a basis for a generalized approach for a downward refracting, turbulent atmosphere.[18]

The present article and the companion article[21] present two alternate approaches to the computation of the time-averaged sound pressure level, in the limit of weak fluctuations. Preliminary reports of the work described in the articles were presented in Refs. 20 and 22. The approach in the companion article is a generalization of the original approach of Clifford and Lataitis to moving random media with anisotropic turbulence. The approach in the present article is based on the Fourier–Stieltjes representation[10,11,13] of moving random media with anisotropic turbulence. This representation is based on a spectral decomposition of refractive-index fluctuations in the turbulent atmosphere. The Fourier–Stieltjes approach yields relatively simple mathematical expressions, which allows us to derive asymptotic approximations of the solution. Further, the Fourier–Stieltjes approach can be extended to the computation of statistical distributions of the sound pressure level, as will be shown in this article. A comparison will be made with the method developed by Hidaka *et al.*[23] for computing statistical distributions. A comparison will also be made with results of numerical simulations with the parabolic equation (PE) method,[24,25] taking into account the effect of the axisymmetric approximation in the PE method on the turbulent atmosphere.[26]

II. GEOMETRY AND ATMOSPHERE

A geometry is considered with a harmonic monopole source and a receiver in a turbulent atmosphere above a ground surface [see Fig. 1(a)]. It is assumed that the ground surface is rigid (the extension to a finite-impedance ground surface will be discussed in Sec. IV E). A rectangular *xyz* coordinate system is used, with the source at position $(0,0,z_s)$ and the receiver at position $(L,0,z_r)$. The ground surface is at $z=0$. We assume $L \gg z_s, z_r$.

The turbulent atmosphere is modeled as a random medium, with an effective sound speed[14] $c_{\text{eff}}(x,y,z) = c_0\sqrt{T(x,y,z)/T_0} + u(x,y,z)$. Here, the first term is the adiabatic sound speed in a medium with temperature distribution $T(x,y,z)$ (with c_0 the value of the adiabatic sound speed at the average temperature T_0), and the second term is the wind speed component u in the x direction. It is assumed that the time-averaged atmosphere is nonrefracting, so that the refractive index for a moving medium[14] $n = c_0/c_{\text{eff}}$ can be written as $n(x,y,z) = 1 + \mu(x,y,z)$, where the fluctuation μ has zero time average, $\langle \mu \rangle = 0$. Assuming that the temperature fluctuation $T_1 \equiv T - T_0$ and the wind speed component u are small, one finds the relation $\mu(x,y,z) = -\frac{1}{2}T_1(x,y,z)/T_0 - u(x,y,z)/c_0$. This relation also follows rigorously from the parabolic equation for sound propagation in a moving medium.[14,21] Averaging over time will be performed by ensemble averaging over different random realizations of the turbulent atmosphere. It is assumed that the turbulence is homogeneous.

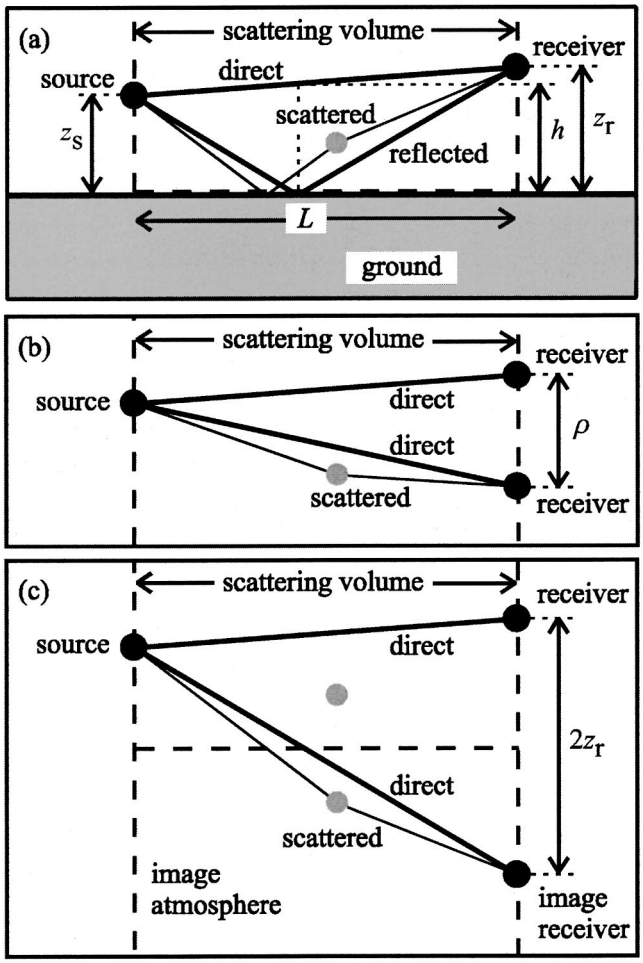


FIG. 1. Geometry of sound propagation from a source to a receiver in a turbulent atmosphere above a ground surface. (a) shows the physical system. (b) and (c) illustrate two methods for computing the fluctuating part of the sound field, the heuristic method A (b) and the rigorous method B (c). Thick solid lines represent the sound rays corresponding with the direct and reflected waves. Thin solid lines represent an example of a wave that reaches the receiver by single scattering from a turbulent inhomogeneity in the atmosphere. Indicated are horizontal range L, source height z_s, receiver height z_r, maximum vertical separation h between the two sound rays in the physical system shown in (a), vertical separation ρ between the two receivers in the unbounded system shown in (b), and vertical separation $2z_r$ between the two receivers in the unbounded system shown in (c).

III. GENERAL EXPRESSIONS

The fluctuating sound pressure field at the receiver is written as the sum of the direct field and the field reflected by the ground surface:

$$p = p_1 + p_2 \tag{1}$$

with

$$p_j = p_0 \frac{\exp(i[kr_j - \omega t] + \psi_j)}{r_j/r_0} \quad \text{for } j = 1,2, \tag{2}$$

where

$$\psi_j \equiv \chi_j + iS_j = \ln(p_j/p_{j,0}) \tag{3}$$

with χ_j the log-amplitude fluctuation and S_j the phase fluctuation, and $p_{j,0}$ the value of p_j in a nonturbulent atmosphere

$(\psi_j = 0)$. Here, p_0 is a constant acoustic pressure, ω is the angular frequency, t is the time, $k = \omega/c_0$ is the wave number, $r_0 = 1$ m, $r_1 = \sqrt{L^2 + (z_s - z_r)^2}$ and $r_2 = \sqrt{L^2 + (z_s + z_r)^2}$.

Following Clifford and Lataitis,[19] we assume $\tau_\omega \ll \tau_t$, where $\tau_\omega = 2\pi/\omega$ is the period of the harmonic waves and τ_t is a characteristic period of the turbulence. In other words, we assume that the relevant turbulent fluctuations are slow compared with the harmonic variations of the acoustic waves.

The actual sound pressure is equal to the real part of the complex sound pressure p given by Eq. (1), and is denoted as $p_r(t)$. The (fluctuating) short-time average of the squared sound pressure is $\overline{p_r^2(t)} = \frac{1}{2} pp^*$, an average over time τ_1, with $\tau_\omega \ll \tau_1 \ll \tau_t$. The corresponding (fluctuating) relative sound pressure level (i.e., sound pressure level relative to free field) is $\Delta L_p = 10 \lg (\overline{p_r^2(t)}/\overline{p_{r1,0}^2(t)})$, with $\overline{p_{r1,0}^2(t)} = \frac{1}{2} p_0^2 r_0^2 / r_1^2$. The long-time average is $\langle \overline{p_r^2(t)} \rangle = \langle \frac{1}{2} pp^* \rangle$, an average over time τ_2, with $\tau_2 \gg \tau_t$. The corresponding (time-averaged) relative sound pressure level is $\langle \Delta L_p \rangle = 10 \lg (\langle \overline{p_r^2(t)} \rangle / \langle \overline{p_{r1,0}^2(t)} \rangle)$.

From Eqs. (1)–(3) one finds the relation[1,23]

$$\Delta L_p = 10 \lg \left(\exp(2\chi_1) + \frac{r_1^2}{r_2^2} \exp(2\chi_2) + 2 \frac{r_1}{r_2} \times \exp(\chi_1 + \chi_2) \cos(kr_1 - kr_2 + \Delta S) \right) \tag{4}$$

with $\Delta S = S_1 - S_2$. This relation will be used in Sec. V for the computation of statistical distributions of the sound pressure level. For the generalization of Eq. (4) to a finite-impedance ground surface, see Sec. IV E. For the long-time average $\langle \Delta L_p \rangle$, we find from Eqs. (1)–(3)

$$\langle \Delta L_p \rangle = 10 \lg \left\{ \langle \exp(2\chi_1) \rangle + \frac{r_1^2}{r_2^2} \langle \exp(2\chi_2) \rangle + \frac{r_1}{r_2} [\exp(i[kr_1 - kr_2]) \langle \exp(\psi_1 + \psi_2^*) \rangle + \exp(-i[kr_1 - kr_2]) \langle \exp(\psi_1^* + \psi_2) \rangle] \right\}. \tag{5}$$

To evaluate this expression, we first note that conservation of energy implies $\langle \exp(2\chi_j) \rangle = 1$. Using the relation[11,19] $\langle \exp(2\chi_j) \rangle = \exp(2\sigma_{\chi_j}^2 + 2\langle \chi_j \rangle)$, with $\sigma_{\chi_j}^2 = \langle (\chi_j - \langle \chi_j \rangle)^2 \rangle$, we find the relation $\langle \chi_j \rangle = -\sigma_{\chi_j}^2$. Here, we have used the fact that a Gaussian distribution can be assumed[13,27] for χ_j; the same applies to S_j. A more complex derivation yields the relation $\langle S_j \rangle = -\langle \chi_j S_j \rangle$ [see Eqs. (B7) and (B8) of Ref. 21]. Terms of the order $\langle \mu^4 \rangle$ and higher have been neglected here. We find

$$\langle \Delta L_p \rangle = 10 \lg \left(1 + \frac{r_1^2}{r_2^2} + 2 \frac{r_1}{r_2} \cos(kr_1 - kr_2 + \langle \chi_2 S_1 \rangle - \langle \chi_1 S_2 \rangle) \Gamma \right) \tag{6}$$

with coherence factor Γ given by

$$\Gamma = \exp(\langle \chi_1 \chi_2 \rangle - \tfrac{1}{2}\langle \chi_1^2 \rangle - \tfrac{1}{2}\langle \chi_2^2 \rangle + \langle S_1 S_2 \rangle - \tfrac{1}{2}\langle S_1^2 \rangle - \tfrac{1}{2}\langle S_2^2 \rangle). \tag{7}$$

Again, terms of the order $\langle \mu^4 \rangle$ and higher have been neglected here. The terms $\langle \chi_j S_k \rangle$ in the argument of the cosine function in Eq. (6) are small for weak fluctuations[19] (see Sec. IV F) and will be neglected unless indicated otherwise. The generalization of Eq. (6) for a finite-impedance ground surface is given in Refs. 16, 19 and 21 (see also Sec. IV E). For a rigid ground surface, Eqs. (6) and (7) are equivalent to Eqs. (10) and (11) of the companion paper.[21]

The computation of $\langle \Delta L_p \rangle$ has been reduced to the computation of the covariances $\langle \chi_j S_k \rangle$, $\langle \chi_j \chi_k \rangle$ and $\langle S_j S_k \rangle$ $(j,k = 1,2)$. We will consider two different methods for the computation, a heuristic and a rigorous one. These methods will be referred to as methods A and B, respectively. Both methods are based on the Rytov approximation and therefore limited to weak fluctuations. The two methods are illustrated in Fig. 1. In both methods, the physical system with the ground surface and a single receiver is replaced by an unbounded system without a ground surface and two receivers. The direct and reflected rays in the physical system are replaced by two direct rays to the receivers in the unbounded system. Below, methods A and B are explained in more detail. The rigorous method B is explained first, as this makes the explanation of the heuristic method A easier. Method A is similar to the method of Daigle *et al.*[15–18]

In the rigorous method B, the unbounded system consists of the original physical system above the ground surface and its image below the ground surface. The image atmosphere below the ground surface is an exact mirror image of the atmosphere above the ground surface (here we assume a rigid ground surface; we will come back to this assumption in Sec. IV). The image receiver is at height $z = -z_r$, so the vertical separation between the two receivers is $2z_r$. The sound pressure at the receiver in the physical system [Fig. 1(a)] is exactly equal to the sum of the sound pressures at the two receivers in the unbounded system [Fig. 1(c)]. This follows from the symmetry, but can also be shown mathematically using the integral representation of the Helmholtz equation for a random medium (see Appendix A of Ref. 21). As an example, Fig. 1(a) shows the path of a wave that is first reflected by the ground and next scattered by a turbulent inhomogeneity. Figure 1(c) shows the corresponding path along the image inhomogeneity in the image atmosphere. The symmetry implies that the contributions of the scattered waves to the sound pressure are equal in Figs. 1(a) and (c). As a consequence, the covariances in Eqs. (6) and (7) can be computed for the system in Fig. 1(c). This will be done in Sec. IV.

The heuristic method A ignores the symmetry of the atmosphere in the unbounded system. As a consequence, we can apply the theory of wave propagation in unbounded turbulent media. In this theory, the covariances $\langle \chi_j \chi_k \rangle$ and $\langle S_j S_k \rangle$ are usually denoted as correlation functions $B_\chi(\rho)$ and $B_S(\rho)$, respectively, where ρ is the separation between the two receivers in the unbounded system. We have $\langle \chi_1^2 \rangle \approx \langle \chi_2^2 \rangle = B_\chi(0)$ and $\langle S_1^2 \rangle \approx \langle S_2^2 \rangle = B_S(0)$. The coherence factor (7) becomes in this case

$$\Gamma(\rho)=\exp[B_\chi(\rho)-B_\chi(0)+B_S(\rho)-B_S(0)]. \tag{8}$$

Expressions for the functions $B_\chi(\rho)$ and $B_S(\rho)$ are given in Sec. IV. For the argument ρ we use the maximum vertical separation $h=2z_sz_r/(z_s+z_r)$ between the two sound rays in the physical system, the direct ray and the ray reflected by the ground surface [see Fig. 1(a)]. The *mean* vertical separation between the two rays is then equal to $\frac{1}{2}h$, both in the physical system and in the unbounded system. It should be noted that Daigle *et al.*[15] found best agreement with experimental data by setting $\rho=\frac{1}{2}h$ instead of $\rho=h$. It should be noted further that the sound pressure at the receiver in the physical system [Fig. 1(a)] is not equal to the sum of the sound pressures at the two receivers in the system in Fig. 1(b). The heuristic method A only approximates the coherence factor in Eq. (6) by the expression in Eq. (8). For the other quantities in Eq. (6) (r_1 and r_2), the physical system in Fig. 1(a) is used.

In both methods, the scattering volume is limited to the region between the source and the receiver (see Fig. 1). This means that backscattering is neglected.

It follows from Eqs. (1), (2) and (6) that the coherence factor can also be defined as

$$\Gamma=\frac{\langle p_1p_2^*\rangle+\langle p_1^*p_2\rangle}{\langle p_1p_2^*\rangle_0+\langle p_1^*p_2\rangle_0}, \tag{9}$$

where index 0 is used for a nonturbulent atmosphere. We have identified the pressures p_1 and p_2 as the pressures at the two receivers in an unbounded system [Figs. 1(b) and (c)]. Therefore, the quantity $\langle p_1p_2^*\rangle$ can be identified as the mutual coherence function for an unbounded medium.[11,13] Thus, the coherence factor Γ is equal to the real part of the mutual coherence function, normalized to unity for a nonturbulent atmosphere.

IV. TIME-AVERAGED SOUND PRESSURE LEVEL

In this section, the heuristic method A and the rigorous method B for computing the time-averaged relative sound pressure level $\langle\Delta L_p\rangle$ are described. First, a distinction is made between isotropic turbulence and axisymmetric turbulence.

A. Isotropic turbulence and axisymmetric turbulence

The turbulent atmosphere is represented by the refractive-index fluctuation $\mu(\vec{r})$. It is assumed that the turbulence is homogeneous,[13] so that the correlation function of $\mu(\vec{r})$ can be written as $B_\mu(\vec{r})\equiv\langle\mu(\vec{r}_1)\mu(\vec{r}_2)\rangle$ with $\vec{r}=\vec{r}_1-\vec{r}_2$. The spectral density $\Phi_\mu(\vec{K})$ of $\mu(\vec{r})$ is defined by

$$\Phi_\mu(\vec{K})=(2\pi)^{-3}\int B_\mu(\vec{r})e^{-i\vec{K}\cdot\vec{r}}\,d\vec{r}. \tag{10}$$

For homogeneous and isotropic turbulence, we have $B_\mu(\vec{r})=B_\mu(r)$, with $r\equiv|\vec{r}|=\sqrt{x^2+y^2+z^2}$, where $\vec{r}=(x,y,z)$. This implies $\Phi_\mu(\vec{K})=\Phi_\mu(K)$, with $K\equiv|\vec{K}|$. Homogeneous and isotropic turbulence will be referred to simply as isotropic turbulence.

In computational methods for sound propagation in a turbulent atmosphere, such as the parabolic equation (PE) method,[24,25] one nearly always assumes that the system has axial symmetry around the vertical axis through the source. The computation of sound propagation can then be performed in two dimensions. The assumption of axial symmetry, however, also affects the turbulent atmosphere.[26] For the purpose of comparisons between analytical results and PE results, it is therefore of interest to have an expression for the spectral density of an axisymmetric turbulent atmosphere. To derive this expression, we represent an axisymmetric turbulent atmosphere by a field $\mu(\vec{r})$ which is independent of y and homogeneous and isotropic in the xz plane, so that $B_\mu(\vec{r})=B_\mu(\sqrt{x^2+z^2})$. In this case Eq. (10) gives, with the notation $\vec{K}=(\kappa_x,\kappa_y,\kappa_z)$,

$$\Phi_\mu(\vec{K})=F_\mu(\sqrt{\kappa_x^2+\kappa_z^2})\,\delta(\kappa_y), \tag{11}$$

where $\delta(\kappa_y)$ is the Dirac delta function and

$$F_\mu(\sqrt{\kappa_x^2+\kappa_z^2})=(2\pi)^{-2}\int\int B_\mu(\sqrt{x^2+z^2})\times\exp(-i[\kappa_xx+\kappa_zz])\,dx\,dz \tag{12}$$

is a function closely related to the two-dimensional spectral density of $\mu(\vec{r})$.[13] Equation (11) will be used for comparisons with PE results.

B. Computational method A

The heuristic method A for computing $\langle\Delta L_p\rangle$ is based on Eq. (6), with the coherence factor $\Gamma(\rho)$ given by Eq. (8) and $\rho=h$ (see Sec. III). In this section, expressions are given for the correlation functions B_χ and B_S in Eq. (8) and for the coherence factor Γ, first for homogeneous turbulence, next for isotropic turbulence and for axisymmetric turbulence. The expressions for isotropic turbulence are known from the literature, but are included here for comparison with the expressions for axisymmetric turbulence.

Using the notation $B_1\equiv B_\chi$ and $B_2\equiv B_S$, the Rytov method yields the following general expression for B_j ($j=1,2$) for homogeneous turbulence,[13]

$$B_j(\vec{\rho})=2\pi L\int_0^1 d\tau\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp(i\vec{\kappa}\cdot\vec{\rho}\tau)\times H_j^2[L\tau(1-\tau),\kappa]\Phi_\mu(\vec{\kappa})\,d\kappa_y\,d\kappa_z \tag{13}$$

with $\vec{\rho}$ and $\vec{\kappa}=(\kappa_y,\kappa_z)$ two-dimensional vectors (indicated by the arrows over the symbols), $\kappa=\sqrt{\kappa_y^2+\kappa_z^2}$, $H_1(x,\kappa)=k\sin(\frac{1}{2}x\kappa^2/k)$, $H_2(x,\kappa)=k\cos(\frac{1}{2}x\kappa^2/k)$, and $\Phi_\mu(\vec{\kappa})$ the spectral density $\Phi_\mu(\vec{K})$ defined by Eq. (10) and evaluated for $\vec{K}=(0,\kappa_y,\kappa_z)$. We note that the notation in Eq. (8) assumes isotropic or axisymmetric turbulence, but this equation can be generalized to homogeneous turbulence simply by replacing ρ by $\vec{\rho}$ and using $\vec{\rho}=h\vec{e}_z$ with $\vec{e}_z$ a unit vector in the z direction. Using Eq. (13), we find for homogeneous turbulence

$$\Gamma = \exp\left(-2\pi L k^2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \Phi_\mu(\vec{\kappa}) \mathrm{f}(\kappa_z)\, d\kappa_y\, d\kappa_z \right) \quad (14)$$

with

$$\mathrm{f}(\kappa_z) = 1 - \frac{\sin(\kappa_z \rho)}{\kappa_z \rho}, \quad (15)$$

where $\rho \equiv |\vec{\rho}|$.

For the case of isotropic turbulence, a transformation to polar coordinates κ, θ_κ is applied in Eq. (13). This gives[14,28]

$$B_j(\rho) = 4\pi^2 L \int_0^1 d\tau \int_0^\infty J_0(\kappa\rho\tau) \times H_j^2[L\tau(1-\tau),\kappa]\Phi_\mu(\kappa)\kappa\, d\kappa \quad (16)$$

with Bessel function J_0. The coherence factor given by Eq. (8) becomes[14,28]

$$\Gamma(\rho) = \exp\left(-4\pi^2 L k^2 \times \int_0^1 d\tau \int_0^\infty [1 - J_0(\kappa\rho\tau)]\Phi_\mu(\kappa)\kappa\, d\kappa \right). \quad (17)$$

As an example, we consider a Gaussian correlation function of the refractive-index fluctuations

$$B_\mu(r) = \mu_0^2 \exp\left(-\frac{r^2}{a^2} \right) \quad (18)$$

with μ_0 the standard deviation of μ and a the correlation length. This expression is a simple approximation for an atmosphere with temperature fluctuations only[14] (with $\mu_0^2 = \langle T_1^2 \rangle / 4T_0^2$). In this case, one finds the following expressions for the correlation functions of the log-amplitude ($j=1$) and phase ($j=2$) fluctuations[14,20,28]

$$B_j(\rho) = \frac{1}{4}\pi^{1/2} L k^2 \mu_0^2 a^3 \int_0^1 \left[F\left(\rho\tau, 0, \frac{1}{4}a^2 \right) + (-1)^j F\left[\rho\tau, L\tau(1-\tau)/k, \frac{1}{4}a^2 \right] \right] d\tau \quad (19)$$

with the function $F(\alpha,\beta,\gamma)$ given by[20]

$$F(\alpha,\beta,\gamma) \equiv \int_0^\infty J_0(\alpha\kappa)\cos(\beta\kappa^2)\exp(-\gamma\kappa^2)\kappa\, d\kappa = \frac{1}{2A}\exp\left(-\frac{\alpha^2\gamma}{4A} \right)\left(\gamma\cos\frac{\alpha^2\beta}{4A} + \beta\sin\frac{\alpha^2\beta}{4A} \right) \quad (20)$$

with the notation $A \equiv \gamma^2 + \beta^2$. Using the relation $F(\rho\tau,0,\frac{1}{4}a^2) - F(0,0,\frac{1}{4}a^2) = 2a^{-2}[\exp(-\rho^2\tau^2/a^2) - 1]$, the following expression is obtained for the coherence factor:[14,28]

$$\Gamma(\rho) = \exp\left(-\pi^{1/2} L k^2 \mu_0^2 a \left[1 - \tfrac{1}{2}\,\pi^{1/2}\,\frac{\mathrm{erf}(\rho/a)}{(\rho/a)} \right] \right), \quad (21)$$

where $\mathrm{erf}(x) = 2\pi^{-1/2}\int_0^x \exp(-\tau^2)\, d\tau$ is the error function.

For the case of axisymmetric turbulence, we use Eq. (11) in Eq. (13) and find (denoting κ_z as κ)

$$B_j(\rho) = 4\pi L \int_0^1 d\tau \int_0^\infty \cos(\kappa\rho\tau) \times H_j^2[L\tau(1-\tau),\kappa]F_\mu(\kappa)\, d\kappa. \quad (22)$$

The coherence factor given by Eq. (8) becomes

$$\Gamma(\rho) = \exp\left(-4\pi L k^2 \times \int_0^1 d\tau \int_0^\infty [1 - \cos(\kappa\rho\tau)]F_\mu(\kappa)\, d\kappa \right). \quad (23)$$

For the Gaussian correlation function (18) we find

$$B_j(\rho) = \frac{1}{2} L k^2 \mu_0^2 a^2 \int_0^1 \left[I\left(\rho\tau, 0, \frac{1}{4}a^2 \right) + (-1)^j I\left[\rho\tau, L\tau(1-\tau)/k, \frac{1}{4}a^2 \right] \right] d\tau \quad (24)$$

with the function $I(\alpha,\beta,\gamma)$ given by[20]

$$I(\alpha,\beta,\gamma) = \int_0^\infty \cos(\alpha\kappa)\cos(\beta\kappa^2)\exp(-\gamma\kappa^2)\, d\kappa = \frac{\sqrt{\pi}}{2^{3/2}A^{1/4}}\exp\left(-\frac{\alpha^2\gamma}{4A} \right) \times \left(\sqrt{1 + \frac{\gamma}{\sqrt{A}}}\cos\frac{\alpha^2\beta}{4A} + \sqrt{1 - \frac{\gamma}{\sqrt{A}}}\sin\frac{\alpha^2\beta}{4A} \right). \quad (25)$$

Using the relation $I(\rho\tau,0,\frac{1}{4}a^2) - I(0,0,\frac{1}{4}a^2) = \pi^{1/2}a^{-1} \times [\exp(-\rho^2\tau^2/a^2) - 1]$, the expression (21) is obtained for the coherence factor. Thus, the coherence factor for axisymmetric turbulence is identical to the coherence factor for isotropic turbulence, in the case of the Gaussian correlation function (18).

C. Computational method B

The rigorous method B for computing $\langle \Delta L_p \rangle$ is based on the Rytov perturbation method[13] applied to the system shown in Fig. 1(c), with an atmosphere that is symmetric with respect to $z=0$. The fluctuating variables ψ_1 and ψ_2 in Eq. (2) can be computed with the Rytov solution evaluated at the receiver position $\vec{r} = (L,0,z_r)$ and the image receiver position $\vec{r}_i = (L,0,-z_r)$, respectively.

We first give the general Rytov solution for an unbounded system with an arbitrary turbulent atmosphere, a monopole source at position $\vec{r}_s = (0,0,z_s)$ and a receiver at position $\vec{r} = (L,y,z)$. The turbulent atmosphere is represented by the random field $\mu(x,y,z)$. The corresponding fluctuating sound pressure field p is represented by the quantity $\psi = \ln p/p_0$, where p_0 is the undisturbed sound pressure. Application of the Rytov perturbation method to the Helmholtz equation for a nonmoving random medium,[13] or to the

parabolic equation for a moving random medium,[14,21,22] yields the following first-order solution for ψ:

$$\psi(\vec{r})=\int_{V'} h(\vec{r},\vec{r}')\mu(\vec{r}')\,d\vec{r}', \tag{26}$$

where the integral is over the scattering volume V' between the source and the receiver, i.e., the volume of all points $\vec{r}'=(x',y',z')$ with $0\leqslant x'\leqslant L$ (backscattering is neglected). The function $h(\vec{r},\vec{r}')$ is given by

$$h(\vec{r},\vec{r}')=\frac{k^2}{2\pi(x'/L)(L-x')}\times\exp\left[i\frac{k}{2}\frac{|(\vec{\rho}'-\vec{\rho}_s)-(x'/L)(\vec{\rho}-\vec{\rho}_s)|^2}{(x'/L)(L-x')}\right] \tag{27}$$

with $\vec{\rho}=(y,z)$, $\vec{\rho}'=(y',z')$ and $\vec{\rho}_s=(0,z_s)$. The fluctuating field $\mu(\vec{r}')$ is represented by the following stochastic Fourier–Stieltjes integral,[10,11,13]

$$\mu(\vec{r}')=\int \exp(i\vec{\kappa}\cdot\vec{\rho}')\,d\nu(x',\vec{\kappa}) \tag{28}$$

with $\vec{\kappa}=(\kappa_y,\kappa_z)$ and random amplitudes $d\nu$ that satisfy the relation

$$\langle d\nu(x',\vec{\kappa})d\nu^*(x'',\vec{\kappa}')\rangle=F_{\mu,2}(x'-x'',\vec{\kappa})\,\delta(\vec{\kappa}-\vec{\kappa}')d\vec{\kappa}d\vec{\kappa}', \tag{29}$$

where $F_{\mu,2}(x,\vec{\kappa})$ is the two-dimensional spectral density of $\mu(\vec{r})$. Equation (28) reduces to an ordinary Fourier integral if one introduces[29] an amplitude density $A_\mu(x',\vec{\kappa})$ with $d\nu(x',\vec{\kappa})=A_\mu(x',\vec{\kappa})d\vec{\kappa}$. In the derivation below we use the Fourier–Stieltjes representation (28), however, in order to follow the corresponding derivation in Ref. 13 for an unbounded system without the symmetry in the atmosphere as in Fig. 1(c). In this way, we limit the length of the derivation by referring to Ref. 13 for identical steps in the two derivations.

In the system shown in Fig. 1(c), the atmosphere for $z<0$ is a mirror image of the atmosphere for $z>0$, so we use the field $\mu(x',y',|z'|)$ instead of $\mu(x',y',z')$ in Eq. (26). Using Eq. (28), Eq. (26) becomes

$$\psi(L,y,z)=\int_0^L\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}dx'\,dy'\,dz'\int \exp(i\kappa_y y'+i\kappa_z|z'|)h(\vec{r},\vec{r}')\,d\nu(x',\kappa_y,\kappa_z) \tag{30}$$

with

$$h(\vec{r},\vec{r}')=\frac{k^2}{2\pi(x'/L)(L-x')}\times\exp\left[i\frac{k}{2}\frac{[y'-(x'/L)y]^2}{(x'/L)(L-x')}\right]\times\exp\left[i\frac{k}{2}\frac{[z'-z_1]^2}{(x'/L)(L-x')}\right], \tag{31}$$

where $z_1\equiv(x'/L)z+(1-x'/L)z_s$. Equations (30) and (31) are equivalent to Eqs. (13)–(15) in the companion paper.[21] The form of the equations is different, however, because we have used the Fourier–Stieltjes representation (28) here. The integrals over y' and z' in Eq. (30) can be evaluated (see the Appendix). This gives

$$\psi(L,y,z)=\int_0^L dx'\int E(x',\kappa_y,\kappa_z,y,z)\,d\nu(x',\kappa_y,\kappa_z) \tag{32}$$

where

$$E(x',\kappa_y,\kappa_z,y,z)=ik\exp\left(-i\frac{(x'/L)(L-x')}{2k}\kappa^2\right)\exp(i\kappa_y y x'/L)\times\{\exp(i\kappa_z z_1)\tfrac{1}{2}[1-(1-i)F_F(\nu_1)]+\exp(-i\kappa_z z_1)\tfrac{1}{2}[1-(1-i)F_F(\nu_2)]\} \tag{33}$$

with $\nu_1=-z_1\sqrt{\alpha}+\frac{1}{2}\kappa_z/\sqrt{\alpha}$, $\nu_2=z_1\sqrt{\alpha}+\frac{1}{2}\kappa_z/\sqrt{\alpha}$, where $\alpha=k/[2(x'/L)(L-x')]$, and $F_F(x)=C(x\sqrt{2/\pi})+iS(x\sqrt{2/\pi})$, with C and S the Fresnel integrals (see the Appendix). The log-amplitude fluctuation χ and the phase fluctuation S are obtained from ψ by the relations $\chi=(\psi+\psi^*)/2$ and $S=(\psi-\psi^*)/(2i)$. Since μ is real, we have $d\nu^*(x',-\vec{\kappa})=d\nu(x',\vec{\kappa})$, so that Eq. (32) gives (with the substitution $\vec{\kappa}\rightarrow-\vec{\kappa}$)

$$\psi^*(L,y,z)=\int_0^L dx'\int E^*(x',-\kappa_y,-\kappa_z,y,z)\times d\nu(x',\kappa_y,\kappa_z). \tag{34}$$

We obtain

$$\chi(L,y,z)=\int_0^L dx'\int E_1(x',\kappa_y,\kappa_z,y,z)\,d\nu(x',\kappa_y,\kappa_z), \tag{35}$$

$$S(L,y,z)=\int_0^L dx'\int E_2(x',\kappa_y,\kappa_z,y,z)\,d\nu(x',\kappa_y,\kappa_z), \tag{36}$$

with

$$E_1(x',\kappa_y,\kappa_z,y,z)=\tfrac{1}{2}[E(x',\kappa_y,\kappa_z,y,z)+E^*(x',-\kappa_y,-\kappa_z,y,z)], \tag{37}$$

$$E_2(x',\kappa_y,\kappa_z,y,z)=\frac{1}{2i}[E(x',\kappa_y,\kappa_z,y,z)-E^*(x',-\kappa_y,-\kappa_z,y,z)]. \tag{38}$$

This gives, using Eq. (29) [for details, see the corresponding derivation in Ref. 13 for an unbounded system without the symmetry in the atmosphere as in Fig. 1(c)],

$$\langle\chi(L,y,z_{r1})\chi(L,y,z_{r2})\rangle=2\pi\int_0^L d\eta\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}E_1(\eta,\kappa_y,\kappa_z,y,z_{r1})\times E_1^*(\eta,\kappa_y,\kappa_z,y,z_{r2})\Phi_\mu(\vec{\kappa})\,d\kappa_y\,d\kappa_z, \tag{39}$$

$$\langle S(L,y,z_{r1})S(L,y,z_{r2})\rangle = 2\pi\int_0^L d\eta\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E_2(\eta,\kappa_y,\kappa_z,y,z_{r1}) \times E_2^*(\eta,\kappa_y,\kappa_z,y,z_{r2})\Phi_\mu(\vec{\kappa})\,d\kappa_y\,d\kappa_z, \quad (40)$$

$$\langle \chi(L,y,z_{r1})S(L,y,z_{r2})\rangle = 2\pi\int_0^L d\eta\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E_1(\eta,\kappa_y,\kappa_z,y,z_{r1}) \times E_2^*(\eta,\kappa_y,\kappa_z,y,z_{r2})\Phi_\mu(\vec{\kappa})\,d\kappa_y\,d\kappa_z, \quad (41)$$

where $\Phi_\mu(\vec{\kappa})$ is the spectral density $\Phi_\mu(\vec{K})$ defined by Eq. (10) and evaluated for $\vec{K}=(0,\kappa_y,\kappa_z)$. For convenience, the variable η running from 0 to L may be replaced by the dimensionless variable $\tau=\eta/L$ running from 0 to 1. With the expressions (39)–(41), the covariances of log-amplitude fluctuations χ_j and phase fluctuations S_j in Eqs. (6) and (7) can be computed:

$\langle\chi_1\chi_1\rangle$ and $\langle S_1S_1\rangle$ are obtained by setting $z_{r1}=z_r$ and $z_{r2}=z_r$,

$\langle\chi_2\chi_2\rangle$ and $\langle S_2S_2\rangle$ are obtained by setting $z_{r1}=-z_r$ and $z_{r2}=-z_r$,

$\langle\chi_1\chi_2\rangle$ and $\langle S_1S_2\rangle$ are obtained by setting $z_{r1}=\pm z_r$ and $z_{r2}=\mp z_r$,

$\langle\chi_1S_2\rangle$ is obtained by setting $z_{r1}=z_r$ and $z_{r2}=-z_r$,

$\langle\chi_2S_1\rangle$ is obtained by setting $z_{r1}=-z_r$ and $z_{r2}=z_r$.

Here, z_r is the height of the receiver and $-z_r$ is the height of the image receiver. With these expressions for the covariances, the time-averaged relative sound pressure level $\langle\Delta L_p\rangle$ given by Eq. (6) can be computed.

For isotropic turbulence, we have $\Phi_\mu(\vec{\kappa})=\Phi_\mu(\kappa)$ in Eqs. (39)–(41). For axisymmetric turbulence, we have from Eq. (11) the relation $\Phi_\mu(\vec{\kappa})=F_\mu(\kappa_z)\delta(\kappa_y)$, and after substitution in Eqs. (39)–(41) the integral over κ_y can be performed analytically. As a consequence, the evaluation of Eqs. (39)–(41) requires the numerical computation of triple integrals for isotropic turbulence, and double integrals for axisymmetric turbulence. With current computers this numerical problem is easily solved.

The rigorous computational method previously presented is an improvement with respect to the approach of Clifford and Lataitis.[19] Their approach neglects the ground reflection of scattered waves, or equivalently, the scattering from the image atmosphere below the ground surface.

D. Asymptotic approximations

We introduce the dimensionless parameters $D\equiv L/a^2k$ and $\zeta\equiv(z_r+z_s)/a$ (the parameter D is called the wave parameter). For the correlation length a near the ground we have $a\sim 1$ m,[15,17] so that $\zeta\sim 1$ if $(z_r+z_s)\sim 1$ m. We consider two limiting cases: (1) $D\ll\zeta$ and (2) $D\gg\zeta$.

In the first case, we have $\nu_2\approx-\nu_1\approx(z_1/a)/\sqrt{2(\tau-\tau^2)D}$, and we find

$$E_1=k\sin(\tfrac{1}{2}\kappa^2a^2[\tau-\tau^2]D)\exp(i\kappa_y y\tau)\exp(i\kappa_z|z_1|),$$
$$E_2=k\cos(\tfrac{1}{2}\kappa^2a^2[\tau-\tau^2]D)\exp(i\kappa_y y\tau)\exp(i\kappa_z|z_1|). \quad (42)$$

Substitution of these expressions into Eqs. (39) and (40) yields $\langle\chi_1\chi_2\rangle=B_\chi(\rho)$, $\langle S_1S_2\rangle=B_S(\rho)$, $\langle\chi_j^2\rangle=B_\chi(0)$ and $\langle S_j^2\rangle=B_S(0)$, with B_χ and B_S given by Eq. (13) and $\rho=h$, where $h=2z_rz_s/(z_r+z_s)$. For the coherence factor we find Eq. (14) with function $f(\kappa_z)$ given by Eq. (15). This means that methods A and B yield equal values for the coherence factor Γ, in the limit $D\ll\zeta$. We note that Eqs. (14) and (15) are equivalent to Eq. (32) of the companion paper.[21]

In the second case, $D\gg\zeta$, we can set $\nu_1\approx\nu_2\approx\frac{1}{2}\kappa_z a\sqrt{2(\tau-\tau^2)D}$. In this case, we find for the variances of phase and log-amplitude fluctuations:

$$\langle S_j^2\rangle=\langle\chi_j^2\rangle=2\pi Lk^2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\Phi_\mu(\vec{\kappa}) \times\left[\frac{1}{2}+\frac{1}{2}\,\frac{\sin(\pm 2\kappa_z z_r)-\sin(2\kappa_z z_s)}{2\kappa_z(\pm z_r-z_s)}\right]d\kappa_y\,d\kappa_z, \quad (43)$$

where the plus sign is for $j=1$ and the minus sign is for $j=2$, and for the covariances:

$$\langle S_1S_2\rangle=\langle\chi_1\chi_2\rangle=2\pi Lk^2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\Phi_\mu(\vec{\kappa}) \times\frac{1}{2}\left[\frac{\sin(2\kappa_z z_r)}{2\kappa_z z_r}+\frac{\sin(2\kappa_z z_s)}{2\kappa_z z_s}\right]d\kappa_y\,d\kappa_z. \quad (44)$$

For the coherence factor we find the same expression (14) as in the first case, but now with a different function $f(\kappa_z)$:

$$f(\kappa_z)=1-\frac{\sin(2\kappa_z z_r)}{2\kappa_z z_r}-\frac{\sin(2\kappa_z z_s)}{2\kappa_z z_s} +\frac{1}{2}\,\frac{\sin(2\kappa_z z_r)-\sin(2\kappa_z z_s)}{2\kappa_z z_r-2\kappa_z z_s} +\frac{1}{2}\,\frac{\sin(2\kappa_z z_r)+\sin(2\kappa_z z_s)}{2\kappa_z z_r+2\kappa_z z_s}. \quad (45)$$

For $z_r\ll a$ and $z_s\ll a$, the sine functions can be expanded, and we find $f(\kappa_z)\approx\frac{2}{15}\kappa_z^4z_r^2z_s^2$.

In this case, methods A and B yield different values of Γ, in general. The differences prove to be small, however. This is explained in the following. In the limiting cases $z_r\gg z_s$ and $z_r\ll z_s$, methods A and B yield equal values of Γ [Eqs. (15) and (45) coincide in these cases]. For the case $z_s\sim z_r$, we distinguish three cases: (*i*) $\zeta\ll 1$ (or $z_s\sim z_r\ll a$), (*ii*) $\zeta\gg 1$ (or $z_s\sim z_r\gg a$) and (*iii*) $\zeta\sim 1$ (or $z_s\sim z_r\sim a$). For $\zeta\ll 1$, both methods yield a value of Γ close to unity, as can be shown by expanding the sine functions in Eqs. (15) and (45), and making use of the restriction to weak fluctuations. For $\zeta\gg 1$, the functions $f(\kappa_z)$ given by Eqs. (15) and (45) oscillate rapidly around the value of unity, except in a narrow region near $\kappa_z=0$; in this case, we can set $f(\kappa_z)=1$ in Eq. (14), both for method A and for method B. For $\zeta\sim 1$, we have compared numerical values of Γ computed with methods A and B for several cases, and found negligible differences.

We conclude that methods A and B yield approximately equal values of Γ for $D\ll\zeta$ and $D\gg\zeta$. From numerical com-

putations we concluded that this is true also in the case $D \sim \zeta$. Therefore, methods A and B yield approximately equal values of Γ in all cases. We note, however, that for $D \gg 1$, $z_r/a \to 0$ and $z_s/a \to 0$, the variances predicted by method A [Eq. (13)] are a factor of 2 smaller than the variances predicted by method B [Eq. (43)], since in this case we have $\langle \chi_j^2 \rangle = \langle S_j^2 \rangle = 2B_\chi(0) = 2B_S(0)$. In this limiting case, the variances are a measure of the fluctuations of the total sound pressure (see Fig. 4 in Sec. IV F).

Finally, it is interesting to make a comparison with the asymptotic approximations presented by Clifford and Lataitis.[19] From Eq. (11) of Clifford and Lataitis, it follows that the coherence factor is given by expression (14) in the present article, with the following function $f(\kappa_z)$:

$$f(\kappa_z) = \alpha_{CL}\left[1 - \frac{\sin(\kappa_z \rho)}{\kappa_z \rho}\right] \tag{46}$$

with $\rho = h$, where $h = 2z_r z_s/(z_r + z_s)$, and $\alpha_{CL} = 1$ for $D \ll 1$ and $\alpha_{CL} = \frac{1}{2}$ for $D \gg 1$ ($\zeta \sim 1$ was assumed). In the first case, the result of Clifford and Lataitis is identical with the results of both methods A and B. In this case, the omission of the ground reflection of scattered waves by Clifford and Lataitis has no effect on the coherence factor. In the second case, the result of Clifford and Lataitis is different from the results of both methods A and B. To verify our solution with method B, we repeated the derivation with method B, this time ignoring the ground reflection of scattered waves; the result was exactly the result of Clifford and Lataitis, for $D \gg 1$. This proves that the two approaches yield the same result, provided equal scattering volumes are used (including image scattering volumes). We conclude that one should use $\alpha_{CL} = 1$ in Eq. (46) for all values of the wave parameter D (this is in fact method A), instead of $\alpha_{CL} = 1$ for $D \ll 1$ and $\alpha_{CL} = \frac{1}{2}$ for $D \gg 1$.

E. Generalization to a finite-impedance ground surface

The foregoing was limited to a rigid ground surface. A finite ground impedance can be taken into account by including a spherical-wave reflection coefficient $Q \equiv |Q| \exp(i\theta)$ as a factor in the pressure p_2, i.e., on the right-hand side of Eq. (2) for $j = 2$. As a result, the second and third terms in the argument of the logarithmic function in Eqs. (4) and (6) acquire factors $|Q|^2$ and $|Q|$, respectively, and the argument of the cosine function acquires a term $-\theta$.[16,19,21,23] In addition, the ground impedance might have an effect on the coherence factor Γ in Eq. (6). From numerical computations we conclude, however, that this effect is small (see, e.g., Fig. 3 in Sec. IV F), and that the rigid-ground approaches to the computation of Γ, which were presented in the previous sections, provide a good approximation also for a finite-impedance ground (the approximation is expected to be more accurate for $D \ll 1$ than for $D \gg 1$; see later in this section). This can be explained as follows. First, we note that the direct and reflected waves in Fig. 1(a) are perturbed predominantly by scattered waves that propagate within a narrow region around the unperturbed paths. In other words, only scattered waves with small scattering angles contribute significantly to the perturbation. As a consequence, the direct wave is predominantly perturbed by scattered waves without a ground reflection, and the reflected wave is predominantly perturbed by scattered waves with a single ground reflection [see Fig. 1(a)]. The contribution of the latter scattered waves to the pressure p_2 in Eq. (1) is equal to $p_2 - p_{2,0} = p_{2,0}\exp(\psi_2) - p_{2,0} \approx p_{2,0}\psi_2$. This implies that the factor Q included in $p_{2,0}$ accounts not only for the effect of ground impedance on the unperturbed wave, but also for the effect of ground impedance on the scattered waves. Here we assume that the reflection coefficient Q for the unperturbed wave also applies to the scattered waves. This is a reasonable assumption, since the significant scattered waves do not deviate very much from the unperturbed path. We note, however, that this assumption is more accurate for $D \ll 1$ (geometrical-acoustics limit) than for $D \gg 1$.

The above arguments do not apply to the situation in which both the source and the receiver are very close to the ground ($z_s/a \ll 1$ and $z_r/a \ll 1$). In this situation, scattered waves with a single ground reflection also contribute to the perturbation of the direct wave, and also multiply scattered waves with multiple ground reflections may contribute to the perturbation of both waves. On the other hand, the coherence factor Γ approaches the value of unity if the source height and/or the receiver height approach zero (the unperturbed direct and reflected paths coincide for $z_s = 0$ and/or $z_r = 0$). Therefore, effects of turbulence on the time-averaged sound pressure level may be ignored if the source and/or the receiver are very close to the ground.

F. Numerical examples

For the numerical examples presented here, the Gaussian correlation function (18) of the refractive-index fluctuations was used, with correlation length $a = 1.1$ m and variance $\mu_0^2 = 10^{-5}$. The values of a and μ_0^2 are based on the work of Daigle *et al.*[15,17]

Figure 2 shows the time-averaged relative sound pressure level $\langle \Delta L_p \rangle$ for three situations with a rigid ground surface: $(f, z_s) = (500, 2.4)$, (1000, 1.5) and (2000, 0.9) (with frequency f in Hz and z_s in m). We used equal heights for the source and the receiver, i.e., $z_s = z_r$. The three situations correspond with $(D, \zeta) = (2.2 \text{ to } 4.0, 4.4)$, (0.9 to 1.8, 2.7) and (0.2 to 0.7, 1.6), respectively. The lines in the figure represent method A for $\rho = h$, $\rho = \frac{1}{2}h$ and $\rho = 2h$ (see Fig. 1), the symbols represent results of computations performed with the Crank–Nicholson parabolic equation (PE) method.[24,25] Results of method B are not shown in Fig. 2, but are indistinguishable from the PE results. The agreement of method A with the PE results is best for $\rho = h$. The relative sound pressure level is too low for $\rho = \frac{1}{2}h$ and too high for $\rho = 2h$. The agreement for $\rho = h$ confirms that $\rho = h$ is a good choice for the argument of the coherence factor in method A. For comparison, we included results for a nonturbulent atmosphere in Fig. 2. As expected, turbulence causes a reduction of the depth of the ground interference minimum. For the PE computations we used a grid spacing of 0.1λ (λ is the wavelength) and a system height of 2000 grid spacings, including an absorbing top layer of 1000 grid spacings to

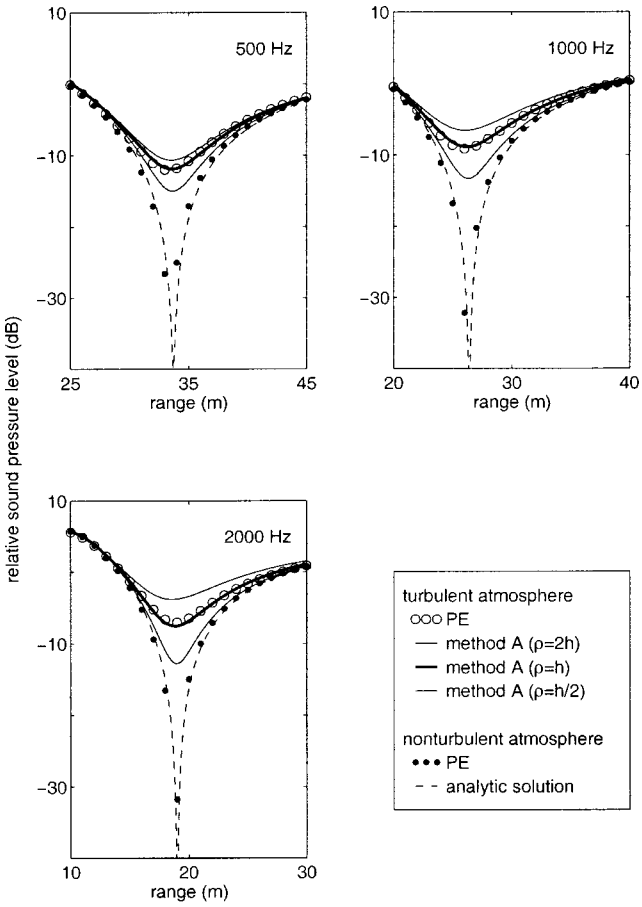


FIG. 2. Time-averaged relative sound pressure level $\langle \Delta L_p \rangle$ as a function of range L, for three situations with a rigid ground surface: $(f, z_s) = (500, 2.4)$, (1000,1.5) and (2000,0.9) (with frequency f in Hz, and $z_s = z_r$ in m). Results of method B (not shown) are indistinguishable from the PE results for a turbulent atmosphere.

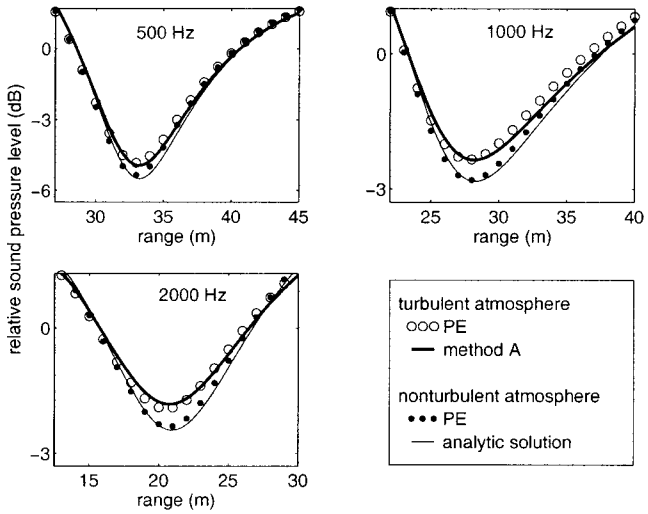


FIG. 3. Time-averaged relative sound pressure level $\langle \Delta L_p \rangle$ as a function of range L, for three situations with a finite-impedance ground surface: $(f, z_s, Z) = (500, 4.0, 6.97 + 5.13i)$, $(1000, 2.5, 5.67 + 3.14i)$ and (2000,1.4, 5.05+1.75i) (with f in Hz, $z_s = z_r$ in m and Z the normalized ground impedance). For method A, $\rho = h$ was used. Results of method B (not shown) are indistinguishable from the results of method A.

eliminate reflections from the top surface. The PE results were obtained by averaging over 500 random realizations of the turbulent atmosphere. Since the PE method assumes that the system has axial symmetry around the vertical axis through the source, the turbulent field is axisymmetric in this case. Therefore, the lines were also computed for axisymmetric turbulence.[26] The difference between the levels for axisymmetric turbulence and isotropic turbulence, however, vanishes in these cases, with a Gaussian correlation function of the refractive-index fluctuations.

Figure 3 shows that there is also good agreement between PE and method A (with $\rho = h$) for three situations with a finite-impedance ground surface: $(f, z_s, Z) = (500, 4.0, 6.97 + 5.13i)$, $(1000, 2.5, 5.67 + 3.14i)$ and $(2000, 1.4, 5.05 + 1.75i)$ (with f in Hz, $z_s = z_r$ in m), corresponding with $(D, \zeta) = (2.4$ to 4.0, 7.3), (1.0 to 1.8, 4.5) and (0.3 to 0.7, 2.5), respectively. Here, Z is the normalized ground impedance. The values of Z are typical for grassland.[30] Results of method B are not shown in Fig. 3, but are indistinguishable from results of method A. The effect of turbulence on the time-averaged sound pressure level is considerably smaller than in the previous situations with a rigid ground surface. This is because the interference minima in a nonturbulent atmosphere are deeper for a rigid ground than for a finite-impedance ground. For a rigid ground the direct and reflected waves have about equal amplitude at the receiver, and therefore the destructive interference is stronger than for a finite-impedance ground, where the amplitudes of the two waves are not equal.

The comparison shown in Fig. 3 was performed for relatively small values of range L, for computational efficiency. This is another reason for the fact that the effect of turbulence is relatively small in Fig. 3. With increasing range, the effect of turbulence becomes larger.

We determined the magnitude of the sum $\delta S \equiv \langle \chi_2 S_1 \rangle - \langle \chi_1 S_2 \rangle$ in the argument of the cosine function in Eq. (6), for the following situations with a rigid ground surface: $(f, z_s, L, \mu_0^2) = (1000, 2, 40, 10^{-5})$, $(250, 2, 40, 10^{-5})$, $(1000, 2, 40, 3\times 10^{-6})$, $(1000, 2, 10, 10^{-5})$, and $(1000, 5, 40, 10^{-5})$ (with f in Hz, z_s and L in m), with $0 \leqslant z_r \leqslant 4$ m. We found that in all situations $|\delta S| \leqslant 0.03$. For practical applications, the sum δS may be neglected in Eq. (6).

Figure 4 shows the variances $\langle S_1 S_1 \rangle$ and $\langle \chi_1 \chi_1 \rangle$ as a function of range L, for a system with source height $z_s = 0$, receiver height $z_r = 0.0333$ m, a rigid ground surface and frequency 1000 Hz. Source height $z_s = 0$ was chosen because the direct and reflected waves coincide in this case, so that $p_1 = p_2$, $\psi_1 = \psi_2$ and $\Gamma = 1$. This implies that ψ_1 follows directly from the pressure p, since $p_1 = \frac{1}{2} p$. Thus, we can compute ψ_1 from PE results and compare $\langle S_1 S_1 \rangle$ and $\langle \chi_1 \chi_1 \rangle$ with the predictions of methods A and B. In this case, χ_1 and S_1 are the log-amplitude and phase fluctuations not only of p_1 but also of the total pressure p (since $p_1 = \frac{1}{2} p$). Figure 4 shows that PE results for the variances (based on 1500 random realizations) are in good agreement with results of method B, up to a range of about $L = 100$ m. For $L > 100$ m, the log-amplitude variance $\langle \chi_1 \chi_1 \rangle$ saturates at a value of about 0.3 (this value agrees with experimental data in Ref. 17). Apparently, the region of weak fluctuations, i.e., the region where the Rytov solution is valid,[13,17,27,31] is limited to $L < 100$ m, in this case. The phase variance $\langle S_1 S_1 \rangle$

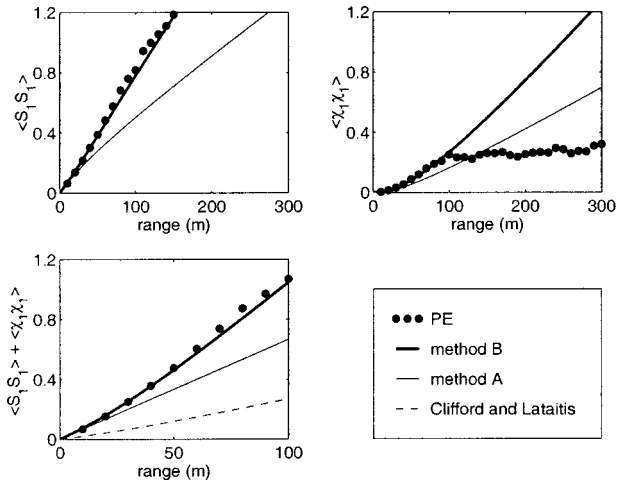


FIG. 4. The variances $\langle S_1 S_1 \rangle$ and $\langle \chi_1 \chi_1 \rangle$ and the sum $\langle S_1 S_1 \rangle + \langle \chi_1 \chi_1 \rangle$ as a function of range L, for a system with source height $z_s = 0$, receiver height $z_r = 0.0333$ m, a rigid ground surface, and frequency 1000 Hz.

increases linearly with range. The graph of $\langle S_1 S_1 \rangle$ is limited to $\langle S_1 S_1 \rangle \leqslant 1.2$; for $\langle S_1 S_1 \rangle > 1.2$, wraparound effects occur in the phase of the complex pressure computed with PE. The figure shows that method A predicts values for the variances that are too small, in this case. Also shown in the figure is a graph of the sum $\langle S_1 S_1 \rangle + \langle \chi_1 \chi_1 \rangle$ for $L \leqslant 100$ m, including a curve computed with Eq. (9) of Clifford and Lataitis.[19] We note that the sum $\langle S_1 S_1 \rangle + \langle \chi_1 \chi_1 \rangle$ is equal for isotropic turbulence and axisymmetric turbulence. The method of Clifford and Lataitis yields values that are too small in this case, as a consequence of the neglect of the ground reflection of scattered waves.

V. STATISTICAL DISTRIBUTIONS OF THE SOUND PRESSURE LEVEL

Turbulent fluctuations of the wind velocity and the temperature in the atmosphere cause fluctuations of the sound pressure level. It is of interest to calculate these fluctuations of the sound pressure level. Using Eq. (4), statistical distributions of the fluctuating relative sound pressure level can be computed from statistical distributions of the fluctuating variables χ_1, χ_2 and $\Delta S \equiv S_1 - S_2$. In this section, two methods are described to compute these distributions, for isotropic or axisymmetric turbulence.

A. Computational method 1: Gaussian distributions of χ_1, χ_2 and ΔS

Computational method 1 is based on the assumption of *independent* Gaussian distributions for χ_1, χ_2 and ΔS, and is similar to the method developed by Hidaka *et al.*[23] From this assumption, statistical distributions of χ_1, χ_2 and ΔS are easily constructed.[20] For the variance $\sigma_{\Delta S}^2 = \langle (\Delta S)^2 \rangle$ of the phase fluctuation difference ΔS, we use the relation[32] $\sigma_{\Delta S}^2 = 2B_S(0) - 2B_S(\rho)$, with free-field correlation function $B_S(\rho)$ and $\rho = h$ (see Secs. III and IV). For the variances $\sigma_{\chi_j}^2$ of the log-amplitude fluctuations χ_1 and χ_2, we use the relation $\sigma_{\chi_j}^2 = B_\chi(0)$, with free-field correlation function $B_\chi(\rho)$ (see Sec. IV). For the mean values of χ_1 and χ_2, we use the relation $\langle \chi_j \rangle = -\sigma_{\chi_j}^2$. For the mean value of ΔS, we use the relation $\langle S_j \rangle = -\langle \chi_j S_j \rangle$, which yields a value of zero for $\langle \Delta S \rangle$, if we assume $\langle \chi_1 S_1 \rangle \approx \langle \chi_2 S_2 \rangle$. Computational method 1 is more efficient than the rigorous computational method 2 described next.

B. Computational method 2: Monte Carlo evaluation of rigorous Rytov solution

Computational method 2 is based on a direct computation of the rigorous Rytov solution (32) for the fluctuating fields ψ_1 and ψ_2 in Eq. (2), making use of random realizations of the corresponding fluctuating field μ. This approach may be considered as Monte Carlo sampling of the Rytov solution.

The discrete approximation of Eq. (28) for the field μ is

$$\mu(\vec{r}) = \sum_{n=1}^{N} \exp(i\vec{\kappa}_n \cdot \vec{\rho})\, d\nu(x, \vec{\kappa}_n). \tag{47}$$

The discrete approximation of Eq. (32) for the field ψ is given by an analogous expression. Using a random number generator, sets of $\vec{\kappa}_n$ and $d\nu(x, \vec{\kappa}_n)$ can be generated that yield realizations of the fields μ and ψ that satisfy two conditions: (i) μ must be real, and (ii) the correlation function $\langle \mu(\vec{r} + \vec{s}) \mu(\vec{r}) \rangle$ must be equal to a given function $B_\mu(s)$. The first condition is satisfied by replacing each pair $\vec{\kappa}_n, d\nu(x, \vec{\kappa}_n)$ by two pairs: $\vec{\kappa}_n, \frac{1}{2} d\nu(x, \vec{\kappa}_n)$ and $-\vec{\kappa}_n, \frac{1}{2} d\nu^*(x, -\vec{\kappa}_n)$. The second condition is handled by writing the field μ as a superposition of discrete random Fourier modes.[25] This requires a distinction between isotropic turbulence and axisymmetric turbulence.

1. Isotropic turbulence

Three-dimensional isotropic fields $\mu(x,y,z)$ are obtained by using the following expressions:

$$\vec{K}_n \equiv (\kappa_{nx}, \kappa_{ny}, \kappa_{nz}) = (\kappa_n \cos\phi_n \sin\theta_n, \kappa_n \sin\phi_n \sin\theta_n, \kappa_n \cos\theta_n), \tag{48}$$

$$d\nu(x, \vec{\kappa}_n) = \sqrt{8\pi\Delta\kappa}\sqrt{\Phi_\mu(\kappa_n)}\,\kappa_n \exp(i\kappa_{nx} x + i\alpha_n), \tag{49}$$

where $\kappa_n \equiv |\vec{K}_n| = n\Delta\kappa$, ϕ_n and α_n are random angles between 0 and 2π, and $\cos\theta_n$ is a random number between 1 and -1. We have $\vec{\kappa}_n = (\kappa_{ny}, \kappa_{nz})$ in Eq. (47), so that $\vec{\kappa}_n \cdot \vec{\rho} = \kappa_{ny} y + \kappa_{nz} z$. For the case of a Gaussian correlation function $B_\mu(s) = \mu_0^2 \exp(-s^2/a^2)$, corresponding with a Gaussian spectral density $\Phi_\mu(\kappa) = \mu_0^2 a^3 \exp(-\kappa^2 a^2/4)/8\pi^{3/2}$, we use $\Delta\kappa = 0.1/a$ and $N = 80$.

To prove that this yields the correct correlation function $B_\mu(s)$, we write the spectral decomposition of $B_\mu(s)$ as follows (valid for homogeneous, isotropic random fields μ):

$$\begin{aligned} B_\mu(s) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \cos(\vec{K}\cdot\vec{s})\Phi_\mu(\kappa)\, d\vec{K} \\ &= \int_0^{2\pi}\int_0^{\pi}\int_0^{\infty} \cos(\vec{K}\cdot\vec{s})\Phi_\mu(\kappa)\kappa^2\, d\kappa \sin\theta\, d\theta\, d\phi \\ &= \left\langle 4\pi \int_0^{\infty} \cos(\vec{K}\cdot\vec{s})\Phi_\mu(\kappa)\kappa^2\, d\kappa \right\rangle, \end{aligned} \tag{50}$$

where angular brackets denote averaging over the angles θ and ϕ. The integral is approximated by a finite sum:

$$B_\mu(s) \approx \left\langle 4\pi\Delta\kappa \sum_n \cos(\vec{K}_n \cdot \vec{s}) \Phi_\mu(\kappa_n) \kappa_n^2 \right\rangle. \quad (51)$$

The same result is obtained from Eqs. (47) and (49):

$$\langle \mu(\vec{r}+\vec{s}) \mu(r) \rangle$$

$$= \left\langle 8\pi\Delta\kappa \sum_n \sqrt{\Phi_\mu(\kappa_n)} \kappa_n \cos(\vec{K}_n \cdot \vec{r} + \alpha_n + \vec{K}_n \cdot \vec{s}) \right.$$

$$\left. \times \sum_m \sqrt{\Phi_\mu(\kappa_m)} \kappa_m \cos(\vec{K}_m \cdot \vec{r} + \alpha_m) \right\rangle. \quad (52)$$

Here, we have used the condition that μ must be real. Now the averaged products $\cos(\vec{K}_n \cdot \vec{r} + \alpha_n + \vec{K}_n \cdot \vec{s}) \cos(\vec{K}_m \cdot \vec{r} + \alpha_m)$ vanish unless $n = m$. We have

$$\langle \cos(\vec{K}_n \cdot \vec{r} + \alpha_n + \vec{K}_n \cdot \vec{s}) \cos(\vec{K}_n \cdot \vec{r} + \alpha_n) \rangle$$

$$= \langle \cos^2(\vec{K}_n \cdot \vec{r} + \alpha_n) \cos(\vec{K}_n \cdot \vec{s}) \rangle$$

$$- \langle \sin(\vec{K}_n \cdot \vec{r} + \alpha_n) \cos(\vec{K}_n \cdot \vec{r} + \alpha_n) \sin(\vec{K} \cdot \vec{s}) \rangle$$

$$= \tfrac{1}{2} \langle \cos(\vec{K}_n \cdot \vec{s}) \rangle. \quad (53)$$

This gives Eq. (51), which completes the proof.

2. Axisymmetric turbulence

Axisymmetric turbulence is represented by two-dimensional fields $\mu(x,z)$. Two-dimensional fields $\mu(x,z)$ can be deduced from three-dimensional fields $\mu(x,y,z)$, but are obtained more efficiently with the following expressions:

$$\vec{K}_n \equiv (\kappa_{nx}, \kappa_{nz}) = (\kappa_n \cos\phi_n, \kappa_n \sin\phi_n), \quad (54)$$

$$d\nu(x, \vec{\kappa}_n) = \sqrt{4\pi\Delta\kappa} \sqrt{F_\mu(\kappa_n)\kappa_n} \exp(i\kappa_{nx} x + i\alpha_n), \quad (55)$$

where $\kappa_n \equiv |\vec{K}_n| = n\Delta\kappa$, and ϕ_n and α_n are random angles between 0 and 2π. We have $\vec{\kappa}_n = \kappa_{nz}$ in Eq. (47), so that $\vec{\kappa}_n \cdot \vec{\rho} = \kappa_{nz} z$. For the case of a Gaussian correlation function $B_\mu(s) = \mu_0^2 \exp(-s^2/a^2)$, corresponding with a Gaussian spectral density $F_\mu(\kappa) = \mu_0^2 a^2 \exp(-\kappa^2 a^2/4)/4\pi$, we use $\Delta\kappa = 0.1/a$ and $N = 60$.

The proof that this yields the correct correlation function $B_\mu(s)$ is analogous to the three-dimensional case. We approximate the spectral decomposition of $B_\mu(s)$ as follows:

$$B_\mu(s) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cos(\vec{K} \cdot \vec{s}) F_\mu(\kappa) d\vec{K}$$

$$\approx \left\langle 2\pi\Delta\kappa \sum_n \cos(\vec{K}_n \cdot \vec{s}) F_\mu(\kappa_n) \kappa_n \right\rangle, \quad (56)$$

where angular brackets denote averaging over the angle ϕ. The same result is obtained from Eqs. (47) and (55).

C. Numerical example

We define separate variances for the positive and negative deviations of the relative sound pressure level from its long-time averaged value $\langle \Delta L_p \rangle$: $\sigma_+^2 = \langle (\Delta L_p - \langle \Delta L_p \rangle)^2 \rangle_+$ and $\sigma_-^2 = \langle (\Delta L_p - \langle \Delta L_p \rangle)^2 \rangle_-$, where averaging is over ΔL_p greater than $\langle \Delta L_p \rangle$ and ΔL_p less than $\langle \Delta L_p \rangle$, respectively. The positive deviations from the long-time average are represented by the level $\Delta L_+ \equiv \langle \Delta L_p \rangle + \sigma_+$, the negative deviations by the level $\Delta L_- \equiv \langle \Delta L_p \rangle - \sigma_-$.

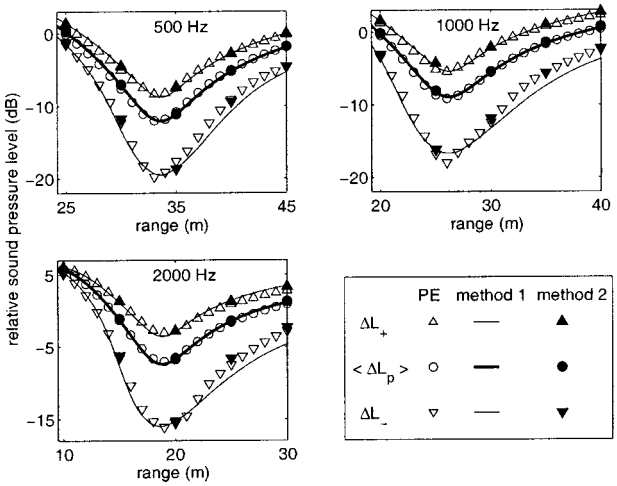


FIG. 5. Levels $\langle \Delta L_p \rangle$, ΔL_+ and ΔL_- as a function of range L, for axisymmetric turbulence, for three situations with a rigid ground surface: $(f, z_s) = (500, 2.4)$, $(1000, 1.5)$ and $(2000, 0.9)$ (with frequency f in Hz, and $z_s = z_r$ in m).

Figure 5 shows computational results of the levels $\langle \Delta L_p \rangle$, ΔL_+ and ΔL_- as a function of range L for axisymmetric turbulence with the Gaussian correlation function (18) with $a = 1.1$ m and $\mu_0^2 = 10^{-5}$. The figure shows results computed with PE (see Sec. IV F) and methods 1 and 2, for the three situations with a rigid ground surface which were also considered in Fig. 2: $(f, z_s) = (500, 2.4)$, $(1000, 1.5)$ and $(2000, 0.9)$ (with f in Hz and $z_s = z_r$ in m). There is good agreement between PE results and the results of method 2. The results of method 1 for ΔL_- deviate slightly. For comparison, we also performed computations with method 1 for isotropic turbulence; we found that the deviations from the curves for axisymmetric turbulence shown in Fig. 5 are very small. Results of method 2 were obtained by averaging over 5000 random realizations of the turbulent atmosphere. A similar agreement between PE and methods 1 and 2 as shown in Fig. 5 was found for the three situations with a finite-impedance ground which were considered in Fig. 3.

VI. CONCLUSIONS

A new, rigorous approach has been presented to the computation of the fluctuating field of a monopole source in a nonrefracting, turbulent atmosphere above a ground surface. For the time-averaged sound pressure level, we have derived an expression that contains triple integrals. From this expression, we have derived simpler expressions for asymptotic cases. The approach presented here is an improvement of the rigorous approach of Clifford and Lataitis,[19] who neglected the ground reflection of scattered waves. In Ref. 19, different solutions were presented for the two asymptotic cases of large and small values of the wave parameter. By means of the improved approach in the

present article, it has been shown that the solution in Ref. 19 for small values of the wave parameter is in fact valid for all values of the wave parameter.

For a rigid ground surface, the approach presented in this article is completely rigorous. For a finite-impedance ground surface, it has been shown that the approach is still a good approximation, provided the expression for the time-averaged sound pressure level is modified in the same way as for a nonturbulent atmosphere. In other words, the coherence factor in the expression for the time-averaged sound pressure level, which accounts for the coherence loss between the direct wave and the wave reflected by the ground, is approximately independent of the ground impedance.

For most practical computations of the time-averaged sound pressure level, the heuristic approach of Daigle *et al.*[15–17] appears sufficiently accurate. The advantage of this approach is that it requires less numerical effort, and, moreover, it can be generalized to a downward refracting atmosphere.[18] We have shown that the best choice for the argument of the coherence factor in the heuristic approach is the maximum separation between the two sound rays between the source and the receiver. In contrast, Daigle *et al.*[15] found better agreement of the heuristic approach with experimental data by setting the argument equal to half the maximum separation. The reason for this is not clear; inaccuracy in the description of the turbulent atmosphere may have played a role.

By considering various asymptotic cases, we have shown that equal results for the time-averaged sound pressure level are obtained in good approximation with the following three approaches: (*i*) the heuristic approach of Daigle *et al.*[15–17] (*ii*) the rigorous approach presented in this article, and (*iii*) the improved asymptotic solution of Clifford and Lataitis[19] (i.e., the solution for small values of the wave parameter).

Further, we have shown that for a comparison between PE results and analytical results, a distinction should be made between axisymmetric turbulence and isotropic turbulence. Only in the case of a Gaussian correlation function of the refractive-index fluctuations, the time-averaged sound pressure level is equal for axisymmetric turbulence and isotropic turbulence. In general, results of the axisymmetric PE method should be compared with analytical results for axisymmetric turbulence.

We have also applied the rigorous approach presented in this article to the computation of statistical distributions of the sound pressure level. Good agreement was obtained with PE results. For most practical computations, however, the method developed by Hidaka *et al.*[23] appears sufficiently accurate. This method requires considerably less numerical effort.

All computational methods considered in this article are limited to weak fluctuations, except for the PE method. The PE results indicate a saturation of the variance of log-amplitude fluctuations at a value of about 0.3, in a system with the source and the receiver close to the ground surface. Numerical simulation methods such as the PE method are currently the only tools that provide accurate predictions for the region of strong fluctuations.[33] The above-mentioned distinction between axisymmetric turbulence and isotropic turbulence should be kept in mind here.

ACKNOWLEDGMENTS

Vladimir Ostashev's research was supported by an NRC-ETL Research Associateship and by the U.S. Army Research Office under Contract No. DAAG 55-98-1-0463.

APPENDIX: THE INTEGRALS OVER y' AND z' IN EQ. (30)

The Fresnel integrals are $C(z)=\int_0^z \cos[(\pi/2)\tau^2]\,d\tau$ and $S(z)=\int_0^z \sin[(\pi/2)\tau^2]\,d\tau$. From Eqs. (7.4.38) and (7.4.39) of Ref. 34, the following relations are easily derived:

$$\int_{-\infty}^{\infty} \exp\left(i[aw^2+bw]\right) dw = \sqrt{\frac{\pi}{2a}}(1+i)\exp\left(-i\,\frac{b^2}{4a}\right), \tag{A1}$$

$$\int_{u}^{\infty} \exp(i[aw^2+bw])\,dw = \sqrt{\frac{\pi}{2a}}(1+i)\exp\left(-i\,\frac{b^2}{4a}\right)\frac{1}{2} \times\left\{1-(1-i)\left[C\left(\nu\sqrt{\frac{2}{\pi}}\right)+iS\left(\nu\sqrt{\frac{2}{\pi}}\right)\right]\right\}, \tag{A2}$$

with $\nu=u\sqrt{a}+\frac{1}{2}b/\sqrt{a}$.

The integral over y' in Eq. (30) can be evaluated by using Eq. (A1):

$$\int_{-\infty}^{\infty} \exp(i\kappa_y y')\exp\left(ik\frac{[y'-(x'/L)y]^2}{2(x'/L)(L-x')}\right)dy' = \exp(i\kappa_y y x'/L)\sqrt{\frac{\pi}{k}\frac{x'}{L}(L-x')}(1+i) \times\exp\left(-i\,\frac{(x'/L)(L-x')}{2k}\kappa_y^2\right). \tag{A3}$$

The integral over z' in Eq. (30) is written as

$$\int_{-\infty}^{\infty} \exp(i\kappa_z|z'|)\exp\left(ik\frac{(z'-z_1)^2}{2(x'/L)(L-x')}\right)dz' = I_1+I_2 \tag{A4}$$

with

$$I_1=\int_0^{\infty} \exp(i\kappa_z z')\exp\left(ik\frac{(z'-z_1)^2}{2(x'/L)(L-x')}\right)dz', \tag{A5}$$

$$I_2=\int_{-\infty}^{0} \exp(-i\kappa_z z')\exp\left(ik\frac{(z'-z_1)^2}{2(x'/L)(L-x')}\right)dz'. \tag{A6}$$

These integrals can be evaluated by using Eq. (A2):

$$I_1 = \exp(i\kappa_z z_1)\sqrt{\frac{\pi}{k}\frac{x'}{L}(L-x')}(1+i) \times \exp\left(-i\frac{(x'/L)(L-x')}{2k}\kappa_z^2\right)\frac{1}{2} \times [1-(1-i)F_F(\nu_1)], \quad \text{(A7)}$$

$$I_2 = \exp(-i\kappa_z z_1)\sqrt{\frac{\pi}{k}\frac{x'}{L}(L-x')}(1+i) \times \exp\left(-i\frac{(x'/L)(L-x')}{2k}\kappa_z^2\right)\frac{1}{2} \times [1-(1-i)F_F(\nu_2)] \quad \text{(A8)}$$

with $\nu_1 = -z_1\sqrt{\alpha} + \frac{1}{2}\kappa_z/\sqrt{\alpha}$, $\nu_2 = z_1\sqrt{\alpha} + \frac{1}{2}\kappa_z/\sqrt{\alpha}$, where $\alpha = k/[2(x'/L)(L-x')]$, and $F_F(x) = C(x\sqrt{2/\pi}) + iS(x\sqrt{2/\pi})$.

[1] U. Ingard and G. C. Maling, Jr., ''On the effect of atmospheric turbulence on sound propagated over ground,'' J. Acoust. Soc. Am. **35**, 1056–1058 (1963).
[2] P. H. Parkin and W. E. Scholes, ''The horizontal propagation of sound from a jet engine close to the ground, at Radlett,'' J. Sound Vib. **1**, 1–13 (1964).
[3] P. H. Parkin and W. E. Scholes, ''The horizontal propagation of sound from a jet engine close to the ground, at Hatfield,'' J. Sound Vib. **2**, 353–374 (1965).
[4] S. Canard-Caruana, S. Lewy, J. Vermorel, and G. Parmentier, ''Long range sound propagation near the ground,'' Noise Control Eng. J. **34**, 111–119 (1990).
[5] U. Ingard, ''A review of the influence of meteorological conditions on sound propagation,'' J. Acoust. Soc. Am. **25**, 405–411 (1953).
[6] M. Delany, ''Sound propagation in the atmosphere: a historical review,'' Acustica **38**, 201–223 (1977).
[7] J. E. Piercy, T. F. W. Embleton, and L. C. Sutherland, ''Review of noise propagation in the atmosphere,'' J. Acoust. Soc. Am. **61**, 1403–1418 (1977).
[8] T. F. W. Embleton, ''Tutorial on sound propagation outdoors,'' J. Acoust. Soc. Am. **100**, 31–48 (1996).
[9] L. C. Sutherland and G. A. Daigle, ''Atmospheric sound propagation,'' in *Encyclopedia of Acoustics*, edited by M. J. Crocker (Wiley, New York, 1997), pp. 341–365.
[10] V. I. Tatarski, *Wave Propagation in a Turbulent Medium* (McGraw-Hill, New York, 1961), pp. 3–26, 122–156, 173–188.
[11] V. I. Tatarskii, *The Effects of the Turbulent Atmosphere on Wave Propagation* (Israel Program for Scientific Translations, Jerusalem, 1971), pp. 1–33, 218–258, 414–423.
[12] L. A. Chernov, *Wave Propagation in a Random Medium* (McGraw-Hill, New York, 1960), pp. 5–11, 58–83.
[13] A. Ishimaru, *Wave Propagation and Scattering in Random Media* (Academic, New York, 1978), pp. 347–387, 407–422, 492, 513–519.
[14] V. E. Ostashev, *Acoustics in Moving Inhomogeneous Media* (E&FN Spon, London, 1997), pp. 16, 67, 193–218.
[15] G. A. Daigle, J. E. Piercy, and T. F. W. Embleton, ''Effects of atmospheric turbulence on the interference of sound waves near a hard boundary,'' J. Acoust. Soc. Am. **64**, 622–630 (1978).
[16] G. A. Daigle, ''Effects of atmospheric turbulence on the interference of sound waves above a finite impedance boundary,'' J. Acoust. Soc. Am. **65**, 45–49 (1979).
[17] G. A. Daigle, J. E. Piercy, and T. F. W. Embleton, ''Line-of-sight propagation through atmospheric turbulence near the ground,'' J. Acoust. Soc. Am. **74**, 1505–1513 (1983).
[18] A. L'Espérance, P. Herzog, G. A. Daigle, and J. R. Nicolas, ''Heuristic model for outdoor sound propagation based on an extension of the geometrical ray theory in the case of a linear sound speed profile,'' Appl. Acoust. **37**, 111–139 (1992).
[19] S. F. Clifford and R. J. Lataitis, ''Turbulence effects on acoustic wave propagation over a smooth surface,'' J. Acoust. Soc. Am. **73**, 1545–1550 (1983).
[20] E. M. Salomons, ''The fluctuating field of a monopole source in a turbulent atmosphere above a ground surface. Time-averaged sound pressure level and statistical distributions,'' in *Proceedings Eighth International Symposium on Long-Range Sound Propagation*, 9–11 September 1998 (The Pennsylvania State University, University Park, 1998), pp. 326–351.
[21] V. E. Ostashev, E. M. Salomons, S. F. Clifford, R. J. Lataitis, D. K. Wilson, Ph. Blanc-Benon, and D. Juvé, ''Sound propagation in a turbulent atmosphere near the ground. A parabolic-equation approach,'' J. Acoust. Soc. Am. **109**, 1894–1908 (2001).
[22] V. E. Ostashev and G. H. Goedecke, ''Interference of direct and ground reflected waves in a turbulent atmosphere,'' in *Proceedings Eighth International Symposium on Long-Range Sound Propagation*, 9–11 September 1998 (The Pennsylvania State University, University Park, 1998), pp. 313–325.
[23] T. Hidaka, K. Kageyama, and S. Masuda, ''Fluctuation of spherical wave propagating over a ground through atmospheric turbulence,'' J. Acoust. Soc. Jpn. (E) **6**, 247–256 (1985).
[24] K. E. Gilbert, R. Raspet, and X. Di, ''Calculation of turbulence effects in an upward-refracting atmosphere,'' J. Acoust. Soc. Am. **87**, 2428–2437 (1990).
[25] P. Chevret, Ph. Blanc-Benon, and D. Juvé, ''A numerical model for sound propagation through a turbulent atmosphere near the ground,'' J. Acoust. Soc. Am. **100**, 3587–3599 (1996).
[26] E. M. Salomons, ''Fluctuations of spherical waves in a turbulent atmosphere: effect of the axisymmetric approximation in computational methods,'' J. Acoust. Soc. Am. **108**, 1528–1534 (2000).
[27] S. M. Rytov, Yu. A. Kravtsov, and V. I. Tatarskii, *Principles of Statistical Radiophysics. Volume 4: Wave Propagation Through Random Media* (Springer, Berlin, 1989), pp. 67–71.
[28] V. E. Ostashev, F. Gerdes, V. Mellert, and R. Wandelt, ''Propagation of sound in a turbulent medium. II. Spherical waves,'' J. Acoust. Soc. Am. **102**, 2571–2578 (1997).
[29] S. M. Rytov, Yu. A. Kravtsov, and V. I. Tatarskii, *Principles of Statistical Radiophysics. Volume 2: Correlation Theory of Random Processes* (Springer, Berlin, 1988), p. 12.
[30] K. Attenborough, ''Acoustic impedance models for outdoor ground surfaces,'' J. Sound Vib. **99**, 521–544 (1985).
[31] L. R. Brownlee, ''Rytov's method and large fluctuations,'' J. Acoust. Soc. Am. **53**, 156–161 (1973).
[32] V. I. Tatarskii, ''Review of scintillation phenomena,'' in *Wave Propagation in Random Media (Scintillation)*, edited by V. I. Tatarskii, A. Ishimaru, and V. U. Zavorotny (I.O.P. and SPIE, Washington, DC, 1993), pp. 2–15.
[33] J. Martin, ''Simulation of wave propagation in random media: theory and applications,'' in *Wave Propagation in Random Media (Scintillation)*, edited by V. I. Tatarskii, A. Ishimaru, and V. U. Zavorotny (I.O.P. and SPIE, Washington, DC, 1993), pp. 463–486.
[34] M. Abramowitz and I. A. Stegun (eds.), *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables* (Dover, New York, 1972).

Sound propagation in a turbulent atmosphere near the ground: A parabolic equation approach

Vladimir E. Ostashev[a)]
NOAA/Environmental Technology Laboratory, 325 Broadway, Boulder, Colorado 80305 and Physics Department, New Mexico State University, Las Cruces, New Mexico 88003

Erik M. Salomons
TNO Institute of Applied Physics, P.O. Box 155, 2600 AD Delft, The Netherlands

Steven F. Clifford and Richard J. Lataitis
NOAA/Environmental Technology Laboratory, 325 Broadway, Boulder, Colorado 80305

D. Keith Wilson
U.S. Army Research Laboratory, 2800 Powder Mill Road, Adelphi, Maryland 20783-1197

Philippe Blanc-Benon and Daniel Juvé
Centre Acoustique, équipe LMFA/UMR CNRS 5509, Ecole Centrale de Lyon, 36 Avenue Guy de Collongue, BP 163, 69131 Ecully, France

(Received 3 September 1999; accepted for publication 3 January 2001)

The interference of the direct wave from the point source to the receiver and the wave reflected from the impedance ground in a turbulent atmosphere is studied. A parabolic equation approach for calculating the sound pressure p at the receiver is formulated. Then, the parabolic equation is solved by the Rytov method yielding expressions for the complex phases of direct and ground-reflected waves. Using these expressions, a formula for the mean squared sound pressure $\langle|p|^2\rangle$ is derived for the case of anisotropic spectra of temperature and wind velocity fluctuations. This formula contains the ''coherence factor,'' which characterizes the coherence between direct and ground-reflected waves. It is shown that the coherence factor is equal to the normalized coherence function of a spherical sound wave for line-of-sight propagation. For the case of isotropic turbulence, this result allows one to obtain analytical formulas for $\langle|p|^2\rangle$ for the Kolmogorov, Gaussian, and von Karman spectra of temperature and wind velocity fluctuations. Using these formulas, the effects of temperature and wind velocity fluctuations, and the effects of different spectra of these fluctuations on the mean squared sound pressure, are numerically studied. Also the effect of turbulent anisotropy on the interference of direct and ground reflected waves is numerically studied. Finally, it is shown that the mean squared sound pressure $\langle|p|^2\rangle$ calculated for the von Karman spectrum of temperature fluctuations agrees well with experimental data obtained in a laboratory experiment. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1356022]

PACS numbers: 43.28.Fp, 43.20.Bi, 43.20.Fn [LCS]

I. INTRODUCTION

For many problems of atmospheric acoustics, a sound field p is the sum of the direct wave from the source to the receiver and that reflected from the ground; see Fig. 1. The interference between these waves results in the interference maxima and minima of p. Quite often, the source and receiver are close to the ground while the horizontal distance between them is in the range from several tens to several hundred meters. In this case, the amplitude of the sound field p can be very small at interference minima in the absence of atmospheric turbulence. But turbulence is always present in the atmosphere. It results in random fluctuations in phase and amplitude of the direct and ground-reflected waves, and partial or complete destruction of the coherence between these waves. This destruction can cause an increase of the mean squared sound pressure $\langle|p|^2\rangle$ at the interference minima by tens of dB. The most recent theory of this phenomenon has been developed by Clifford and Lataitis,[1] who had generalized the results of previous research; e.g., see Ref. 2 and the references presented there.

Clifford and Lataitis used the Rytov method to calculate phase and log-amplitude fluctuations of the direct and ground-reflected waves due to refractive index fluctuations. They followed Tatarskii,[3] who considered the case when refractive index fluctuations are caused by temperature fluctuations. Furthermore, Clifford and Lataitis calculated $\langle|p|^2\rangle$ by an energy-conserving approach. Finally, they presented a formula for $\langle|p|^2\rangle$ for a Gaussian correlation function of refractive index fluctuations. These results made a significant contribution to the development of atmospheric acoustics and have been used by many specialists in the field.

Since the publication of the Clifford and Lataitis paper,[1] atmospheric acoustics has been developed significantly. It has become clear, e.g., Refs. 4, 5, that the effects of wind velocity fluctuations on the statistical moments of a sound field are usually greater than the effects of temperature fluctuations. Furthermore, while the Gaussian spectrum of turbu-

a)Electronic mail: vostashe@nmsu.edu

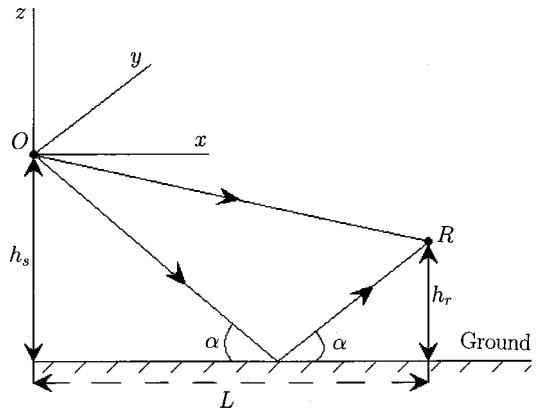


FIG. 1. The geometry of the problem. The source is located at the origin O of the coordinate system x, y, z, and the receiver is located at the point R. Line segments with arrows indicate directions of propagation of the direct and ground-reflected waves. Source and receiver heights above the ground are h_s and h_r, and the horizontal distance between them is L.

lence was widely used in atmospheric acoustics in the 1970s and 1980s, nowadays the more realistic Kolmogorov and von Karman spectra are employed, e.g., Refs. 4, 6. It has also been shown[5,7] that anisotropy of atmospheric turbulence may significantly affect the statistical moments of a sound field. Finally, the parabolic equation method has been used for analytical and numerical calculations of the statistical moments of a sound field, e.g., Refs. 4, 8, 9, 10.

The goal of our current research is to generalize the Clifford and Lataitis theory[1] by taking into account the aforementioned developments in atmospheric acoustics. In this paper, sound propagation over the impedance ground in the atmosphere with both temperature and velocity fluctuations is described on the basis of the parabolic equation. The Rytov method is used to approximately solve the parabolic equation and derive an equation for $\langle|p|^2\rangle$, which is valid for anisotropic spectra of temperature and velocity fluctuations. Then, we derive formulas for $\langle|p|^2\rangle$ for isotropic turbulence with the Kolmogorov, Gaussian, and von Karman spectra of temperature and velocity fluctuations. Furthermore, we numerically study the effects of temperature and velocity fluctuations, the effects of different spectra of turbulence, and the effects of turbulent anisotropy on $\langle|p|^2\rangle$. Finally, we show that the theoretical results for $\langle|p|^2\rangle$ for the von Karman spectrum of temperature fluctuations are in a good agreement with experimental data obtained in a large anechoic chamber in Ecole Centrale de Lyon.

Preliminary reports on a generalization of the Clifford and Lataitis theory[1] were simultaneously and independently presented at the 8th International Symposium on Long Range Sound Propagation.[11,12] These presentations gave two different approaches for the problem: a parabolic equation approach,[11] and an approach based on the spectral representation of refractive-index fluctuations.[12] The collaborative effort toward further development of these approaches resulted in two papers: the present paper and the companion one.[13] The papers complement each other, and explain the connection between results obtained by these approaches.

The paper is organized as follows. In Sec. I, the basic equations of the parabolic equation approach are presented. In Sec. II, expressions for $\langle|p|^2\rangle$ are derived. In Sec. III, the interference of the direct and ground-reflected waves in a turbulent atmosphere is studied numerically; furthermore, some of theoretical results obtained are compared with experimental data. And in Sec. V, the results obtained are summarized.

II. BASIC EQUATIONS

A. Geometry of the problem

Figure 1 shows the geometry of the problem, with a source and receiver in a turbulent atmosphere near an impedance ground surface. The source and receiver heights above the ground are h_s and h_r, respectively. The horizontal distance between them is L. The path length of the wave from the source to the receiver is given by $R_s=\sqrt{L^2+(h_s-h_r)^2}$, and that of the wave reflected from the ground is $R_i=\sqrt{L^2+(h_s+h_r)^2}$. The source is located at the origin of the Cartesian coordinate system x, y, z. The z-axis is directed upward, and the x-axis is in the horizontal direction from source to receiver. The ground surface coincides with the horizontal plane $z=-h_s$. The mean value of temperature T is constant and is equal to T_0, and the mean value of the wind velocity vector $\mathbf{v}$ is zero.

In the absence of atmospheric turbulence, the sound pressure p at the receiver is the sum of the sound fields of the direct and ground-reflected waves (e.g., Refs. 14, 15):

$$p=p_s^{(0)}+p_i^{(0)}=\frac{\exp(ikR_s)}{R_s}+Q\frac{\exp(ikR_i)}{R_i}. \tag{1}$$

Here, $p_s^{(0)}$ and $p_i^{(0)}$ correspond to the direct and reflected waves, the time-dependence of p is omitted, k is the sound wavenumber, and Q is the spherical-wave reflection coefficient which can also be interpreted as the apparent strength of the image source. For the locally reacting surface and in the approximation of near-grazing sound propagation when $R_s\sim R_i\gg h_s,h_r$, the coefficient Q is given by

$$Q=\frac{\sin\alpha-\beta+2\beta[1+i\sqrt{\pi}de^{-d^2}\operatorname{erfc}(-id)]}{\sin\alpha+\beta}. \tag{2}$$

Here, β is the admittance of the surface, α is the grazing angle of the wave incident on the ground (see Fig. 1), $\operatorname{erfc}(x)=(2/\sqrt{\pi})\int_x^\infty\exp(-t^2)dt$ is the complementary error function, and $d=\sqrt{ikR_i/2}[\beta+(h_s+h_r)/R_i]$ is the numerical distance. The range of applicability of Eq. (2) for Q is studied in detail in Refs. 14, 15, 16.

Due to fluctuations in the temperature $\tilde{T}(\mathbf{R})$ and velocity $\mathbf{v}(\mathbf{R})$ in the atmosphere, where $\mathbf{R}=(x,y,z)$, the sound field p becomes a random field. In many applications, it is important to know the mean squared sound pressure $\langle|p|^2\rangle$, where $\langle\,\rangle$ denotes an ensemble average.

B. Reformulation of the geometry of the problem

The impedance ground located at $z=-h_s$ complicates significantly the calculation of $\langle|p|^2\rangle$. This calculation would be simpler in a free space. Our main idea is that calculation of $\langle|p|^2\rangle$ over the impedance ground can be approximately

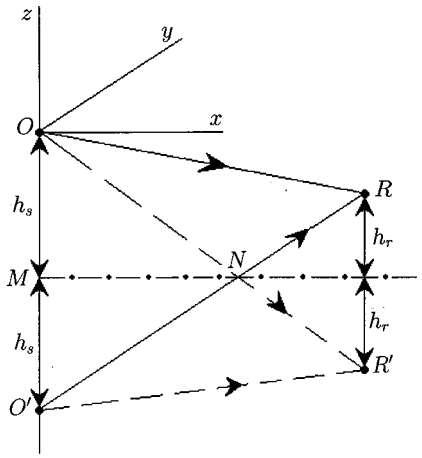


FIG. 2. The new geometry of the problem which includes the image source. Source, image source, receiver, and image receiver are located at the points O, O', R, and R', respectively. The ground is being removed, and a random medium is filling the whole space. The dash–dotted line coincides with the surface of the ground in the geometry of Fig. 1. Line segments with arrows indicate directions of propagation of different waves from the source and the image source.

replaced by the calculation for the geometry shown in Fig. 2. In this geometry, the ground is removed, the random medium fills the whole space and is symmetrical with respect to the plane $z=-h_s$, and the source at the point $\mathbf{R}_s=(0,0,0)$ is supplemented by an image source of the strength Q located at the point $\mathbf{R}_i=(0,0,-2h_s)$.

We now give a mathematical formulation of the problem, the geometry of which is shown in Fig. 2. If the random fields $\tilde{T}$ and $\mathbf{v}$ are relatively weak, the sound pressure p due to the source and image source can be sought as a solution of the parabolic equation:[4]

$$2ik\frac{\partial p}{\partial x}+\Delta_\perp p+2k^2\left(1+\frac{\mu_{\rm mov}}{2}\right)p=0, \tag{3}$$

where $\Delta_\perp=(\partial^2/\partial y^2)+(\partial^2/\partial z^2)$, and $\mu(\mathbf{R})$ describes the random medium in the geometry of Fig. 2. $\mu(\mathbf{R})$ can be expressed in the following form:

$$\mu(x,y,z)=\Theta(z+h_s)\epsilon_{\rm mov}(x,y,z) +[1-\Theta(z+h_s)]\epsilon_{\rm mov}(x,y,-z-2h_s). \tag{4}$$

Here, $\Theta(z)=1$ if $z\geqslant 0$, $\Theta(z)=0$ if $z<0$, and $\epsilon_{\rm mov}(\mathbf{R})=-\tilde{T}(\mathbf{R})/T_0-2v_x(\mathbf{R})/c_0$ is twice the fluctuations in the acoustic refractive index in a moving atmosphere in the geometry of Fig. 1, where v_x is the fluctuations of the wind velocity component in the x-direction, and c_0 is the mean value of the adiabatic sound speed. It follows from Eq. (4) that the random field $\mu=\epsilon$ above the plane $z=-h_s$ and is symmetrical with respect to this plane. We assume that the random field $\epsilon_{\rm mov}$ has a Gaussian distribution, $\langle\epsilon_{\rm mov}\rangle=0$, and that $\tilde{T}$ and $\mathbf{v}$ are uncorrelated. The latter assumption is valid for isotropic turbulence.[3]

The initial condition to the parabolic equation (3) is formulated in the plane $x=0$:

$$p(x=0,\mathbf{r})=\frac{2\pi i}{k}\,\delta(\mathbf{r}-\mathbf{r}_s)+\frac{2\pi iQ}{k}\,\delta(\mathbf{r}-\mathbf{r}_i), \tag{5}$$

where $\mathbf{r}=(y,z)$ are the transverse coordinates, $\mathbf{r}_s=(0,0)$, $\mathbf{r}_i=(0,-2h_s)$, and $\delta(\mathbf{r})$ is the delta-function. The first term on the right-hand side of Eq. (5) corresponds to the point source located at $\mathbf{R}_s=(0,0,0)$, while the second term corresponds to the image source of the strength Q located at $\mathbf{R}_i=(0,0,-2h_s)$.

In the limiting case $\epsilon_{\rm mov}=0$, the parabolic equation (3) with the initial condition (5) can be easily solved. The result is

$$p(x,\mathbf{r})=p_s^{(0)}+p_i^{(0)}=\frac{\exp\left(ikx+\dfrac{ik(\mathbf{r}-\mathbf{r}_s)^2}{2x}\right)}{x} +Q\frac{\exp\left(ikx+\dfrac{ik(\mathbf{r}-\mathbf{r}_i)^2}{2x}\right)}{x}, \tag{6}$$

where $p_s^{(0)}$ and $p_i^{(0)}$ are the fields of the waves from the source and image source to the receiver. This formula for p coincides with Eq. (1) in the approximation of near-grazing sound propagation.

The parabolic equation (3) with the initial condition (5) and the random field $\mu(\mathbf{R})$ given by Eq. (4) are the starting equations of our approach for calculating $\langle|p|^2\rangle$. Other statistical moments of a sound field can also be calculated on the basis of Eqs. (3)–(5). It is shown in Appendix A that, when calculating p, the new geometry of the problem shown in Fig. 2 is equivalent to the original one in Fig. 1 if $Q=1$. It is also shown in Appendix A that, for $Q=1$, both geometries are equivalent to the geometry in Ref. 13. The latter geometry is similar to that in Fig. 2; however, instead of the image source, an image receiver is used. Since in the limiting case $\epsilon_{\rm mov}=0$, Eqs. (3)–(5) correctly describe sound propagation near the impedance ground, we expect that they can be used to approximately describe sound propagation if the atmospheric turbulence is relatively weak. The main advantage of the geometry in Fig. 2 is that the parabolic equation (3) allows us to use various methods for calculating $\langle|p|^2\rangle$ and other statistical moments of the sound field, namely, geometrical acoustics, Rytov method, Markov approximation, Feynman path-integral technique, diagram technique, etc. These methods are well developed in theories of wave propagation in unbounded random media.[3,4,17,18] Note that in the theories of waves in random media, an approach for calculating the statistical moments of a field p which satisfies Eq. (3) is usually called the parabolic equation approach or method, e.g., see Refs. 4, 17, 18.

III. RYTOV METHOD

A. Mean squared sound pressure

For the geometry in Fig. 2, the sound field p at the receiver is the sum of the field due to the source and the field due to the image source: $p(\mathbf{R})=p_s(\mathbf{R})+p_i(\mathbf{R})$, where the subscripts s and i stand for source and image source. Sound waves from both sources exhibit phase and amplitude fluctuations in the random medium. Following the Rytov method, we express p_s and p_i in the forms $p_s=p_s^{(0)}e^{\psi_s}$ and $p_i=p_i^{(0)}e^{\psi_i}$. Here, $p_s^{(0)}$ and $p_i^{(0)}$ are the sound fields in an

atmosphere without random inhomogeneities and are given by the first and second terms in Eq. (6) [or in Eq. (1)], respectively. Furthermore, $\psi_s=\chi_s+iS_s$ are fluctuations in the complex phase of the wave emitted by the source, and $\psi_i=\chi_i+iS_i$ are those in the wave emitted by the image source, where χ and S with the corresponding indexes are fluctuations in the log-amplitude and phase of these waves. Thus, the Rytov method allows us to express the sound field p at the receiver in the form

$$p(\mathbf{R})=p_s^{(0)}(\mathbf{R})e^{\psi_s(\mathbf{R})}+p_i^{(0)}(\mathbf{R})e^{\psi_i(\mathbf{R})}. \tag{7}$$

Using the Rytov method, χ_j and S_j can be expressed as a series in ϵ_{mov}: $\chi_j=O(\epsilon_{\mathrm{mov}})+O(\epsilon_{\mathrm{mov}}^2)+\cdots$, and $S_j=O(\epsilon_{\mathrm{mov}})+O(\epsilon_{\mathrm{mov}}^2)+\cdots$. Hereinafter, the subscript j stands for s or i. It can be shown from these series that χ_j and S_j have Gaussian distributions, and that

$$\begin{aligned}\overline{\chi_j}&=O(\langle\epsilon_{\mathrm{mov}}^2\rangle)+O(\langle\epsilon_{\mathrm{mov}}^4\rangle)+\cdots,\\ \overline{S_j}&=O(\langle\epsilon_{\mathrm{mov}}^2\rangle)+O(\langle\epsilon_{\mathrm{mov}}^4\rangle)+\cdots.\end{aligned} \tag{8}$$

Furthermore, $\langle\chi_j^2\rangle$ and $\langle S_j^2\rangle$ can also be expressed as series in even powers of $\langle\epsilon_{\mathrm{mov}}^2\rangle$, analogous to series (8). In the derivations below, we keep terms of order $\langle\epsilon_{\mathrm{mov}}^2\rangle$ and neglect those proportional to $\langle\epsilon_{\mathrm{mov}}^4\rangle$.

Substituting the values of $p_s^{(0)}$ and $p_i^{(0)}$ from Eq. (1) into Eq. (7), squaring both sides of the resulting formula, and performing ensemble averaging, one obtains the formula for the mean squared sound pressure:

$$\begin{aligned}\langle|p|^2\rangle=&\frac{1}{R_s^2}\langle e^{\psi_s+\psi_s^*}\rangle+\frac{|Q|^2}{R_i^2}\langle e^{\psi_i+\psi_i^*}\rangle\\ &+\frac{|Q|}{R_sR_i}[e^{i\Omega}\langle e^{\psi_s^*+\psi_i}\rangle+e^{-i\Omega}\langle e^{\psi_s+\psi_i^*}\rangle],\end{aligned} \tag{9}$$

where $Q=|Q|e^{i\Omega}$. In Eq. (9), $\langle e^{\psi_s+\psi_s^*}\rangle=\langle e^{\psi_i+\psi_i^*}\rangle=1$. These equalities are consequences of the energy conservation in the parabolic equation approximation.[1,17] The calculation of the terms $\langle e^{\psi_s^*+\psi_i}\rangle$ and $\langle e^{\psi_s+\psi_i^*}\rangle$ in Eq. (9) is presented in Appendix B. Substitution of these terms into Eq. (9) yields

$$\begin{aligned}\langle|p|^2\rangle=&\frac{1}{R_s^2}+\frac{|Q|^2}{R_i^2}+\frac{2|Q|C}{R_sR_i}\cos[(R_i-R_s)k+\Omega\\ &+\langle\chi_sS_i\rangle-\langle\chi_iS_s\rangle+O(\langle\epsilon_{\mathrm{mov}}^4\rangle)].\end{aligned} \tag{10}$$

Here, the ''coherence'' factor C describes the coherence between the wave emitted by source and that emitted by image source, and is given by

$$\begin{aligned}C=\exp[&\langle\chi_s\chi_i\rangle+\langle S_sS_i\rangle-\tfrac{1}{2}(\langle\chi_s^2\rangle+\langle\chi_i^2\rangle+\langle S_s^2\rangle+\langle S_i^2\rangle)\\ &+O(\langle\epsilon_{\mathrm{mov}}^4\rangle)].\end{aligned} \tag{11}$$

In the case $Q=1$, Eqs. (10) and (11) are equivalent to Eqs. (6) and (7) from Ref. 13. Equations (10) and (11) differ from analogous equations in Ref. 1 only by the term $\langle\chi_iS_i\rangle-\langle\chi_sS_s\rangle$ in the argument of the cosine, which is present in Ref. 1 and is not present in Eq. (10). The reason for this difference is explained in Appendix B. This small difference does not affect the final result since in the Rytov approximation the term $\langle\chi_iS_i\rangle-\langle\chi_sS_s\rangle$ can be omitted.

B. Log-amplitude and phase fluctuations

To obtain the values of χ_s, χ_i, S_s, and S_i in Eqs. (10) and (11), we substitute $p_j=p_j^{(0)}e^{\psi_j}$ into Eq. (3) and use the first Rytov approximation for ψ_j. The result is (e.g., Ref. 19)

$$\begin{aligned}\psi_j(x,\mathbf{r})=&\frac{k^2}{4\pi}\int_0^x dx'\int_{-\infty}^{\infty}dy'\int_{-\infty}^{\infty}dz'\,\mu(x',\mathbf{r}')\\ &\times\frac{\exp\left[ik(x-x')+\dfrac{ik(\mathbf{r}-\mathbf{r}')^2}{2(x-x')}\right]}{x-x'}\frac{p_j^{(0)}(x',\mathbf{r}')}{p_j^{(0)}(x,\mathbf{r})}.\end{aligned} \tag{12}$$

We now rewrite formula (12) for the wave emitted by source and that emitted by the image source. For both waves, we have $x=L$, and $\mathbf{r}=(0,h_r-h_s)$. We also take into account that μ is given by Eq. (4). Substituting $p_s^{(0)}$ from Eq. (6) into Eq. (12), one obtains the complex amplitude of the wave emitted by the source:

$$\begin{aligned}\psi_s=&\frac{k^2L}{4\pi}\int_0^L\frac{dx}{x(L-x)}\int_{-\infty}^{\infty}dy\left[\int_{-h_s}^{\infty}dz\,\epsilon_{\mathrm{mov}}(x,y,z)\right.\\ &\left.+\int_{-\infty}^{-h_s}dz\,\epsilon_{\mathrm{mov}}(x,y,-z-2h_s)\right]\\ &\times\exp\left\{\frac{ikL[y^2+(z+(h_s-h_r)x/L)^2]}{2x(L-x)}\right\}.\end{aligned} \tag{13}$$

Substituting $p_i^{(0)}$ from Eq. (6) into Eq. (12), one obtains the complex amplitude ψ_i of the wave emitted by the image source. It is worthwhile to express ψ_i as a sum of two terms corresponding to two terms in Eq. (4): $\psi_i=\psi_1+\psi_2$, where

$$\begin{aligned}\psi_1=&\frac{k^2L}{4\pi}\int_0^L\frac{dx}{x(L-x)}\int_{-\infty}^{\infty}dy\int_{-\infty}^{-h_s}dz\,\epsilon_{\mathrm{mov}}(x,y,-z-2h_s)\\ &\times\exp\left\{\frac{ikL[y^2+(z+2h_s-(h_s+h_r)x/L)^2]}{2x(L-x)}\right\},\end{aligned} \tag{14}$$

$$\begin{aligned}\psi_2=&\frac{k^2L}{4\pi}\int_0^L\frac{dx}{x(L-x)}\int_{-\infty}^{\infty}dy\int_{-h_s}^{\infty}dz\,\epsilon_{\mathrm{mov}}(x,y,z)\\ &\times\exp\left\{\frac{ikL[y^2+(z+2h_s-(h_s+h_r)x/L)^2]}{2x(L-x)}\right\}.\end{aligned} \tag{15}$$

In Eqs. (13)–(15) for ψ_s and ψ_i, the limits of integration over y and z are from $-\infty$ to ∞. However, integration volumes can be much narrower. Indeed, it is well known[17] that in the geometrical acoustics approximation the statistical moments of phase and log-amplitude fluctuations are given by an integral along the path of the wave from source to receiver, with the integrand proportional to the correlation function of random inhomogeneities in a medium. In the geometry in Fig. 2, there are two such paths, OR and $O'R$. Therefore, the volumes essential for integration in Eqs. (13)–(15) are the tubes along OR and $O'R$, with transverse scales of order l, where l is a scale of the correlation function of medium inhomogeneities. Note that the geometrical acoustics approximation is valid if the wave parameter $D=L/(kl^2)$ is much less than 1.

Let us simplify Eqs. (13)–(15) in this approximation. Hereinafter, we assume that $l \ll h_s, h_r$. In this case, the integration volumes along OR and $O'R$ are relatively thin tubes with respect to h_s, h_r, and L. Furthermore, the integration volume along OR is always located in the region $z > -h_s$. Therefore, the second term in the square brackets in Eq. (13) can be neglected, and the formula for the complex phase ψ_s is simplified:

$$\psi_s = \frac{k^2 L}{4\pi} \int_0^L \frac{dx}{x(L-x)} \int_{-\infty}^{\infty} dy \int_{-h_s}^{\infty} dz\, \epsilon_{\mathrm{mov}}(x,y,z) \times \exp\left\{ \frac{ikL[y^2 + (z + (h_s - h_r)x/L)^2]}{2x(L-x)} \right\}. \quad (16)$$

The integration volume along $O'R$ can be split into two volumes. The first one is along the line segment $O'N$, and the second one is along the line segment NR (see Fig. 2). These two integration volumes correspond to ψ_1 and ψ_2. Since the integration volume along $O'N$ is relatively thin, the limits of integration over x in Eq. (14) for ψ_1 can be taken from 0 to L_1, where $L_1 = L/(h_r/h_s + 1)$ is the length of the line segment MN. Introducing a new integration variable $z' = -z - 2h_s$ in Eq. (14) and then denoting $z' = z$, one obtains

$$\psi_1 = \frac{k^2 L}{4\pi} \int_0^{L_1} dx \int_{-\infty}^{\infty} dy \int_{-h_s}^{\infty} dz\, \frac{\epsilon_{\mathrm{mov}}(x,\mathbf{r})}{x(L-x)} \times \exp\left\{ \frac{ikL[y^2 + (z + (h_s + h_r)x/L)^2]}{2x(L-x)} \right\}. \quad (17)$$

Analogously, the lower limit of integration in Eq. (15) can be approximately replaced by L_1:

$$\psi_2 = \frac{k^2 L}{4\pi} \int_{L_1}^{L} dx \int_{-\infty}^{\infty} dy \int_{-h_s}^{\infty} dz\, \frac{\epsilon_{\mathrm{mov}}(x,\mathbf{r})}{x(L-x)} \times \exp\left\{ \frac{ikL[y^2 + (z + 2h_s - (h_s + h_r)x/L)^2]}{2x(L-x)} \right\}. \quad (18)$$

Now consider the opposite limiting case of Fraunhofer diffraction when $D \gg 1$. In this case, the volumes essential for integration in Eqs. (13)–(15) are again the tubes along OR and $O'R$ but their transverse scales are of order $\sqrt{x/k}$ and are much greater than l, at least for $x \sim L$. Nevertheless, the integration volume along OR is still located in the region $z > -h_s$ if $h_0 > \sqrt{L/k}$. Here, for simplicity, we assume that source and receiver are at the same height above the ground, and denote $h_0 = h_s = h_r$. In this case, Eq. (16) is a good approximation for ψ_s. For the considered case $D \gg 1$, it can be analogously shown that ψ_1 and ψ_2 can be approximated by Eqs. (17) and (18) if $h_0 > \sqrt{L/k}$.

If velocity fluctuations are zero, Eqs. (16)–(18) coincide with the equations for ψ_s and ψ_i presented in Ref. 1. Our derivation clearly shows which terms are neglected when deriving these equations.

C. Derivation of a formula for the coherence factor *C*

Equations (16)–(18) for ψ_s and $\psi_i = \psi_1 + \psi_2$ will be used to calculate the statistical moments in Eqs. (10) and (11). Note that in Eqs. (10) and (11), $\chi_s = \mathrm{Re}\,\psi_s$; $S_s = \mathrm{Im}\,\psi_s$; $\chi_i = \mathrm{Re}\,\psi_i = \mathrm{Re}\,\psi_1 + \mathrm{Re}\,\psi_2 = \chi_1 + \chi_2$; and $S_i = \mathrm{Im}\,\psi_i = \mathrm{Im}\,\psi_1 + \mathrm{Im}\,\psi_2 = S_1 + S_2$. Here, $\chi_1 = \mathrm{Re}\,\psi_1$, $\chi_2 = \mathrm{Re}\,\psi_2$, $S_1 = \mathrm{Im}\,\psi_1$, and $S_2 = \mathrm{Im}\,\psi_2$.

With these notations, and taking into account that ψ_1 and ψ_2 have a negligible correlation, one obtains the following formula for the coherence factor C:

$$C = \exp[\langle \chi_s \chi_1 \rangle + \langle S_s S_1 \rangle + \langle \chi_s \chi_2 \rangle + \langle S_s S_2 \rangle - \tfrac{1}{2}[\langle \chi_s^2 \rangle + \langle S_s^2 \rangle + \langle \chi_1^2 \rangle + \langle S_1^2 \rangle + \langle \chi_2^2 \rangle + \langle S_2^2 \rangle] + O(\langle \epsilon_{\mathrm{mov}}^4 \rangle)]. \quad (19)$$

Calculations of the statistical moments in Eq. (19) are straightforward but involved. The result is simpler when the statistical moments are grouped in pairs. The derivation of the expression for a sum $\langle \chi_s \chi_2 \rangle + \langle S_s S_2 \rangle$ is presented in Appendix C. The derivation is similar but not identical to the derivation in Ref. 1, where the case of isotropic turbulence is considered. The derivation of expressions for other pairs of the statistical moments in Eq. (19) are analogous to that in Appendix C. As a result, one obtains the following formulas for pairs of the statistical moments in Eq. (19):

$$\langle \chi_s \chi_1 \rangle + \langle S_s S_1 \rangle = \frac{k^2 L}{16 h_r} \int_0^{2h_r L_1/L} dz\, b_{\mathrm{eff}}(0,z) + \frac{k^2}{8\pi} \int_0^{L_1} dx \left[\int_{-\infty}^{-2xh_r/L} d\xi\, \frac{\sin\left(\frac{\xi kL(\xi/2 - m_1)}{x(L-x)} \right)}{\xi} b_{\mathrm{eff}}(0, \xi + 2h_r x/L) - \int_{-2xh_r/L}^{\infty} d\xi\, \frac{\sin\left(\frac{\xi kL(\xi/2 + m_2)}{x(L-x)} \right)}{\xi} b_{\mathrm{eff}}(0, \xi + 2h_r x/L) \right], \quad (20)$$

$$\langle \chi_s \chi_2 \rangle + \langle S_s S_2 \rangle = \frac{k^2 L}{16 h_s} \int_0^{2h_s(1 - L_1/L)} dz\, b_{\mathrm{eff}}(0,z) + \frac{k^2}{8\pi} \int_{L_1}^{L} dx \left[\int_{-\infty}^{-2h_s(1 - x/L)} d\xi\, \frac{\sin\left(\frac{\xi kL(\xi/2 - m_1)}{x(L-x)} \right)}{\xi} \times b_{\mathrm{eff}}(0, \xi + 2h_s(1 - x/L)) - \int_{-2h_s(1 - x/L)}^{\infty} d\xi\, \frac{\sin\left(\frac{\xi kL(\xi/2 - m_2)}{x(L-x)} \right)}{\xi} b_{\mathrm{eff}}(0, \xi + 2h_s(1 - x/L)) \right], \quad (21)$$

$$\langle \chi_s^2 \rangle + \langle S_s^2 \rangle = \frac{k^2 L b_{\text{eff}}(0,0)}{8} + \frac{k^2}{4\pi} \int_0^L dx \int_0^\infty d\xi\, b_{\text{eff}}(0,\xi) \times \frac{\sin\left(\dfrac{\xi k L(-\xi/2 - m_1)}{x(L-x)} \right)}{\xi}, \tag{22}$$

$$\langle \chi_1^2 \rangle + \langle S_1^2 \rangle = \frac{k^2 L_1 b_{\text{eff}}(0,0)}{8} + \frac{k^2}{4\pi} \int_0^{L_1} dx \times \int_0^\infty d\xi\, b_{\text{eff}}(0,\xi) \frac{\sin\left(\dfrac{\xi k L(-\xi/2 - m_2)}{x(L-x)} \right)}{\xi}, \tag{23}$$

$$\langle \chi_2^2 \rangle + \langle S_2^2 \rangle = \frac{k^2 (L - L_1) b_{\text{eff}}(0,0)}{8} + \frac{k^2}{4\pi} \int_{L_1}^L dx \times \int_0^\infty d\xi\, b_{\text{eff}}(0,\xi) \frac{\sin\left(\dfrac{\xi k L(m_2 - \xi/2)}{x(L-x)} \right)}{\xi}. \tag{24}$$

In these equations,

$$m_1(x) = -(1 - x/L) h_s - x h_r / L, \quad m_2(x) = -(1 - x/L) h_s + x h_r / L, \tag{25}$$

and the function $b_{\text{eff}}(\mathbf{r})$ is given by[4]

$$b_{\text{eff}}(\mathbf{r}) = \int_{-\infty}^{\infty} B_{\text{eff}}(x, \mathbf{r}) dx = 2\pi \int d^2 K_\perp\, e^{i \mathbf{K}_\perp \cdot \mathbf{r}} \Phi_{\text{eff}}(0, \mathbf{K}_\perp). \tag{26}$$

Here, $B_{\text{eff}}(\mathbf{R})$ is the correlation function of the random field ϵ_{mov}, $\Phi_{\text{eff}}(\mathbf{K})$ is the three-dimensional spectral density of ϵ_{mov} and is related to B_{eff} by the Fourier transform: $B_{\text{eff}}(\mathbf{R}) = \int d^3 K e^{i \mathbf{K} \cdot \mathbf{R}} \Phi_{\text{eff}}(\mathbf{K})$, where $\mathbf{K} = (K_x, \mathbf{K}_\perp)$ is the wave vector, and $\mathbf{K}_\perp = (K_y, K_z)$. When deriving Eqs. (20)–(24), we used the Markov approximation: $B_{\text{eff}}(x, \mathbf{r}) = \delta(x) b_{\text{eff}}(\mathbf{r})$.

The correlation function B_{eff} can be expressed in terms of the correlation functions B_T and B_{xx} of the random fields $\tilde{T}$ and v_x, and the spectral density Φ_{eff} can be expressed in terms of the spectral densities Φ_T and Φ_{xx} of these random fields:[4]

$$B_{\text{eff}}(\mathbf{R}) = \frac{B_T(\mathbf{R})}{T_0^2} + \frac{4 B_{xx}(\mathbf{R})}{c_0^2}, \quad \Phi_{\text{eff}}(\mathbf{K}) = \frac{\Phi_T(\mathbf{K})}{T_0^2} + \frac{4 \Phi_{xx}(\mathbf{K})}{c_0^2}. \tag{27}$$

The term $\langle \chi_s S_i \rangle - \langle \chi_i S_s \rangle$ in the argument of cosine in Eq. (10) can be calculated analogously to the terms on the left-hand sides in Eqs. (20)–(24). It can be shown that the former term is much less than 1, and hence can be ignored. The same result is obtained in Refs. 1, 13.

Equations (20)–(24) can be simplified further. Let us consider the order of magnitudes of different terms in the argument of the sine in Eq. (22). First, the integration variable x satisfies the following relationship: $x(L-x) \sim L^2$. Second, since ξ is an argument of the function $b_{\text{eff}}(0,\xi)$ in the integrand in Eq. (22), it is of the order of the scale l of random inhomogeneities. Third, it follows from Eq. (25) that the function $m_1(x) \sim -h_m$, where h_m is a maximum of source and receiver heights [i.e., $h_m = \max(h_s, h_r)$]. Therefore, taking into account that $h_m \gg l$, we have $-\xi/2 - m_1 \sim h_m$. These estimates allow us to obtain the following relationship for the ratio of the sine and ξ in Eq. (22), which is denoted by s:

$$s = \frac{\sin\left(\dfrac{\xi k L(-\xi/2 - m_1)}{x(L-x)} \right)}{\xi} \sim \frac{\sin\left(\dfrac{\xi k h_m}{L} \right)}{\xi} = \frac{\sin\left(\dfrac{\xi}{l} \dfrac{h_m}{lD} \right)}{\xi}. \tag{28}$$

We assume that $D \ll h_m / l$. Since $D = L/(kl^2)$, this inequality is fulfilled for a relatively small distance L of sound propagation and/or for relatively high sound frequency $f = k c_0/(2\pi)$. The inequality $D \ll h_m/l$ is always valid in the geometrical acoustics approximation. It is also valid beyond the limits of this approximation because $h_m \gg l$. Since $D \ll h_m/l$, the argument of the sine in Eq. (28) is much greater than 1, and s is approximately equal to $\pi \delta(\xi)$.[20] Substituting the value of s into Eq. (22) and calculating the integrals over ξ and x, it can be shown that the second term on the right-hand side of this equation is equal to the first one. Analogously, it can be shown that for the considered case $D \ll h_m/l$, the second terms on the right-hand sides of Eqs. (20), (21), (23), and (24) are equal to the first terms.

Thus, the statistical moments on the left-hand sides of Eqs. (20)–(24) are just twice the first terms on the right-hand sides of these equations. Substituting the values of these statistical moments into Eq. (19) yields

$$C = \exp\left\{ -\frac{k^2 L}{4} \left[b_{\text{eff}}(0,0) - \frac{1}{2h_r} \int_0^{2h_r L_1/L} dz\, b_{\text{eff}}(0,z) - \frac{1}{2h_s} \int_0^{2h_s(1 - L_1/L)} dz\, b_{\text{eff}}(0,z) \right] \right\}. \tag{29}$$

Equation (29) is a generalization of formula for C in Ref. 1 for the case of both temperature and velocity fluctuations, and for the case of anisotropic spectra of these fluctuations. Furthermore, it will be shown at the end of the next subsection that C is given by Eq. (29) for all values of D if the Rytov approximation is valid. On the other hand, in Ref. 1 C is given by Eq. (29) only for $D \ll 1$ and is given by a different formula if $D \geqslant 1$. This difference is due to the fact that ground reflection of scattered waves is ignored in Ref. 1 as it is indicated in Refs. 12, 13. In the rest of this paper, we will obtain results [except Eq. (40)] that were not considered in Ref. 1.

D. Analysis of the coherence factor *C*

Taking into account the value of $L_1 = L/(h_r/h_s + 1)$, it can be shown that in Eq. (29) the upper limits of integration are the same: $2h_r L_1/L = 2h_s(1 - L_1/L) = 2h_s h_r/(h_s + h_r) = h$. Here, these limits are denoted by h. In the geometry in Fig. 1 and in the approximation of near-grazing propagation,

h is a maximum separation between the direct ray from source to receiver and that reflected from the ground. The use of h allows us to simplify Eq. (29):

$$C=\exp\left\{-\frac{k^2L}{4h}\int_0^h[b_{\rm eff}(0,0)-b_{\rm eff}(0,z)]dz\right\}. \tag{30}$$

Substituting $b_{\rm eff}$ from the right-hand side of Eq. (26) into Eq. (30), one obtains an expression for C in terms of $\Phi_{\rm eff}$:

$$C=\exp\left\{-\frac{\pi k^2L}{2h}\int_0^h dz\int d^2K_\perp\,\Phi_{\rm eff}(0,\mathbf{K}_\perp)[1-e^{iK_zz}]\right\}. \tag{31}$$

Accounting for the fact that $\Phi_{\rm eff}(0,\mathbf{K}_\perp)$ is an even function and calculating the integral over z yields

$$C=\exp\left\{-\frac{\pi k^2L}{2}\int d^2K_\perp\,\Phi_{\rm eff}(0,\mathbf{K}_\perp)\left[1-\frac{\sin(K_zh)}{K_zh}\right]\right\}. \tag{32}$$

This formula for C coincides with Eqs. (14) and (15) from Ref. 13 obtained in that reference by a different approach.

We now show that Eq. (30) is closely related to the coherence function of a spherical sound wave for line-of-sight propagation. A formula for the coherence function of a spherical electromagnetic wave for line-of-sight propagation is well known, e.g., Ref. 18. An approach presented in Ref. 4 allows us to generalize this formula for the case of a spherical sound wave propagating in a turbulent atmosphere with temperature and velocity fluctuations:

$$\begin{aligned}\Gamma(\mathbf{r};\mathbf{r}_+)&=\langle p(x,\mathbf{r}')p^*(x,\mathbf{r}'')\rangle\\&=\frac{\exp(ik\mathbf{r}\cdot\mathbf{r}_+/L)}{L^2}\\&\quad\times\exp\left\{-\frac{k^2}{4}\int_0^L[b_{\rm eff}(0)-b_{\rm eff}(\mathbf{r}\xi/L)]d\xi\right\}.\end{aligned} \tag{33}$$

Here, $\mathbf{r}=\mathbf{r}'-\mathbf{r}''$, and $\mathbf{r}_+=(\mathbf{r}'+\mathbf{r}'')/2$. We denote by $\hat{\Gamma}$ a normalized coherence function: $\hat{\Gamma}(\mathbf{r})=\Gamma(\mathbf{r};\mathbf{r}_+)/\Gamma^{(0)}(\mathbf{r};\mathbf{r}_+)$, where $\Gamma^{(0)}$ is the coherence function in a medium without random inhomogeneities. It follows from Eq. (33) that $\hat{\Gamma}$ is given by the second exponent on the right-hand side of this equation and does not depend on $\mathbf{r}_+$. Let $\mathbf{r}=(0,h)$, i.e., the two points of observation $\mathbf{r}'$ and $\mathbf{r}''$ are located on a vertical line at a distance h. Then, a comparison of Eq. (30) and the second exponent in Eq. (33) reveals that

$$C=\hat{\Gamma}(0,h)\equiv\frac{\Gamma(0,h;\mathbf{r}_+)}{\Gamma^{(0)}(0,h;\mathbf{r}_+)}. \tag{34}$$

Thus, the coherence factor C is equal to the normalized coherence function of a spherical sound wave for line-of-sight propagation.

Equations (30)–(34) are valid for anisotropic turbulence. Let us now consider the case of isotropic turbulence. In this case, $\Phi_{\rm eff}(0,\mathbf{K}_\perp)$ is given by Ref. 4:

$$\Phi_{\rm eff}(0,\mathbf{K}_\perp)=\frac{\Phi_T(K_\perp)}{T_0^2}+\frac{4F(K_\perp)}{c_0^2}, \tag{35}$$

where $F(K_\perp)$ is the three-dimensional spectral density of velocity fluctuations related to Φ_{xx} by $\Phi_{xx}(\mathbf{K})=(1-K_x^2/K^2)F(K)$. Substituting Eq. (35) into Eq. (31) and calculating the integral over the azimuthal directions of the vector $\mathbf{K}_\perp$, one obtains

$$C=\exp\left\{-\pi^2k^2L\int_0^1 dt\int_0^\infty\left[\frac{\Phi_T(K)}{T_0^2}+\frac{4F(K)}{c_0^2}\right]\times[1-J_0(hKt)]KdK\right\}, \tag{36}$$

where J_0 is the Bessel function of zero the order.

The relationship (34) holds for isotropic turbulence also. For such turbulence, the coherence function $\Gamma(\mathbf{r};\mathbf{r}_+)$ is presented in Ref. 4 for the Kolmogorov, Gaussian, and von Karman spectra of temperature and velocity fluctuations. These spectra have been most widely used in atmospheric acoustics. Equation (34) and formulas for $\Gamma(\mathbf{r};\mathbf{r}_+)$ from Ref. 4 allow us to calculate the coherence factor C for these spectra. For the Kolmogorov spectra of temperature and velocity fluctuations, $\Phi_T(K)$ and $F(K)$ are given by[3,4]

$$\Phi_T^{\rm K}(K)=AC_T^2K^{-11/3},\quad F^{\rm K}(K)=(11/6)AC_v^2K^{-11/3}. \tag{37}$$

Here, the superscript ''K'' refers to the Kolmogorov spectrum, $A\approx0.0330$, and C_T^2 and C_v^2 are the structure parameters of temperature and velocity fluctuations. The Kolmogorov spectrum is a realistic spectrum of isotropic turbulence in the inertial subrange (e.g., Sec. 6.2.3 in Ref. 4). It does not describe a spectrum of turbulence due to eddies larger than those in the inertial subrange. For the Kolmogorov spectrum, the coherence factor C can be found with the use of Eq. (34) and Eq. (7.83) for $\Gamma(\mathbf{r};\mathbf{r}_+)$ from Ref. 4:

$$C^K=\exp\left\{-k^2h^{5/3}L\frac{3BC_T^2}{8T_0^2}-k^2h^{5/3}L\frac{11BC_v^2}{4c_0^2}\right\}, \tag{38}$$

where the coefficient $B\approx0.364$.

For the Gaussian spectra of temperature and velocity fluctuations, $\Phi_T(K)$ and $F(K)$ are given by[4]

$$\Phi_T^{\rm G}(K)=\frac{\sigma_T^2l^3}{8\pi^{3/2}}e^{-K^2l^2/4},\quad F^{\rm G}(K)=\frac{\sigma_v^2K^2l^5}{32\pi^{3/2}}e^{-K^2l^2/4}, \tag{39}$$

where superscript ''G'' refers to the Gaussian spectrum, σ_T^2 and σ_v^2 are the variances of temperature and velocity fluctuations, and l is the scale of random inhomogeneities. For this spectrum, the coherence factor C can be obtained by making use of Eq. (34) and Eq. (7.107) for $\Gamma(\mathbf{r};\mathbf{r}_+)$ from Ref. 4:

$$C^{\rm G}=\exp\left\{-2\gamma_T^{\rm G}L\left[1-\frac{\sqrt{\pi}l}{2h}\,{\rm erf}\left(\frac{h}{l}\right)\right]-2\gamma_v^{\rm G}L\left[1-\frac{\sqrt{\pi}l}{4h}\,{\rm erf}\left(\frac{h}{l}\right)-\frac{1}{2}e^{-h^2/l^2}\right]\right\}. \tag{40}$$

Here, $\gamma_T^{\rm G}=\sqrt{\pi}k^2\sigma_T^2l/(8T_0^2)$ is the extinction (attenuation) coefficient of the mean sound field due to temperature fluctuations with the Gaussian spectrum, $\gamma_v^{\rm G}=\sqrt{\pi}k^2\sigma_v^2l/(2c_0^2)$ is that due to velocity fluctuations, and erf(h/l) is the error func-

tion. If $\sigma_v^2=0$, Eq. (40) coincides with the corresponding equation in Ref. 1.

For the von Karman spectra of temperature and velocity fluctuations, $\Phi_T(K)$ and $F(K)$ are given by[4]

$$\Phi_T^{\mathrm{vK}}(K)=AC_T^2(K^2+K_0^2)^{-11/6},$$
$$F^{\mathrm{vK}}(K)=\tfrac{11}{6}AC_v^2K^2(K^2+K_0^2)^{-17/6}. \tag{41}$$

Here, the superscript ''vK'' refers to the von Karman spectrum, and K_0 is the wavenumber which is inversely proportional to the outer scale of turbulence. The coherence factor C can be obtained with the use of Eq. (34) and Eq. (7.114) for $\Gamma(\mathbf{r};\mathbf{r}_+)$ from Ref. 4:

$$C^{\mathrm{vK}}=\exp\left\{-\frac{2L}{K_0h}\int_0^{K_0h}dt\left[\gamma_T^{\mathrm{vK}}\left[1-\frac{2^{1/6}t^{5/6}}{\Gamma(5/6)}K_{5/6}(t)\right]\right.\right.$$
$$\left.\left.+\gamma_v^{\mathrm{vK}}\left[1-\frac{2^{1/6}t^{5/6}}{\Gamma(5/6)}\left(K_{5/6}(t)-\frac{t}{2}K_{1/6}(t)\right)\right]\right]\right\}. \tag{42}$$

Here, $K_\nu(t)$ is the modified Bessel function, $\gamma_T^{\mathrm{vK}}=3\pi^2Ak^2K_0^{-5/3}C_T^2/(10T_0^2)$ is the extinction coefficient of the mean sound field due to temperature fluctuations with the von Karman spectrum, and $\gamma_v^{\mathrm{vK}}=6\pi^2Ak^2K_0^{-5/3}C_v^2/(5c_0^2)$ is that due to velocity fluctuations.

At the end of this section, let us consider the range of applicability of Eqs. (30)–(32) and (36). These equations have been derived for the case of geometrical acoustics ($D\ll1$), and for the case of Fraunhofer diffraction ($D\gg1$) if

$$h_0>\sqrt{L/k},\quad D\ll h_m/l. \tag{43}$$

Two facts indicate that Eqs. (30)–(32) and (36) are valid even if inequalities (43) are not fulfilled. First, the results obtained in Ref. 13 show that the range of applicability of Eq. (32) does not require validity of these inequalities. Second, Ref. 21 considers the interference of the direct and ground-reflected electromagnetic waves in an atmosphere with fluctuations ϵ in the dielectric permittivity. The geometry of the problem in Ref. 21 is similar to that in Fig. 1. Starting from the parabolic equation for the electromagnetic field E, using the Feynman path-integral technique, and assuming that the ground is perfectly reflecting ($Q=-1$), the following formula for the mean squared field $\langle|E|^2\rangle$ is obtained in Ref. 21:

$$\langle|E|^2\rangle=\frac{2}{L^2}\left[1-\exp\left(-\frac{\pi k^2L}{4}\int_0^1 d\xi\,H(h\xi)\right)\right.$$
$$\left.\times\cos(2kh_sh_r/L)\right], \tag{44}$$

where

$$H(\mathbf{r})=2\int d^2K_\perp\Phi_\epsilon(0,\mathbf{K}_\perp)[1-e^{i\mathbf{K}_\perp\cdot\mathbf{r}}]. \tag{45}$$

Here, $\Phi_\epsilon(K_x,\mathbf{K}_\perp)$ is the three-dimensional spectral density of ϵ. The range of applicability of Eqs. (44) and (45) is that of the Rytov method. Note that inequalities (43) have not been used when deriving Eqs. (44) and (45).

According to Ref. 4, Eqs. (44) and (45) can be used to obtain a formula for $\langle|p|^2\rangle$ for the case of sound propagation near the ground in the atmosphere with temperature and velocity fluctuations. The formula for $\langle|p|^2\rangle$ is given by the right-hand side of Eq. (44) if $\Phi_\epsilon(0,\mathbf{K}_\perp)$ is replaced by $\Phi_{\mathrm{eff}}(0,\mathbf{K}_\perp)$. This formula coincides with Eq. (10), where C is given by Eq. (36) if in Eq. (10) $|Q|=1$, $\Omega=\pi$, $\langle\chi_sS_i\rangle-\langle\chi_iS_s\rangle=0$, $R_s=R_i=L$, and $R_i-R_s=2h_sh_r/L$. The latter two equalities are always valid in the approximation of near-grazing propagation. Therefore, if $Q=-1$, Eqs. (44) and (45) derived in Ref. 21 lead to the same formula for $\langle|p|^2\rangle$ that is obtained in the present paper.

Thus, we can conclude that the range of applicability of Eqs. (30)–(32) and (36) is not limited by inequalities (43), and is the same as the range of applicability of the Rytov method, which is known in the literature (e.g., Ref. 17). The latter range includes a condition that amplitude fluctuations in the direct wave are not saturated (i.e., $\langle\chi_s^2\rangle\ll1$).

IV. NUMERICAL AND EXPERIMENTAL RESULTS

A. Isotropic turbulence

In this and next subsections, Eqs. (10), (38), (40), and (42) are used to numerically study the effects of atmospheric turbulence on the relative sound pressure level W, which is related to $\langle|p|^2\rangle$ by

$$W=10\log(\langle|p|^2\rangle/p_0^2), \tag{46}$$

where $p_0=1/r_0$ is the amplitude of the sound pressure due to the source at a distance $r_0=1$ m.

In the numerical examples below, the source and receiver heights above the ground are $h_s=6$ m and $h_r=3$ m, respectively; the horizontal distance between them is $L=130$ m; the adiabatic sound speed $c_0=340$ m/s; and the ground is acoustically hard, i.e., $|Q|=1$ and $\Omega=0$. We denote by W_0 the relative sound pressure level in the atmosphere without turbulence. Here W_0 is given by Eqs. (46) and (10) with $C=1$. Figure 3 shows the dependence of W_0 on the sound frequency f. The interference of the direct and ground-reflected waves results in maxima and minima in the dependence of W_0 on f.

The solid line in Fig. 4 reproduces the first minimum of W_0 in a more narrow frequency range from 595 to 635 Hz. The dotted, dash–dotted, and dashed lines correspond to the relative sound pressure levels W_v^{K}, W_v^{G}, and W_v^{vK} in a turbulent atmosphere having velocity fluctuations only ($\tilde{T}=0$), with the Kolmogorov, Gaussian, and von Karman spectra, respectively. Here W_v^{K}, W_v^{G}, and W_v^{vK} are given by Eqs. (46) and (10), where C is given by Eqs. (38), (40), and (42), respectively. In Fig. 4, $l=1$ m and $C_v^2=8\times10^{-8}c_0^2\,\mathrm{m}^{-2/3}$. Furthermore, the values of K_0 and σ_v^2 are chosen from the following relationships:

$$K_0=\frac{2\Gamma(5/6)}{\Gamma(1/3)l},\quad \sigma_v^2=\frac{\Gamma^2(1/3)}{\pi2^{4/3}\sqrt{3}}C_v^2K_0^{-2/3}, \tag{47}$$

where Γ is the gamma-function. These relationships between the parameters of the Gaussian and von Karman spectra of velocity fluctuations are used to compare the statistical moments of a sound field for these spectra, see Eqs. (6.49) and (6.51) from Ref. 4. If the relationships (47) are valid, the

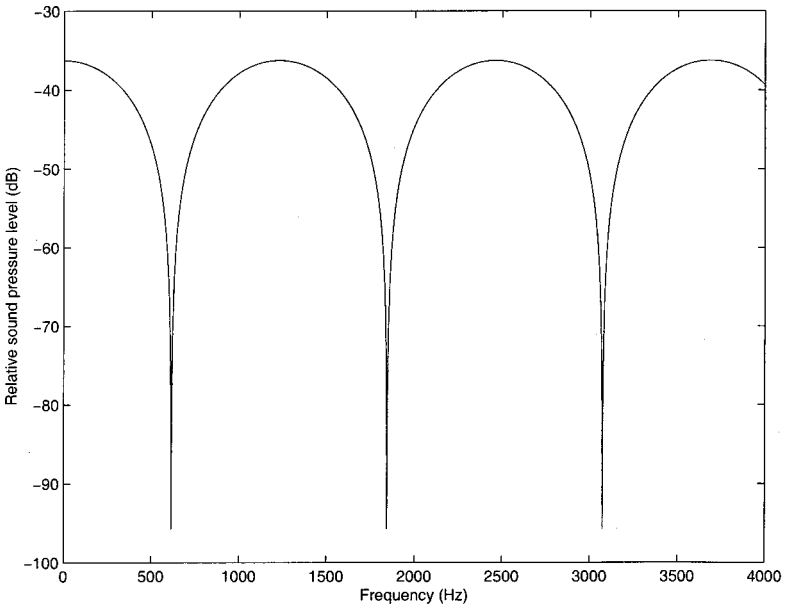


FIG. 3. Relative sound pressure level W_0 versus the frequency f in the atmosphere without turbulence.

integral lengths of the von Karman and Gaussian spectra of velocity fluctuations are equal, and the variances of these spectra are the same. Furthermore, $\gamma_v^{\mathrm{vK}}=\gamma_v^{\mathrm{G}}$ if Eq. (47) is valid.

It follows from Fig. 4 that the relative sound pressure levels for the Gaussian and von Karman spectra, W_v^{G} and W_v^{vK}, are very close to each other. On the other hand, the relative sound pressure level for the Kolmogorov spectrum, W_v^{K}, significantly differs from W_v^{G} and W_v^{vK}. This can be explained by using results obtained in Ref. 22. It follows from this reference that, if $\gamma_v^{\mathrm{vK}}L=\gamma_v^{\mathrm{G}}L<0.5$, the coherence function Γ_v^{vK} of a spherical sound wave for line-of-sight propagation in the atmosphere having velocity fluctuations with the von Karman spectrum nearly coincides with the coherence function Γ_v^{G} for the case of Gaussian spectrum and significantly differs from the coherence function Γ_v^{K} for the Kolmogorov spectrum. On the other hand, if $\gamma_v^{\mathrm{vK}}L=\gamma_v^{\mathrm{G}}L$ is greater than 100, Γ_v^{vK} nearly coincides with Γ_v^{K} and significantly differs from Γ_v^{G}. For the values of the parameters in Fig. 4, $\gamma_v^{\mathrm{vK}}L=\gamma_v^{\mathrm{G}}L\sim 7\times 10^{-4}$. Therefore, in this case $\Gamma_v^{\mathrm{vK}}\sim\Gamma_v^{\mathrm{G}}$, and Γ_v^{vK} significantly differs from Γ_v^{K}. Since the coherence factor C and the coherence function Γ are related by Eq. (34), one can conclude that $W_v^{\mathrm{vK}}\sim W_v^{\mathrm{G}}$, and that W_v^{vK} significantly differs from W_v^{K}. This explains Fig. 4.

Considerations above allow us also to conclude that, if $\gamma_v^{\mathrm{vK}}L=\gamma_v^{\mathrm{G}}L>100$, the coherence factor C for the von Karman spectrum is nearly the same as that for the Kolmogorov spectrum, and differs significantly from C for the Gaussian spectrum. However, in this case the coherence factors C for these spectra are much less than 1 since γ_v^{G} and γ_v^{vK} are in the exponents in Eqs. (40) and (42). Therefore, $W_v^{\mathrm{K}}\approx W_v^{\mathrm{G}}\approx W_v^{\mathrm{vK}}$ because the third term (which is proportional to C) on the right-hand side of Eq. (10) can be ignored.

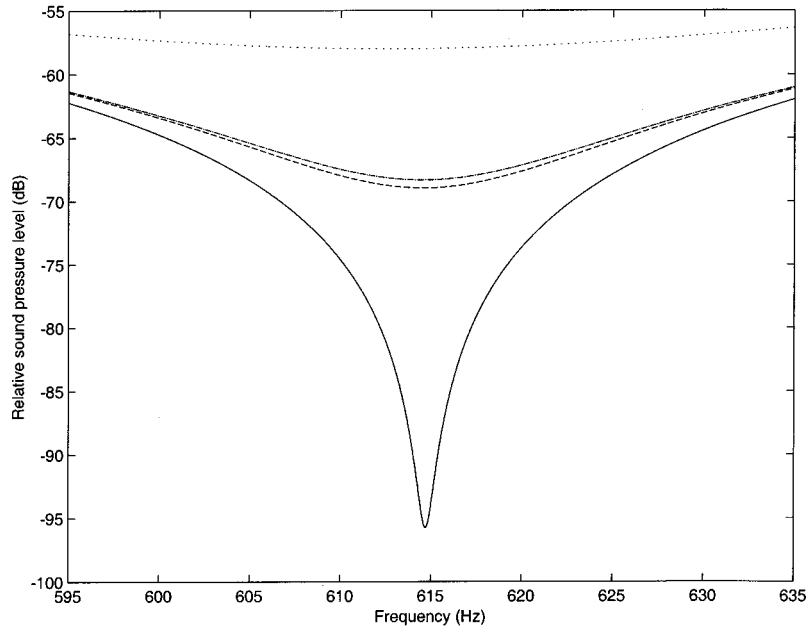


FIG. 4. Relative sound pressure level versus the frequency f. Solid, dashed, dash–dotted, and dotted lines correspond to the cases of atmosphere without turbulence, and with von Karman, Gaussian, and Kolmogorov spectra of velocity fluctuations.

Let us now study the effects of temperature and velocity contributions to the relative sound pressure levels W^{K}, W^{G}, and W^{vK} for the Kolmogorov, Gaussian, and von Karman spectra. Equations (38), (40), and (42) clearly show that these contributions to the coherence factor C and, hence, to W are different. Numerical calculations show that the maximum difference between the sound pressure levels W^{K}, W^{G}, and W^{vK} in the atmosphere with temperature fluctuations only and those in the atmosphere with velocity fluctuations only is a few dB. When comparing temperature and velocity contributions to W^{K}, W^{G}, and W^{vK}, we assumed that $C_T^2/T_0^2 = 4C_v^2/c_0^2$. This relationship is employed to compare the relative contributions from temperature and velocity fluctuations to the statistical moments of a sound field (e.g., Ref. 4). If this relationship were valid, the temperature and velocity contributions to the structure parameter of acoustic refractive index fluctuations would be the same.

B. Anisotropic turbulence

Now let us consider the effects of turbulent anisotropy on the interference of the direct and ground-reflected waves. For anisotropic turbulence, it is convenient to express the velocity component along the x-axis in the following form: $v_x(\mathbf{R}) = v_1(\mathbf{R})\cos\theta + v_2(\mathbf{R})\sin\theta$. Here v_1 is the component of $\mathbf{v}(\mathbf{R})$ in the direction of the mean wind, v_2 is the component crosswind, and θ is the angle between the horizontal component of the mean wind velocity vector and the x-axis. In this case, ϵ_{mov} is reexpressed as $\epsilon_{\mathrm{mov}} = -\tilde{T}/T_0 - 2(v_1\cos\theta + v_2\sin\theta)/c_0$. Using this formula, the following expression for $\Phi_{\mathrm{eff}}(\mathbf{K})$ can be obtained:

$$\Phi_{\mathrm{eff}}(\mathbf{K}) = \frac{\Phi_T(\mathbf{K})}{T_0^2} + \frac{4}{c_0^2}[\Phi_{11}(\mathbf{K})\cos^2\theta + \Phi_{12}(\mathbf{K})\sin(2\theta) + \Phi_{22}(\mathbf{K})\sin^2\theta], \tag{48}$$

and likewise for the correlation function $B_{\mathrm{eff}}(\mathbf{R})$. In Eq. (48), Φ_{11}, Φ_{12}, and Φ_{22} are the three-dimensional spectral densities of the correlation functions $\langle v_1(\mathbf{R})v_1(\mathbf{R}')\rangle$, $\langle v_1(\mathbf{R})v_2(\mathbf{R}')\rangle$, and $\langle v_2(\mathbf{R})v_2(\mathbf{R}')\rangle$, respectively.

We will use Mann's anisotropic rapid-distortion model[23] for the turbulent velocity field. This model was derived from an isotropic von Karman spectrum by calculating distortions to the turbulence resulting from a constant vertical wind shear. Since the spectral equations for Mann's model are rather complicated and have been given elsewhere,[7,23] they will not be repeated here. The geometry of the problem is the same as in Figs. 3 and 4. The application of Mann's model in this paper is the same as in Wilson:[7] the dimensionless shear parameter, Γ (not to be confused with the coherence function), is set to 3.58; the variance parameter is $\sigma_M^2 = 1.52u_*^2$ (where u_* is the friction velocity); and the length scale is $l_M = 0.587z$. Because the length scale in Mann's model is height-dependent, we are forced to assign z to a representative value in order to calculate the length scale. Therefore we somewhat arbitrarily set $z = 4$ m, resulting in $l_M = 2.35$ m. The friction velocity can be determined from the velocity variance. Based on Eq. (47), the variance for our example is $\sigma_v^2 = 0.00542\ \mathrm{m}^2/\mathrm{s}^2$. Taking $\sigma_v^2 = 2.97u_*^2$ in the isotropic von Karman model,[7] one has $u_* = 0.0427$ m/s. (The variance σ_v^2 and friction velocity u_* in this example are characteristic of very low wind atmospheric conditions.)

Calculations for along mean wind and crosswind propagation using the Mann model are shown in Fig. 5. Also shown for comparison is the isotropic von Karman model, with $\sigma_v^2 = 2.97u_*^2$ and $K_0^{-1} = 1.79z = 7.16$ m,[7] where K_0 is a parameter in Eq. (41). The effect of the turbulence on filling in the interference minimum is about 4 dB larger in the along-wind than in the crosswind direction. The prediction for the isotropic von Karman model falls in between. This behavior results from the higher variance and extended integral length scale in the along-wind direction characteristic of a shear layer.

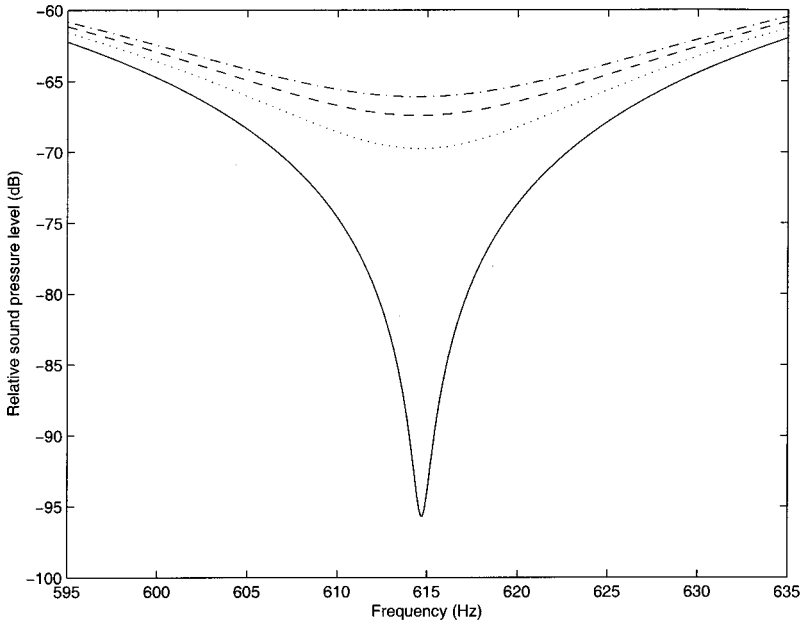


FIG. 5. Relative sound pressure level versus the frequency f. Solid, dashed, dash–dotted, and dotted lines correspond to the cases of the atmosphere without turbulence, and isotropic von Karman, Mann along-wind, and Mann crosswind spectra of velocity fluctuations.

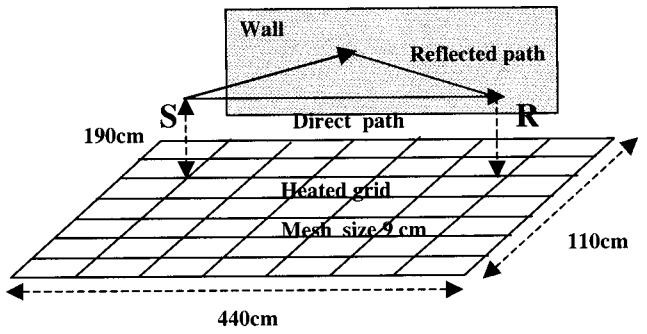


FIG. 6. The geometry of the laboratory experiment. S is a point monochromatic source, and R is a receiver (microphone).

C. Comparison between theory and experiment

Here, we compare mean squared sound pressure $\langle|p|^2\rangle$ calculated for the von Karman spectrum of temperature fluctuations [see Eqs. (10) and (42)] with experimental data obtained in a large anechoic chamber in Ecole Centrale de Lyon.[24]

In the experiment, temperature fluctuations were created by a heated grid located horizontally (see Fig. 6). The size of the grid was 440×110 cm and the size of a mesh in the grid was 9 cm. A spectrum of temperature fluctuations above the grid is approximated very well by the von Karman spectrum $\Phi_T^{\mathrm{vK}}(K)$ given by Eq. (41). The parameters K_0 and C_T^2 of this spectrum, measured in the experiment, were as follows: $K_0=6.8\,\mathrm{m}^{-1}$ and $C_T^2/T_0^2=3.36\times10^{-4}\,\mathrm{m}^{-2/3}$. The value of C_T^2 is normalized by T_0^2 since γ_T^{vK} in Eq. (42) for the coherence factor C^{vK} is proportional to C_T^2/T_0^2.

An acoustically hard wall was vertically placed above the grid. A point monochromatic source and a microphone were also placed above the grid at the distances h_s and h_r from the wall. For this geometry, two waves are impinging on the receiver: the direct wave and that reflected from the wall. In the experiment, time-mean squared pressure $\langle|p|^2\rangle$ was measured as a function of the distance L between the source and microphone. The distance h_s was 10 cm; L varied from 3 cm to 3 m; and the measurements were done for $h_r=7$ and 10 cm, and $f=23.5$, 31, 40, and 75 kHz. For this geometry, the values of the wave parameter D and the variances of log-amplitude and phase fluctuations $\langle\chi_s^2\rangle$ and $\langle S_s^2\rangle$ are close to those for sound propagation in the atmosphere. Furthermore, the relationship $L\gg K_0^{-1}\gg k^{-1}$ between parameters of the experiment usually holds for outdoor sound propagation also. Therefore, the laboratory experiment in question models sound propagation in a turbulent atmosphere.

In the experiment, a reference microphone was located at a distance 2 cm from the source. Using the measured values of $\langle|p|^2\rangle$ and the pressure p_0 at the reference microphone, the relative sound pressure level W was calculated with the use of Eq. (46). Then, the dependence of W versus L was plotted for fixed values of h_r and f. One of these dependences, obtained for $h_r=7$ cm and $f=40$ kHz, is shown in Fig. 7 by the ''+'' symbol. Interference maxima and minima are clearly seen in the figure as well as a decrease of W with an increase of L due to spherical spreading.

In Fig. 7, the solid line is the theoretical predictions of the relative sound pressure level W for the von Karman spectrum of temperature fluctuations. W was calculated with the use of Eqs. (10), (42), and (46), where the values of parameters were taken from the experiment: $Q=1$, $f=40$ kHz, $h_s=10$ cm, $h_r=7$ cm, $K_0=6.8\,\mathrm{m}^{-1}$, and $C_T^2/T_0^2=3.36\times10^{-4}\,\mathrm{m}^{-2/3}$. Figure 7 shows good agreement between theoretical predictions and experimental data.

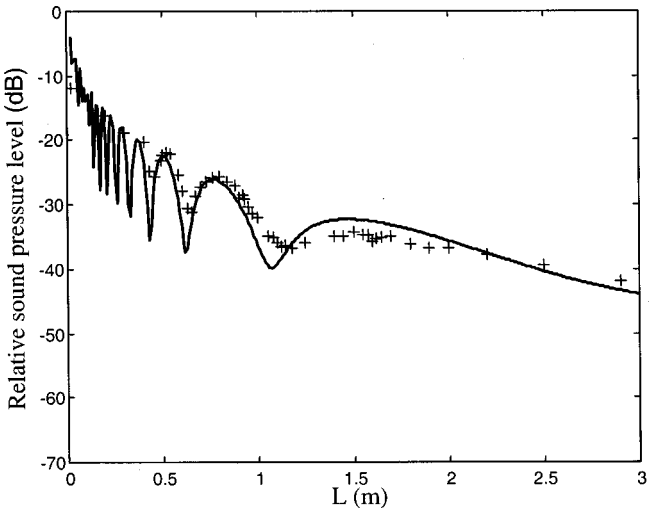


FIG. 7. Relative sound pressure level W versus the distance L between the source and the receiver for $f=40$ kHz, $h_s=10$ cm, and $h_r=7$ cm. The solid line corresponds to theoretical predictions; symbols ''+'' correspond to experimental data obtained in the laboratory experiment.

For other values of h_r and f, theoretical predictions and experimental data also agree well. A more detailed description of the experimental setup and comparison between data and theoretical predictions is present in Refs. 6 and 25.

V. CONCLUSIONS

We have studied sound propagation near the impedance ground in the atmosphere with temperature and velocity fluctuations when a sound field p is the sum of the direct and ground-reflected waves; see Fig. 1. We have shown that, when calculating p, the geometry of Fig. 1 can be replaced by a new geometry shown in Fig. 2. In this new geometry, the ground is removed, a random medium fills the whole space, and the source is supplemented by an image source. The sound field p at the receiver is the sum of the sound field emitted by the source, and that emitted by the image source. We gave a mathematical formulation of the geometry in Fig. 2, which is based on the parabolic equation (3) for p, initial condition (5), and Eq. (4) which determines the random inhomogeneities in a medium. Equations (3), (4), and (5) for calculating p are the first important result obtained in the paper.

We used the Rytov method to solve the parabolic equation (3) and obtained expressions for the complex phases ψ_s and ψ_i of the waves emitted by the source and the image source. We also derived Eq. (10) for the mean squared sound pressure $\langle|p|^2\rangle$. This equation contains the coherence factor C which characterizes the coherence between the waves emitted by the source and the image source. Using expressions for ψ_s and ψ_i, we obtained Eq. (29) for C. Equations

(10) and (29) for $\langle|p|^2\rangle$ and C, respectively, generalize those obtained by Clifford and Lataitis[1] for the case of both temperature and velocity fluctuations, and for the case of anisotropic spectra of these fluctuations. Furthermore, we have shown that C is given by Eq. (29) for all values of the wave parameter D if the Rytov approximation is valid. This generalization is the second important result of the present paper.

Then, we showed that the coherence factor C is equal to the normalized coherence function $\hat{\Gamma}$ of a spherical sound wave for line-of-sight propagation. For the case of isotropic turbulence, the coherence function has been recently calculated for the Kolmogorov, Gaussian, and von Karman spectra of temperature and velocity fluctuations.[4] This allowed us to obtain analytical expressions (38), (40), and (42) for C for these spectra. This is the third important result of the paper.

Using the obtained expressions for C, we numerically compared the relative sound pressure levels W for the Kolmogorov, Gaussian, and von Karman spectra of velocity fluctuations. Then, we numerically studied the relative contributions from temperature and velocity fluctuations to W for the Kolmogorov, Gaussian, and von Karman spectra of these fluctuations. Also, the effects of turbulent anisotropy on the interference of direct and ground-reflected waves was numerically studied. Finally, we showed that the theoretical predictions of the relative sound pressure level for the von Karman spectrum of temperature fluctuations are in a good agreement with experimental data obtained in a laboratory experiment.

ACKNOWLEDGMENTS

This material is partly based upon work that was supported by the U.S. Army Research Office under Contract No. DAAG55-98-1-0463, and NRC-ETL Research Associateship.

APPENDIX A: EQUIVALENCE OF GEOMETRIES IN FIGS. 1 AND 2

In this appendix, we show the equivalence of the geometries in Figs. 1 and 2 for the case $Q=1$. In the analysis below, we assume that a source is located at the point $\mathbf{R}_0=(x_0,y_0,z_0)$ which may not coincide with the origin of the coordinate system. The receiver is located at the point $\mathbf{R}=(x,y,z)$.

Let us derive an integral equation for the sound pressure $p(\mathbf{R},\mathbf{R}_0)$ due to the point source for the geometry in Fig. 1. If there are no random inhomogeneities in the atmosphere, $p(\mathbf{R},\mathbf{R}_0)$ is equal to the Green's function $G(\mathbf{R},\mathbf{R}_0)$ of the problem, given by

$$G(\mathbf{R},\mathbf{R}_0)=G_0(\mathbf{R}-\mathbf{R}_0)+Q(\mathbf{R},\mathbf{R}_0)G_0(\mathbf{R}-\mathbf{R}_0'). \tag{A1}$$

Here, $G_0(\mathbf{R})=R^{-1}e^{ikR}$ is the Green's function in a free space, and $\mathbf{R}_0'=(x_0,y_0,-z_0-2h_s)$ is a point symmetrical to $\mathbf{R}_0=(x_0,y_0,z_0)$ with respect to the ground surface $z=-h_s$. In Eq. (A1), we take into account that Q depends on the positions of the source and the receiver. Equation (A1) is a generalization of Eq. (1) for the case of arbitrary positions of the source and the receiver.

If there are fluctuations in the acoustic refractive index in the atmosphere, given by $\epsilon_{\mathrm{mov}}/2$, the sound field $p(\mathbf{R},\mathbf{R}_0)$ satisfies the following integral equation:

$$p(\mathbf{R},\mathbf{R}_0)=G(\mathbf{R},\mathbf{R}_0)+k^2\int_{-\infty}^{\infty}dx_1\int_{-\infty}^{\infty}dy_1 \times\int_{-h_s}^{\infty}dz_1\,\epsilon_{\mathrm{mov}}(\mathbf{R}_1)G(\mathbf{R},\mathbf{R}_1)p(\mathbf{R}_1,\mathbf{R}_0). \tag{A2}$$

To show that Eq. (A2) correctly describes sound propagation in the turbulent atmosphere near the impedance ground, one can solve this equation by iterations, replacing $p(\mathbf{R}_1,\mathbf{R}_0)$ in the integrand by the right-hand side of this equation. The resulting Born series describes a process of multiple scattering of p in a turbulent atmosphere near the ground.

For the geometry in Fig. 2, an integral equation for the sound pressure $p(\mathbf{R},\mathbf{R}_0)$ can be derived analogously to the derivation of Eq. (A2):

$$p(\mathbf{R},\mathbf{R}_0)=G(\mathbf{R},\mathbf{R}_0)+k^2\int_{-\infty}^{\infty}dx_1\int_{-\infty}^{\infty}dy_1 \times\int_{-\infty}^{\infty}dz_1\,\mu(\mathbf{R}_1)G_0(\mathbf{R}-\mathbf{R}_1)p(\mathbf{R}_1,\mathbf{R}_0). \tag{A3}$$

We point out some similarities and differences between Eqs. (A2) and (A3). If $\epsilon_{\mathrm{mov}}=0$, it follows from Eq. (A3) that $p=G(\mathbf{R},\mathbf{R}_0)$, as it should. Since there is no boundary for the geometry in Fig. 2, the Green's function is equal to $G_0(\mathbf{R}-\mathbf{R}_1)$ rather than to $G(\mathbf{R},\mathbf{R}_0)$; G_0 is in the integrand in Eq. (A3). The integration over z_1 in Eq. (A3) is from $-\infty$ to ∞, while that in Eq. (A2) is from $-h_s$ to ∞.

We now show that Eqs. (A2) and (A3) are equivalent. We denote by I_z the integral over z_1 in Eq. (A3):

$$I_z=\int_{-\infty}^{\infty}dz_1\,\mu(\mathbf{R}_1)G_0(\mathbf{R}-\mathbf{R}_1)p(\mathbf{R}_1,\mathbf{R}_0). \tag{A4}$$

We split this integral into two: the first from $-h_s$ to ∞, the second from $-\infty$ to $-h_s$. In the latter integral, we change the integration variable from z_1 to $z_2=-z_1-h_s$. The limits of integration over z_2 are from $-h_s$ to ∞. Then, denoting $z_2=z_1$, one obtains

$$I_z=\int_{-h_s}^{\infty}dz_1[\mu_{\mathrm{mov}}(\mathbf{R}_1)G_0(\mathbf{R}-\mathbf{R}_1)p(\mathbf{R}_1,\mathbf{R}_0) +\mu_{\mathrm{mov}}(\mathbf{R}_1')G_0(\mathbf{R}-\mathbf{R}_1')p(\mathbf{R}_1',\mathbf{R}_0)], \tag{A5}$$

where $\mathbf{R}_1'=(x_1,y_1,-z_1-2h_s)$ is a point symmetrical to $\mathbf{R}_1=(x_1,y_1,z_1)$ with respect to the plane $z=-h_s$.

It can be shown that $p(\mathbf{R},\mathbf{R}_0)=p(\mathbf{R}',\mathbf{R}_0)$, where $\mathbf{R}'=(x,y,-z-2h_s)$ is a point of the image receiver (see Fig. 2), and $p(\mathbf{R}',\mathbf{R}_0)$ is the sound pressure at this point. Indeed, $p(\mathbf{R},\mathbf{R}_0)$ is the sum of the sound fields of two waves propagating along the line segments OR and $O'R$, while $p(\mathbf{R}',\mathbf{R}_0)$ is the sum of the sound fields of the waves propagating along $O'R'$ and OR'. Taking into account that μ is symmetrical with respect to the plane $z=-h_s$, it is clear that

the line segments OR and $O'R'$ pass through the same inhomogeneities; so do $O'R$ and OR'. This proves that $p(\mathbf{R},\mathbf{R}_0)=p(\mathbf{R}',\mathbf{R}_0)$. Therefore, in Eq. (A5), $p(\mathbf{R}_1,\mathbf{R}_0)=p(\mathbf{R}_1',\mathbf{R}_0)$. Furthermore, using Eq. (4), it is easy to show that in Eq. (A5), $\mu(\mathbf{R}_1)=\mu(\mathbf{R}_1')=\epsilon_{\mathrm{mov}}(\mathbf{R}_1)$.

Using these equalities in Eq. (A5) for I_z and substituting the value of I_z into Eq. (A3) yields

$$p(\mathbf{R},\mathbf{R}_0)=G(\mathbf{R},\mathbf{R}_0)+k^2\int_{-\infty}^{\infty}dx_1\int_{-\infty}^{\infty}dy_1 \times\int_{-h_s}^{\infty}dz_1\,\epsilon_{\mathrm{mov}}(\mathbf{R}_1)p(\mathbf{R}_1,\mathbf{R}_0) \times[G_0(\mathbf{R}-\mathbf{R}_1)+G_0(\mathbf{R}-\mathbf{R}_1')]. \tag{A6}$$

This equation is equivalent to Eq. (A2) if $Q=1$ and differs from it if $Q\neq 1$.

In the geometry of Fig. 2, $p(\mathbf{R},\mathbf{R}_0)$ is the sum of the sound fields propagating along the line segments OR and $O'R$. The latter line segment passes through the same inhomogeneities as the line segment OR' from the source to the image receiver. Therefore, one can conclude that $p(\mathbf{R},\mathbf{R}_0)$ is equal to the sound fields along the line segments OR and OR'. In other words, $p(\mathbf{R},\mathbf{R}_0)$ can be represented as the sum of the sound fields at the receiver and image receiver, which are emitted by the source. This result is used in Ref. 13.

APPENDIX B: CALCULATION OF THE COHERENCE FACTOR *C*

In Appendix B, expressions for $\overline{\chi_s}$, $\overline{\chi_i}$, $\overline{S_s}$, and $\overline{S_i}$ are obtained, which are then used in deriving a formula for the coherence factor C.

The random fields $\psi_s=\chi_s+iS_s$ and $\psi_i=\chi_i+iS_i$ have Gaussian destributions. We remind the reader of a well-known formula (e.g., Refs. 1, 17):

$$\langle\exp(\xi)\rangle=\exp(\tfrac{1}{2}\langle(\xi-\overline{\xi})^2\rangle+\overline{\xi}), \tag{B1}$$

which is valid for a random field ξ having a Gaussian distribution. This formula allows one to express the mean value of e^{ξ} in terms of the mean value and variance of the random field ξ.

The equality $\langle e^{\psi_s+\psi_s^*}\rangle=1$, which is used to derive Eq. (10), can also be written as $\langle e^{2\chi_s}\rangle=1$. Using Eq. (B1), the latter equality can be rewritten in the following form:

$$\exp(2\langle\chi_s^2\rangle-2(\overline{\chi_s})^2+2\overline{\chi_s})=1. \tag{B2}$$

This equation is equivalent to the following equation: $\langle\chi_s^2\rangle-(\overline{\chi_s})^2+\overline{\chi_s}=0$. It follows from the latter equation that

$$\overline{\chi_s}=-\langle\chi_s^2\rangle+O(\langle\epsilon_{\mathrm{mov}}^4\rangle). \tag{B3}$$

Let us now derive a formula for the mean value of phase fluctuations $\overline{S_s}$. Consider the mean sound field of the wave emitted by the source:

$$\langle p_s\rangle=\frac{e^{ikR_s}}{R_s}\langle e^{\chi_s+iS_s}\rangle. \tag{B4}$$

Using Eq. (B1), and taking into account that $(\overline{\chi_s})^2\sim O(\langle\epsilon_{\mathrm{mov}}^4\rangle)$ and $(\overline{S_s})^2\sim O(\langle\epsilon_{\mathrm{mov}}^4\rangle)$ [see Eq. (8)], one obtains from Eq. (B4):

$$\langle p_s\rangle=\frac{e^{ikR_s}}{R_s}\exp\Big(-\frac{1}{2}\langle\chi_s^2\rangle-\frac{1}{2}\langle S_s^2\rangle+i\langle\chi_sS_s\rangle +i\overline{S_s}+O(\langle\epsilon_{\mathrm{mov}}^4\rangle)\Big). \tag{B5}$$

We now compare this equation with Eq. (7.61) from Ref. 4 for the mean sound field calculated by a different approach:

$$\langle p_s\rangle=\frac{e^{ikR_s}}{R_s}\exp\Big(-\frac{1}{2}\langle\chi_s^2\rangle-\frac{1}{2}\langle S_s^2\rangle+O(\langle\epsilon_{\mathrm{mov}}^4\rangle)\Big). \tag{B6}$$

A comparison between Eqs. (B5) and (B6) reveals that

$$\overline{S_s}=-\langle\chi_sS_s\rangle+O(\langle\epsilon_{\mathrm{mov}}^4\rangle). \tag{B7}$$

This formula is also presented in Ref. 3, where it is obtained for a plane wave.

Formulas similar to Eqs. (B3) and (B7) are valid for the wave emitted by the image source

$$\overline{\chi_i}=-\langle\chi_i^2\rangle+O(\langle\epsilon_{\mathrm{mov}}^4\rangle),\quad \overline{S_i}=-\langle\chi_iS_i\rangle+O(\langle\epsilon_{\mathrm{mov}}^4\rangle). \tag{B8}$$

Equations (B3), (B7), and (B8) allow us to calculate the term $\langle e^{\psi_s^*+\psi_i}\rangle$. When calculating this term, we first use Eq. (B1). Then, we replace $\overline{\chi_s}$, $\overline{\chi_i}$, $\overline{S_s}$, and $\overline{S_i}$ by the right-hand sides of Eqs. (B3), (B7), and (B8). The result is

$$\langle e^{\psi_i+\psi_s^*}\rangle=\exp[-\tfrac{1}{2}(\langle\chi_s^2\rangle+\langle S_s^2\rangle+\langle\chi_i^2\rangle+\langle S_i^2\rangle)+\langle\chi_s\chi_i\rangle +\langle S_sS_i\rangle+i(\langle\chi_sS_i\rangle-\langle\chi_iS_s\rangle)+O(\langle\epsilon_{\mathrm{mov}}^4\rangle)]. \tag{B9}$$

The term $\langle e^{\psi_s+\psi_i^*}\rangle$ is a complex conjugation to that $\langle e^{\psi_s^*+\psi_i}\rangle$ and can be obtained from Eq. (B9). Substitution of both terms into Eq. (9) results in Eqs. (10) and (11).

When deriving Eqs. (B9), (10), and (11), one can assume for simplicity that $\overline{S_s}=0$ and $\overline{S_i}=0$. It can be shown that in this case these equations would still hold if the argument of the exponent in Eq. (B9) is augmented by the term $i(\langle\chi_iS_i\rangle-\langle\chi_sS_s\rangle)$, and the argument of cosine in Eq. (10) is added by the term $\langle\chi_iS_i\rangle-\langle\chi_sS_s\rangle$. In this case, Eq. (10) would coincide with the analogous equation from Ref. 1.

APPENDIX C: DERIVATION OF $\langle\chi_s\chi_2\rangle+\langle S_sS_2\rangle$

In Appendix C, we calculate $\langle\chi_s\chi_2\rangle+\langle S_sS_2\rangle$ for the case of anisotropic spectra of temperature and velocity fluctuations.

Expressions for χ_s and χ_2 are given by the real parts of the right-hand sides of Eqs. (16) and (18). Multiplying these expressions and performing ensemble averaging, one obtains

$$\langle \chi_s \chi_2 \rangle = \frac{k^4 L^2}{4^2 \pi^2} \int_0^L dx_1 \int_{-\infty}^{\infty} dy_1 \int_{-h_s}^{\infty} dz_1 \int_{L_1}^{L} dx_2 \int_{-\infty}^{\infty} dy_2 \times \int_{-h_s}^{\infty} dz_2 \frac{B_{\text{eff}}(x_1-x_2, y_1-y_2, z_1-z_2)}{x_1(L-x_1)x_2(L-x_2)} \times \cos\left\{ \frac{kL}{2x_1(L-x_1)} \left[\left(z_1 + (h_s-h_r)\frac{x_1}{L} \right)^2 + y_1^2 \right] \right\} \times \cos\left\{ \frac{kL}{2x_2(L-x_2)} \left[\left(z_2 + 2h_s - (h_s+h_r)\frac{x_2}{L} \right)^2 + y_2^2 \right] \right\}. \tag{C1}$$

Here, $B_{\text{eff}}(x_1-x_2, y_1-y_2, z_1-z_2) = \langle \epsilon_{\text{mov}}(x_1, y_1, z_1) \epsilon_{\text{mov}}(x_2, y_2, z_2) \rangle$ is the correlation function of the random field ϵ_{mov}. We use the Markov approximation for this correlation function: $B_{\text{eff}}(x, \mathbf{r}) = \delta(x) b_{\text{eff}}(\mathbf{r})$. This allows us to calculate the integral over x_2.

Expressions for S_s and S_2 are given by the imaginary parts of the right-hand sides of Eqs. (16) and (18). Therefore, $\langle S_s S_2 \rangle$ is also given by the right side of Eq. (C1) if the two cosines in this equation are replaced by sines. Adding formulas for $\langle \chi_s \chi_2 \rangle$ and $\langle S_s S_2 \rangle$ yields

$$\langle \chi_s \chi_2 \rangle + \langle S_s S_2 \rangle = \frac{k^4 L^2}{4^2 \pi^2} \int_{L_1}^{L} dx \int_{-\infty}^{\infty} dy_1 \int_{-\infty}^{\infty} dy_2 \times \int_{-h_s}^{\infty} dz_1 \int_{-h_s}^{\infty} dz_2 \frac{b_{\text{eff}}(y_1-y_2, z_1-z_2)}{x^2(L-x)^2} \times \cos\left\{ \frac{kL}{2x(L-x)} \left[\left(z_1 + (h_s-h_r)\frac{x}{L} \right)^2 - \left(z_2 + 2h_s - (h_s+h_r)\frac{x}{L} \right)^2 + y_1^2 - y_2^2 \right] \right\}. \tag{C2}$$

In this equation, we introduce new integration variables $u_1 = z_1 + (h_s - h_r)x/L$, and $u_2 = z_2 + 2h_s - (h_s+h_r)x/L$. Equation (C2) becomes

$$\langle \chi_s \chi_2 \rangle + \langle S_s S_2 \rangle = \frac{k^4 L^2}{4^2 \pi^2} \mathrm{Re} \int_{L_1}^{L} dx \int_{-\infty}^{\infty} dy_1 \int_{-\infty}^{\infty} dy_2 \times \int_{m_1}^{\infty} du_1 \int_{-m_2}^{\infty} du_2 \times \frac{b_{\text{eff}}(y_1-y_2, u_1-u_2+2h_s(1-x/L))}{x^2(L-x)^2} \times \exp\left\{ \frac{ikL}{2x(L-x)} [u_1^2 - u_2^2 + y_1^2 - y_2^2] \right\}. \tag{C3}$$

Here, $m_1(x)$ and $m_2(x)$ are given by Eq. (25). Now we introduce new integration variables in Eq. (C3):

$$\xi = u_1 - u_2, \quad \eta = \frac{u_1+u_2}{2}, \quad \xi' = y_1 - y_2, \quad \eta' = \frac{y_1+y_2}{2}. \tag{C4}$$

Then, Eq. (C3) takes the form

$$\langle \chi_s \chi_2 \rangle + \langle S_s S_2 \rangle = \frac{k^4 L^2}{4^2 \pi^2} \mathrm{Re} \int_{L_1}^{L} dx \left\{ \int_{-\infty}^{m_1+m_2} d\xi \int_{m_1 - \xi/2}^{\infty} d\eta + \int_{m_1+m_2}^{\infty} d\xi \int_{-m_2+\xi/2}^{\infty} d\eta \right\} \times \int_{-\infty}^{\infty} d\xi' \int_{-\infty}^{\infty} d\eta' \frac{b_{\text{eff}}(\xi', \xi + 2h_s(1-x/L))}{x^2(L-x)^2} \times \exp\left[\frac{ikL(\xi\eta + \xi'\eta')}{x(L-x)} \right]. \tag{C5}$$

In this equation, the integral over η' is equal to $2\pi\delta(kL\xi'/x(L-x))$. Therefore, the integral over ξ' can be calculated. The result is

$$\langle \chi_s \chi_2 \rangle + \langle S_s S_2 \rangle = \frac{k^3 L}{8\pi} \mathrm{Re} \int_{L_1}^{L} dx \left\{ \int_{-\infty}^{m_1+m_2} d\xi \int_{m_1 - \xi/2}^{\infty} d\eta + \int_{m_1+m_2}^{\infty} d\xi \int_{-m_2+\xi/2}^{\infty} d\eta \right\} \times \frac{b_{\text{eff}}(0, \xi + 2h_s(1-x/L))}{x(L-x)} \exp\left(\frac{ikL\xi\eta}{x(L-x)} \right). \tag{C6}$$

We remind the reader of a useful formula:[1]

$$\int_b^{\infty} d\eta \, e^{ia\eta} = \pi\delta(a) + i\frac{e^{iab}}{a}, \tag{C7}$$

The use of this formula allows us to calculate the integral over η in Eq. (C6). The resulting formula for $\langle \chi_s \chi_2 \rangle + \langle S_s S_2 \rangle$ is given by Eq. (21).

[1] S. F. Clifford and R. J. Lataitis, ''Turbulence effects on acoustic wave propagation over a smooth surface,'' J. Acoust. Soc. Am. **73**, 1545–1550 (1983).

[2] G. A. Daigle, ''Correlation of the phase and amplitude fluctuations between direct and ground-reflected sound,'' J. Acoust. Soc. Am. **68**, 297–302 (1980).

[3] V. I. Tatarskii, *The Effects of the Turbulent Atmosphere on Wave Propagation* (Israel Program for Scientific Translation, Jerusalem, 1971).

[4] V. E. Ostashev, *Acoustics in Moving Inhomogeneous Media* (E & FN SPON, An Imprint of Thompson Professional, London, 1997).

[5] D. K. Wilson and D. W. Thomson, ''Acoustic propagation through anisotropic, surface-layer turbulence,'' J. Acoust. Soc. Am. **96**, 1080–1095 (1994).

[6] Ph. Blanc-Benon, J. Wasier, D. Juvé, and V. E. Ostashev, ''Experimental studies of sound propagation through thermal turbulence near a boundary,'' in *Proceedings of the 29th International Congress and Exhibition on Noise Control Engineering* (Nice, France, 2000).

[7] D. K. Wilson, ''Calculated coherence and extinction of sound waves propagating through anisotropic, shear-induced turbulent velocity fluctuations,'' J. Acoust. Soc. Am. **105**, 658–671 (1999).

[8] P. Chevret, Ph. Blanc-Benon, and D. Juvé, ''A numerical model for sound propagation through a turbulent atmosphere near the ground,'' J. Acoust. Soc. Am. **100**, 3587–3599 (1996).

[9] M. West, K. Gilbert, and R. A. Sack, ''A tutorial on the parabolic equation (PE) model used for long range sound propagation in the atmosphere,'' Appl. Acoust. **37**, 31–49 (1992).
[10] K. E. Gilbert, R. Raspet, and X. Di, ''Calculation of turbulence effects in an upward-retracting atmosphere,'' J. Acoust. Soc. Am. **87**, 2428–2437 (1990).
[11] V. E. Ostashev and G. H. Goedecke, ''Interference of direct and ground reflected waves in a turbulent atmosphere,'' *Proceedings of the 8th International Symposium on Long Range Sound Propagation*, Penn State University, 1998, pp. 313–325.
[12] E. M. Salomons, ''The fluctuating field of a monopole source in a turbulent atmosphere above a ground surface. Time-averaged sound pressure level and statistical distributions,'' in Ref. 11, pp. 326–351.
[13] E. Salomons, V. E. Ostashev, S. Clifford, and R. Lataitis, ''Sound propagation in a turbulent atmosphere near the ground: an approach based on the spectral representation of refractive-index fluctuations,'' J. Acoust. Soc. Am. **109**, 1881–1893 (2001).
[14] K. Attenborough, S. I. Hayek, and J. M. Lawther, ''Propagation of sound above a porous half-space,'' J. Acoust. Soc. Am. **68**, 1493–1501 (1980).
[15] M. A. Nobile and S. I. Hayek, ''Acoustic propagation over an impedance plane,'' J. Acoust. Soc. Am. **78**, 1325–1336 (1985).
[16] X. Di and K. E. Gilbert, ''An exact Laplace transform formulation for a point source above a ground surface,'' J. Acoust. Soc. Am. **93**, 714–720 (1993).
[17] S. M. Rytov, Yu. A. Kravtzov, and V. I. Tatarskii, *Principles of Statistical Radio Physics. Part 4. Wave Propagation Through Random Media* (Springer-Verlag, Berlin, 1989).
[18] A. Ishimaru, *Wave Propagation and Scattering in Random Media* (IEEE Press, New York, 1997).
[19] A. S. Gurvich, A. I. Kon, V. L. Mironov, and S. S. Khmelevtsov, *Laser Radiation in the Turbulent Atmosphere* (Nauka, Moscow, 1976) (in Russian).
[20] Here, we use the following formula: $\lim_{a\to\infty} \sin(a\xi)/\xi = \pi\delta(\xi)$; e.g., see G. A. Korn and T. M. Korn, *Mathematical Handbook* (McGraw-Hill, New York, 1968).
[21] A. V. Kukushkin and I. M. Fuks, ''Reflection formulas for a wave field in a random-nonuniform medium above an ideally reflecting interface,'' Radiophys. Quantum Electron. **31**, 1035–1042 (1980).
[22] V. E. Ostashev, B. Brähler, V. Mellert, and G. H. Goedecke, ''Coherence functions of plane and spherical waves in a turbulent medium with the von Karman spectrum of medium inhomogeneities,'' J. Acoust. Soc. Am. **104**, 727–737 (1998).
[23] J. Mann, ''The spatial structure of neutral atmospheric surface layer turbulence,'' J. Fluid Mech. **273**, 141–168 (1994).
[24] J. Wasier, ''Etude expérimentale des effects d'une frontière sur la propagation des ondes acoustiques à travers une turbulence thermique,'' Ph.D. thesis, ECL 99–46, Ecole Centrale de Lyon, 1999.
[25] Ph. Blanc-Benon and D. Juvé, ''Experimental study of sound propagation above rigid and non rigid boundaries in the presence of turbulence,'' *Proceedings of the 9th International Symposium on Long Range Sound Propagation*, The Hague, 2000 (in press).

Statistical moments of the sound field propagating in a random, refractive medium near an impedance boundary

D. K. Wilson
U.S. Army Research Laboratory, AMSRL-CI-EP, 2800 Powder Mill Road, Adelphi, Maryland 20783

V. E. Ostashev
NOAA/Environmental Technology Laboratory, 325 Broadway, Boulder, Colorado 80305
and Physics Department, New Mexico State University, Las Cruces, New Mexico 88003

(Received 8 October 1999; accepted for publication 26 January 2001)

Propagation of a monochromatic sound field in a refractive and turbulent medium near an impedance boundary is considered. Starting from the parabolic equation for a moving medium and using the Markov approximation, a closed equation for the statistical moments of arbitrary order of the sound-pressure field is derived. Numerical methods for directly solving the first- and second-moment versions of this equation are formulated. The first-moment formulation is very similar to parabolic equations (PEs) that are now widely used to calculate sound fields for particular realizations of a random medium. The second-moment formulation involves a large, fringed tridiagonal matrix, which is solved using iterative refinement and Cholesky factorization. The solution is computationally intensive and currently restricted to low frequencies. As an example, the first and second moments are directly calculated for upwind and downwind propagation of a sound wave through a turbulent atmosphere. For these cases, predictions from the second-moment PE were statistically indistinguishable from the result of 40 random trials calculated with a standard Crank–Nicholson PE, although the second-moment PE yielded smoother results due to its ensemble-average nature. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1359239]

PACS numbers: 43.28.Gq, 43.28.Js, 43.28.Lv [LCS]

I. INTRODUCTION

Most previous numerical calculations of acoustic wave propagation in a refractive medium with random fluctuations have been stochastic in nature. Random behavior in the propagation medium is incorporated by synthesizing an ensemble of realizations, or ''snapshots,'' of the medium, solving separately for the sound–pressure field in each realization, and then calculating statistics of the sound field from the ensemble. For example, in atmospheric acoustics, realizations of the near-ground atmospheric turbulence field are synthesized from a postulated equation for the turbulence spectrum, and then superimposed on a height-dependent mean sound-speed profile. The sound-field calculations are performed using numerical implementations of the standard parabolic equation (PE) approximation.[1–3] A similar approach is employed in ocean acoustics, although the spectrum typically describes internal wave activity rather than turbulence.[4,5]

Clearly, it would be desirable to directly solve a deterministic equation for the statistical moments of a propagating field, if such a closed equation can be derived and an efficient method for solving it devised. Given that many theoretical results for the statistical moments in nonrefractive environments have been derived previously using the parabolic and Markov approximations,[6–9] these approximations would seem a natural starting point in formulating numerical solutions for the moments in more complex environments with refraction. Indeed, Puzenko and Chaevskii,[10] Hill,[11] Bates and Bates,[12] Mandine and Pelissier,[13] and several other authors have previously derived closed equations for the first and second moments of a propagating wavefield that incorporate scattering and refraction. Bates and Bates, as well as Mandine and Pelissier, numerically solved for the first moment.

In this article we derive a closed equation for the statistical moments of the complex sound-pressure field propagating in a refractive and turbulent medium above an impedance boundary. The equation for the first moment is solved analytically for the case of isotropic and homogeneous turbulence. Furthermore, numerical methods are devised and implemented for direct determination of the second, as well as the first, moments of the sound-pressure field in inhomogeneous turbulence. The second moment solution is particularly important because it provides the information needed to calculate the mean-square pressure and the equivalent sound-pressure level, L_{eq}. Energy scattered into refractive shadow zones by turbulence and other random phenomena is included in calculations of the second moment. The coherence of the signals between a pair of sensors, which affects the performance of beamforming arrays, also depends on the second moment. The first moment, which represents the unscattered part of the amplitude, provides insight into the nature of the propagated signal but is generally less important in applications than the second moment.

This article is structured as follows. In Sec. II, a general, closed equation for the statistical moments of propagating sound field is derived and then specialized to the cases of the first and second moments. The results are new in that they are valid in a moving medium above an impedance ground. Our methods for direct numerical solution of the first- and second-moment equations are described in Sec. III. Some

example calculations are provided in Sec. IV. Although our examples in this article are for propagation through near-ground atmospheric turbulence, propagation through other random environments such as the ocean can be handled similarly.

II. THEORY

A. Background

In this article we consider the following problem. A monochromatic sound source with frequency ω is located in the plane $x=0$ in the Cartesian coordinate system $\mathbf{R}=(x,y,z)$. The plane $z=0$ of the coordinate system coincides with the surface of the ground, the z-axis being directed upwards. The x-axis is parallel to the ground and is in the horizontal direction from the source to a receiver. The mean values of the adiabatic sound speed and wind velocity vector in the atmosphere are denoted by $c(\mathbf{R})$ and $\mathbf{V}(\mathbf{R})=(V_x(\mathbf{R}),\mathbf{V}_\perp(\mathbf{R}))$, respectively, where $\mathbf{V}_\perp(\mathbf{R})=(V_y(\mathbf{R}),V_z(\mathbf{R}))$, and V_x, V_y, and V_z are the wind velocity components along the axes of the coordinate system. Furthermore, fluctuations of the adiabatic sound speed and wind velocity component along the x-axis due to atmospheric turbulence are denoted by $\tilde{c}(\mathbf{R})$ and $v_x(\mathbf{R})$.

Propagation of a sound-pressure field $p(\mathbf{R})$ in such a refractive and turbulent atmosphere can be described with the following parabolic equation:[9]

$$2ik_0\frac{\partial p(\mathbf{R})}{\partial x}+\hat{L}_{\mathbf{r}}p(\mathbf{R})+k_0^2\varepsilon_{\mathrm{mov}}(\mathbf{R})p(\mathbf{R})=0. \tag{1}$$

Here, $k_0=\omega/c_0$, and the subscript zero indicates evaluation at $z=0$. The quantity $\varepsilon_{\mathrm{mov}}(\mathbf{R})=-2[\tilde{c}(\mathbf{R})+v_x(\mathbf{R})]/c_0 \simeq -\tilde{T}(\mathbf{R})/T_0-2v_x(\mathbf{R})/c_0$ is twice the fluctuation in the acoustic refractive index in a moving atmosphere, where $\tilde{T}(\mathbf{R})$ is the temperature fluctuation. Furthermore, in Eq. (1) the operator $\hat{L}_{\mathbf{r}}$ is given by

$$\hat{L}_{\mathbf{r}}=\nabla_\perp^2+k_0^2+k^2(\mathbf{R})+\frac{2ik_0\mathbf{V}_\perp(\mathbf{R})\cdot\nabla_\perp}{c_0}, \tag{2}$$

where $\nabla_\perp=(\partial/\partial y,\partial/\partial z)$, $k(\mathbf{R})=k_0c_0/[c(\mathbf{R})+V_x(\mathbf{R})]$ is the sound wavenumber, and the subscript $\mathbf{r}$ indicates that $\hat{L}_{\mathbf{r}}$ operates on the transverse coordinates $\mathbf{r}=(y,z)$. The last term in Eq. (2) accounts for the effect of the mean crosswind on sound propagation. If $v_x=0$ and $\mathbf{V}=0$, Eq. (1) coincides with the standard parabolic equation, which is widely used for waves in random media.[6,7] The range of applicability of the latter equation is well known.[8] In a moving medium, Eq. (1) generalizes this standard parabolic equation by accounting for first-order terms in velocity, i.e., v_x/c_0 and V/c_0. Second and higher order terms such as v_x^2/c_0^2, V^2/c_0^2, $V\tilde{c}/c_0^2$, $v_x\tilde{c}/c_0^2$, etc. are not included in Eq. (1). Note that to first order, k^2 in Eq. (2) can be written as $k^2=k_0^2(c_0^2/c^2-2v_x/c_0)$. In this case, Eq. (1) is essentially the same as Eq. (7.122) from Ostashev,[9] where the range of applicability of this equation is considered in more detail.

The initial condition to Eq. (1) reads

$$p(x=0,\mathbf{r})=p_0(\mathbf{r}), \tag{3}$$

where $p_0(\mathbf{r})$ is the sound field due to the source in the vertical plane at $x=0$. The boundary condition to Eq. (1) is given by

$$\left.\frac{\partial p(x,y,z)}{\partial z}\right|_{z=0}=-ik_0\beta p(x,y,0). \tag{4}$$

Here β is the normalized admittance of the ground, being equal to ρ_0c_0/Z_s, where ρ_0 is the density of the air at $z=0$, and Z_s is the ground impedance.

In the Markov approximation, the spatial correlation function B_{eff} of the random field $\varepsilon_{\mathrm{mov}}$ is approximated as[8,9]

$$\begin{aligned}B_{\mathrm{eff}}(x_1,\mathbf{r}_1;x_2,\mathbf{r}_2)&=\langle\varepsilon_{\mathrm{mov}}(x_1,\mathbf{r}_1)\varepsilon_{\mathrm{mov}}(x_2,\mathbf{r}_2)\rangle\\&=\delta(x_1-x_2)b_{\mathrm{eff}}(x_1,\mathbf{r}_1;x_1,\mathbf{r}_2),\end{aligned} \tag{5}$$

where b_{eff} is called the *transverse* or *projected* correlation function.[21] The range of applicability of the Markov approximation is studied in detail by Rytov *et al.*[8] Note that for most geometries of sound propagation outdoors (including those in Sec. III), this approximation is valid provided that the parabolic equation can be used. The preceding form of B_{eff} allows us to take into account that, near the ground, atmospheric turbulence is inhomogeneous and anisotropic. Hereinafter, $b_{\mathrm{eff}}(x_1,\mathbf{r}_1;x_1,\mathbf{r}_2)$ in Eq. (5) will be denoted as $b_{\mathrm{eff}}(x_1;\mathbf{r}_1;\mathbf{r}_2)$. It follows from Eq. (5) that

$$\begin{aligned}b_{\mathrm{eff}}(x_1;\mathbf{r}_1;\mathbf{r}_2)&=\int_{-\infty}^{\infty}B_{\mathrm{eff}}(x_1,\mathbf{r}_1;x_2,\mathbf{r}_2)dx_2\\&=\int_{-\infty}^{\infty}\left[\frac{B_T(x_1,\mathbf{r}_1;x_2,\mathbf{r}_2)}{T_0^2}+\frac{4B_{xx}(x_1,\mathbf{r}_1;x_2,\mathbf{r}_2)}{c_0^2}\right]dx_2,\end{aligned} \tag{6}$$

where B_T and B_{xx} are the correlation functions of the random fields $\tilde{T}(\mathbf{R})$ and $v_x(\mathbf{R})$.

B. Closed equations for statistical moments of a sound field

We define a general moment, order $n+m$, of the complex sound pressure as

$$\begin{aligned}\Gamma_{n,m}(x;\mathbf{r}_1,\ldots,\mathbf{r}_n;\mathbf{r}_1',\ldots,\mathbf{r}_m')&\\=\langle p(x,\mathbf{r}_1)\ldots p(x,\mathbf{r}_n)p^*(x,\mathbf{r}_1')\ldots p^*(x,\mathbf{r}_m')\rangle,&\end{aligned} \tag{7}$$

where n can be 1,2,..., and m can be 0,1,2,.... Some particular cases of this arbitrary moment function $\Gamma_{n,m}$ are of special interest in applications. The first moment $\Gamma_{1,0}$ is the mean field, representing the unscattered part of the amplitude. The second moment $\Gamma_{1,1}$ is proportional to the wave energy when $\mathbf{r}_1=\mathbf{r}_1'$ and is also related to the two-point coherence function of the wavefield. (In this article, the term ''second moment'' implicitly refers to $\Gamma_{1,1}$, although $\Gamma_{2,0}$ and $\Gamma_{0,2}$ are also valid second moments. The quantity $\Gamma_{1,1}$ is often called the mutual coherence function.) Fourth moments such as $\Gamma_{2,2}$ play an important role in describing the intensity fluctuations (scintillations) of the sound field.

Starting from Eq. (1) and using the Markov approximation (5), the following equation for $\Gamma_{n,m}(x;\mathbf{r}_1,\ldots,\mathbf{r}_n;\mathbf{r}'_1,\ldots,\mathbf{r}'_m)$ is derived in Appendix A:

$$\frac{\partial\Gamma_{n,m}}{\partial x}-\frac{i}{2k_0}[\hat{L}_{\mathbf{r}_1}+\cdots+\hat{L}_{\mathbf{r}_n}-\hat{L}^*_{\mathbf{r}'_1}-\cdots-\hat{L}^*_{\mathbf{r}'_m}]\Gamma_{n,m}+\frac{k_0^2}{8}\left[\sum_{i=1}^{n}\sum_{j=1}^{n}b_{\mathrm{eff}}(x;\mathbf{r}_i;\mathbf{r}_j)+\sum_{i=1}^{m}\sum_{j=1}^{m}b_{\mathrm{eff}}(x;\mathbf{r}'_i;\mathbf{r}'_j)-2\sum_{i=1}^{n}\sum_{j=1}^{m}b_{\mathrm{eff}}(x;\mathbf{r}_i;\mathbf{r}'_j)\right]\Gamma_{n,m}=0. \tag{8}$$

The initial condition to Eq. (8) is formulated in the plane $x=0$ and can be obtained by making use of Eqs. (3) and (7):

$$\Gamma_{n,m}(x=0;\mathbf{r}_1,\ldots,\mathbf{r}_n;\mathbf{r}'_1,\ldots,\mathbf{r}'_m)=p_0(\mathbf{r}_1)\ldots p_0(\mathbf{r}_n)p_0^*(\mathbf{r}'_1)\ldots p_0^*(\mathbf{r}'_m). \tag{9}$$

The boundary conditions to Eq. (8) are derived in Appendix B. They are given by

$$\left.\frac{\partial\Gamma_{n,m}(x;\mathbf{r}_1,\ldots,\mathbf{r}_n;\mathbf{r}'_1,\ldots,\mathbf{r}'_m)}{\partial z_i}\right|_{z_i=0}=-ik_0\beta\Gamma_{n,m}|_{z_i=0},\quad i=1,\ldots,n \tag{10}$$

and

$$\left.\frac{\partial\Gamma_{n,m}(x;\mathbf{r}_1,\ldots,\mathbf{r}_n;\mathbf{r}'_1,\ldots,\mathbf{r}'_m)}{\partial z'_j}\right|_{z'_j=0}=ik_0\beta^*\Gamma_{n,m}|_{z'_j=0},\quad j=1,\ldots,m. \tag{11}$$

Thus, Eq. (8) for $\Gamma_{n,m}$ with the initial condition (9) and boundary conditions (10) and (11) is a closed equation for the statistical moments of a sound field propagating in a refractive and turbulent atmosphere near an impedance ground.

Let us consider, for a moment, line-of-sight sound propagation in a refractive and turbulent atmosphere when sound waves do not interact with the ground. Starting from Eq. (1) and the initial condition (3), and using the same approach as in Appendix A, a closed equation for $\Gamma_{n,m}$ can be derived. It can be shown that this equation coincides with Eq. (8). Furthermore, the initial condition to this equation coincides with Eq. (9). Thus, the statistical moments $\Gamma_{n,m}$ for both sound propagation near the impedance ground and line-of-sight sound propagation satisfy Eq. (8) with the initial condition (9). The difference between the statistical moments $\Gamma_{n,m}$ for these geometries is that, in the first case, $\Gamma_{n,m}$ should satisfy the boundary conditions (10) and (11), while for the second case these boundary conditions are not needed. Note that, for the case of electromagnetic wave propagation, analogous results were obtained in Refs. 10 and 13, where the mean field and the coherence function were considered. Also note that, for homogeneous and isotropic turbulence, when $b_{\mathrm{eff}}(x_1;\mathbf{r}_1;\mathbf{r}_2)=b_{\mathrm{eff}}(\mathbf{r}_1-\mathbf{r}_2)$, Eq. (8) coincides with Eq. (7.123) from Ostashev,[9] where the case of line-of-sight sound propagation was considered.

Later in this article, we will deal with the first moment $\langle p(x;\mathbf{r})\rangle=\Gamma_{1,0}(x;\mathbf{r})$ (mean complex sound field) and the second moment $\langle p(x;\mathbf{r})p^*(x;\mathbf{r}')\rangle=\Gamma_{1,1}(x;\mathbf{r};\mathbf{r}')$. The latter, for simplicity, will be denoted as $P(x;\mathbf{r};\mathbf{r}')$. When considering the first moment, it is convenient to introduce the mean complex amplitude $A(x,\mathbf{r})=\langle p(x,\mathbf{r})\rangle e^{-ik_0x}$. The equation for A can be obtained from Eq. (8) with $n=1$ and $m=0$:

$$\frac{\partial A}{\partial x}-\frac{i}{2k_0}\left[\nabla_\perp^2+k^2(x,\mathbf{r})-k_0^2+\frac{2ik_0\mathbf{V}_\perp(x,\mathbf{r})\cdot\nabla_\perp}{c_0}\right]A+\frac{k_0^2}{8}b_{\mathrm{eff}}(x;\mathbf{r};\mathbf{r})A=0. \tag{12}$$

The initial condition to this equation reads: $A(0,\mathbf{r})=p_0(\mathbf{r})$. The boundary condition for Eq. (12), obtained from Eq. (10), is

$$\left.\frac{\partial A(x,y,z)}{\partial z}\right|_{z=0}=-ik_0\beta A(x,y,z)|_{z=0}. \tag{13}$$

The equation for the second moment $P(x;\mathbf{r};\mathbf{r}')$ can be obtained from Eq. (8) with $n=1$ and $m=1$:

$$\frac{\partial P}{\partial x}-\frac{i}{2k_0}\left[\nabla_\perp^2-\nabla_\perp'^2+k^2(x,\mathbf{r})-k^2(x,\mathbf{r}')+\frac{2ik_0}{c_0}(\mathbf{V}_\perp(x,\mathbf{r})\cdot\nabla_\perp+\mathbf{V}_\perp(x,\mathbf{r}')\cdot\nabla'_\perp)\right]P+\frac{k_0^2}{8}[b_{\mathrm{eff}}(x;\mathbf{r};\mathbf{r})+b_{\mathrm{eff}}(x;\mathbf{r}';\mathbf{r}')-2b_{\mathrm{eff}}(x;\mathbf{r};\mathbf{r}')]P=0, \tag{14}$$

where $\nabla'_\perp=(\partial/\partial y',\partial/\partial z')$. The initial condition to this equation reads: $P(0;\mathbf{r};\mathbf{r}')=p_0(\mathbf{r})p_0^*(\mathbf{r}')$. The boundary conditions to Eq. (14) follow from Eqs. (10) and (11)

$$\left.\frac{\partial P(x;\mathbf{r};\mathbf{r}')}{\partial z}\right|_{z=0}=-ik_0\beta P|_{z=0},\qquad\left.\frac{\partial P(x;\mathbf{r};\mathbf{r}')}{\partial z'}\right|_{z'=0}=ik_0\beta^*P|_{z'=0}. \tag{15}$$

Equations (12) and (14) are too complicated to be solved numerically with current resources. However, it is possible to solve these equations in the 2-D case when the dependence of all functions on y is omitted. We will also assume that the vertical wind speed $V_z=0$, and that the transverse correlation function b_{eff} and wavenumber k do not depend on the range x. Then, from Eqs. (12) and (14) we have

$$\frac{\partial A(x,z)}{\partial x}-\frac{i}{2k_0}\left[\frac{\partial^2}{\partial z^2}+k^2(z)-k_0^2\right]A+\frac{k_0^2}{8}b_{\mathrm{eff}}(z;z')A=0, \tag{16}$$

and

$$\frac{\partial P(x;z;z')}{\partial x}-\frac{i}{2k_0}\left[\frac{\partial^2}{\partial z^2}-\frac{\partial^2}{\partial z'^2}+k^2(z)-k^2(z')\right]P$$
$$+\frac{k_0^2}{8}[b_{\text{eff}}(z;z)+b_{\text{eff}}(z';z')-2b_{\text{eff}}(z;z')]P=0. \quad (17)$$

The boundary conditions to Eqs. (16) and (17) coincide with the boundary conditions (13) and (15), with the dependence on y omitted.

C. Analytical solution for the mean field

For the particular case of isotropic and homogeneous turbulence, Eq. (12) for the mean complex amplitude A can be solved analytically. Indeed, in this case $B_{\text{eff}}(x_1,\mathbf{r}_1,x_2,\mathbf{r}_2)=B_{\text{eff}}(x_1-x_2,\mathbf{r}_1-\mathbf{r}_2)$. Furthermore, it can be shown from Eq. (6) that b_{eff} does not depend on x_1 and depends only on the difference between the vectors $\mathbf{r}_1$ and $\mathbf{r}_2$: $b_{\text{eff}}(x_1;\mathbf{r}_1;\mathbf{r}_2)=b_{\text{eff}}(\mathbf{r}_1-\mathbf{r}_2)$. Therefore, in Eq. (12) $b_{\text{eff}}(x;\mathbf{r};\mathbf{r})=b_{\text{eff}}(0)$. Now Eq. (12) can be solved for A:

$$A(x,\mathbf{r})=A^{(0)}(x,\mathbf{r})e^{-k_0^2 b_{\text{eff}}(0)x/8}. \quad (18)$$

Here $A^{(0)}(x,\mathbf{r})$ is the complex amplitude of the sound field in the atmosphere without turbulence. $A^{(0)}$ satisfies Eq. (12) with the last term on the left-hand side omitted:

$$\frac{\partial A^{(0)}}{\partial x}-\frac{i}{2k_0}\left[\nabla_\perp^2+k^2(\mathbf{R})-k_0^2+\frac{2ik_0\mathbf{V}_\perp(\mathbf{R})\cdot\nabla_\perp}{c_0}\right]A^{(0)}=0, \quad (19)$$

and the boundary condition

$$\left.\frac{\partial A^{(0)}(x,y,z)}{\partial z}\right|_{z=0}=-ik_0\beta A^{(0)}(x,y,0). \quad (20)$$

The initial condition to Eq. (19) reads: $A^{(0)}(0,\mathbf{r})=p_0(\mathbf{r})$.

The first term on the right-hand side of Eq. (18), $A^{(0)}(x,\mathbf{r})$ can be found as a numerical solution of Eq. (19) describing sound propagation near the impedance ground in a refractive atmosphere. Furthermore, the second exponential term in Eq. (18) describes the attenuation of the mean sound field due to turbulence. Therefore, in Eq. (18) the effects of turbulence on $\langle p\rangle$ are separated from those of impedance ground and refraction and hence can be studied independently.

III. NUMERICAL IMPLEMENTATION

A. First moment

Numerical solution of Eq. (16) begins with discretization of the computational domain in the vertical (z-) direction. As is typical for PEs, we use a vertical grid with N points and constant spacing Δz. The discrete levels are therefore $z_n=n\Delta z$, $n=1,2,\ldots,N$. We define

$$b_n=\frac{i[k^2(z_n)-k_0^2]}{2k_0}, \quad h_{mn}=\frac{k_0^2}{8}b_{\text{eff}}(z_m;z_n), \quad (21)$$

and

$$A_n(x)=A(x,z_n),$$

Substituting the central finite difference approximation $\partial^2 A_n/\partial z^2\simeq(A_{n-1}-2A_n+A_{n+1})/\Delta z^2$, Eq. (16) becomes

$$\frac{\partial A_n(x)}{\partial x}=\frac{i}{2k_0\Delta z^2}[A_{n-1}(x)-2A_n(x)+A_{n+1}(x)]$$
$$+(b_m-h_{mm})A_n(x). \quad (22)$$

This result can be conveniently recast in the matrix form

$$\frac{\partial\mathbf{A}(x)}{\partial x}=(\mathbf{E}-\mathbf{F})\mathbf{A}(x), \quad (23)$$

in which

$$\mathbf{A}(x)=[A_1(x)A_2(x)\cdots A_N(x)]^T, \quad (24)$$

$$\mathbf{F}=\text{diag}(h_{11},h_{22},\ldots,h_{NN}), \quad (25)$$

and $\mathbf{E}=(i/2k_0\Delta z^2)\mathbf{T}+\mathbf{D}$, where

$$\mathbf{T}=\begin{bmatrix}-2+\sigma_1 & 1+\sigma_2 & 0 & \cdots & \\ 1 & -2 & 1 & \cdots & \\ 0 & 1 & -2 & 1 & \\ \vdots & \vdots & \ddots & \ddots & \ddots\end{bmatrix} \quad (26)$$

and

$$\mathbf{D}=\text{diag}(b_1,b_2,\ldots,b_N). \quad (27)$$

The superscript T is the matrix transpose. The constants σ_1 and σ_2 implement the impedance lower boundary condition as discussed in Appendix B. The only significant differences between Eq. (23) above and (22) in West *et al.*[14] are the appearance of the diagonal matrix $\mathbf{F}$, which incorporates the 2-D correlation function, and replacement of the deterministic sound field by its first moment. The final step in formulating the solution is to discretize the range coordinate, by introducing the centered difference approximations

$$\left.\frac{\partial\mathbf{A}}{\partial x}\right|_{x=x_{j+1/2}}\simeq\frac{\mathbf{A}(x_{j+1})-\mathbf{A}(x_j)}{\Delta x},$$
$$\mathbf{A}_{x=x_{j+1/2}}\simeq\frac{\mathbf{A}(x_{j+1})+\mathbf{A}(x_j)}{2}. \quad (28)$$

This results in

$$\left[\mathbf{I}_N-\frac{\Delta x}{2}(\mathbf{E}-\mathbf{F})\right]\mathbf{A}(x_{j+1})=\left[\mathbf{I}_N+\frac{\Delta x}{2}(\mathbf{E}-\mathbf{F})\right]\mathbf{A}(x_j), \quad (29)$$

in which $\mathbf{I}_N$ is the $N\times N$ identity matrix. The usual method for advancing the solution at step x_j to x_{j+1} is to multiply $\mathbf{A}(x_j)$ by $\mathbf{M}^-=[\mathbf{I}_N+(\Delta x/2)(\mathbf{E}-\mathbf{F})]$, thereby obtaining a linear matrix equation, in which the matrix $\mathbf{M}^+=[\mathbf{I}_N-(\Delta x/2)(\mathbf{E}-\mathbf{F})]$ premultiplies $\mathbf{A}(x_{j+1})$. Because the matrix $\mathbf{M}^+$ is tridiagonal, numerical implementation of the first-moment equation (16) is efficient and remains essentially the same as the standard PE.

B. Second moment

Following the same procedure as for the first-moment equation, Eq. (17) becomes, when discretized in the vertical coordinate,

$$\frac{\partial P_{mn}(x)}{\partial x} = \frac{i}{2k_0\Delta z^2}[P_{m-1,n}(x) - 2P_{mn}(x) + P_{m+1,n}(x)] - \frac{i}{2k_0\Delta z^2}[P_{m,n-1}(x) - 2P_{mn}(x) + P_{m,n+1}(x)] + (b_m + b_n^* - d_{mn})P_{mn}(x), \tag{30}$$

where $P_{mn}(x) = P(x;z_m;z_n)$, and

$$d_{mn} = d_{\text{eff}}(z_m;z_n) = b_{\text{eff}}(z_m;z_m) + b_{\text{eff}}(z_n;z_n) - 2b_{\text{eff}}(z_m;z_n). \tag{31}$$

The function $d_{\text{eff}}(z_m;z_n)$ is the *transverse*, or *projected* structure function. Equation (30) is very similar to the finite-difference implementation of a two-dimensional diffusion problem,[15] except that the range coordinate x replaces time in our case. In order to propagate the solution for P_{mn}, one must also propagate $P_{m-1,n}$, $P_{m+1,n}$, $P_{m,n-1}$, and $P_{m,n+1}$. These moments in turn require $P_{m-2,n}$, $P_{m-1,n-1}$, $P_{m-1,n+1}$, etc. Therefore a full $N\times N$ matrix containing the P_{mn} must be propagated in range. This fact makes the second-moment equation intrinsically more taxing on computational resources than the first-moment equation, where only an $N\times 1$ vector had to be propagated. Converting Eq. (30) to matrix form results in

$$\frac{\partial \mathbf{P}(x)}{\partial x} = \mathbf{E}\mathbf{P}(x) + \mathbf{P}(x)\tilde{\mathbf{E}} - \mathbf{H}\odot\mathbf{P}(x), \tag{32}$$

in which $\mathbf{P}(x)$ and $\mathbf{H}$ are $N\times N$ matrices containing $P_{mn}(x)$ and d_{mn}, respectively, $\tilde{\mathbf{E}} \equiv (\mathbf{E}^*)^T$, and $\odot$ represents element-by-element (Hadamard) multiplication. Note that $\mathbf{P}$ and $\mathbf{H}$ are Hermitian matrices (i.e., $\mathbf{P} = \tilde{\mathbf{P}}$), although $\mathbf{E}$ is not necessarily so. Discretizing the range coordinate, we arrive at

$$\left(\mathbf{I}_N - \frac{\Delta x}{2}\mathbf{E} + \frac{\Delta x}{2}\mathbf{H}\odot\right)\mathbf{P}(x_{j+1}) - \frac{\Delta x}{2}\mathbf{P}(x_{j+1})\tilde{\mathbf{E}} = \left(\mathbf{I}_N + \frac{\Delta x}{2}\mathbf{E} - \frac{\Delta x}{2}\mathbf{H}\odot\right)\mathbf{P}(x_j) + \frac{\Delta x}{2}\mathbf{P}(x_j)\tilde{\mathbf{E}}. \tag{33}$$

Only the terms $\mathbf{I}_N\mathbf{P}(x_{j+1})$ and $(-\Delta x/2)\mathbf{E}\mathbf{P}(x_{j+1})$ involve pre-multiplication by tridiagonal matrices analogous to the first-moment equation; the term $(-\Delta x/2)\mathbf{P}(x_{j+1})\tilde{\mathbf{E}}$ is post-multiplication by a tridiagonal matrix, and the Hadamard multiplication $(\Delta x/2)\mathbf{H}\odot\mathbf{P}(x_{j+1})$ complicates matters even more. The usual method for solving matrix equations such as Eq. (33) is to rearrange the matrices $\mathbf{P}(x_j)$ and $\mathbf{P}(x_{j+1})$ as column vectors.[15] The result of this rearrangement can be written in a condensed form using the Kronecker matrix product, defined as

$$\mathbf{A}\otimes\mathbf{B} = \begin{bmatrix} A_{11}\mathbf{B} & A_{12}\mathbf{B} & \cdots \\ A_{21}\mathbf{B} & A_{22}\mathbf{B} & \\ \vdots & & \ddots \end{bmatrix}.$$

(The Kronecker product of an $N_1\times N_2$ matrix with an $N_3\times N_4$ matrix has dimensions $N_1N_3\times N_2N_4$.) Using this definition, our rearranged version of Eq. (33) is

$$\left(\mathbf{I}_{N^2} - \frac{\Delta x}{2}\bar{\bar{\mathbf{E}}} + \frac{\Delta x}{2}\bar{\bar{\mathbf{H}}}\right)\overline{\mathbf{P}}(x_{j+1}) = \left(\mathbf{I}_{N^2} + \frac{\Delta x}{2}\bar{\bar{\mathbf{E}}} - \frac{\Delta x}{2}\bar{\bar{\mathbf{H}}}\right)\overline{\mathbf{P}}(x_j), \tag{34}$$

where

$$\bar{\bar{\mathbf{E}}} = \mathbf{I}_N\otimes\mathbf{E} + \mathbf{E}^*\otimes\mathbf{I}_N,$$

$$\bar{\bar{\mathbf{H}}} = \text{diag}(h_{11},h_{21},\ldots,h_{12},h_{22},\ldots,h_{NN}),$$

and

$$\overline{\mathbf{P}}(x_j) = [P_{11}(x_j)P_{21}(x_j)\cdots P_{12}(x_j)P_{22}(x_j)\cdots P_{NN}(x_j)]^T.$$

The matrix pre-multiplying $\overline{\mathbf{P}}(x_{j+1})$ in Eq. (34) is *tridiagonal with fringes*,[15] so called because the elements on the main (zeroth), the ± 1, and the $\pm N$ diagonals are nonzero. It is a huge matrix for any practical problem. Since the upper absorbing layer in an atmospheric PE solution is usually about 30 wavelengths thick, and the vertical spacing of the grid typically is 0.1 wavelengths, N is at least 300. For low-frequency calculations, we might take $N = 1000$ as representative. Hence the matrices in Eq. (34) are about $10^6\times 10^6$; these 10^{12} elements cannot be stored directly in computer memory. Fortunately, the matrix is highly sparse: only about 5×10^6 elements are nonzero, and it is feasible to store these nonzero elements.

Solution of sparse systems such as this is a well studied problem in numerical analysis. The fringes complicate matters significantly in comparison to a purely tridiagonal system. We attempted several conventional methods for solving Eq. (34). The first of these was to simply store the matrix in MATLAB as a sparse matrix, and then solve the equation directly using MATLAB's built-in sparse matrix operations.[16] Except for very low frequencies and thin upper attenuating layers, this method was unreasonably slow and exceeded available computer memory. A second method involved taking explicit advantage of the ''block tridiagonal'' structure of the matrix pre-multiplying $\overline{\mathbf{P}}(x_{j+1})$ in Eq. (34). (That is, when the matrix is viewed as an $N\times N$ matrix whose elements are in turn $N\times N$ matrices, the only nonzero blocks are the matrices on the main and ± 1 diagonals.) While efficient algorithms for solving systems with such a blocked structure can be easily programmed,[17] we found them to be only marginally faster and memory efficient than the built-in MATLAB sparse matrix operations.

Besides the direct methods, iterative methods are available for solving Eq. (34). Among these are the biconjugate gradient method[15] and successive over-relaxation.[15,17] We found the iterative methods converged far too slowly to an accurate solution. Small errors in the solution accumulate into large ones as $\overline{\mathbf{P}}$ is stepped forward in range, rapidly diminishing the fidelity of the refraction calculation.

Our difficulties with these methods led us to consider alternatives involving separate treatment of the refraction and scattering components in the PE. The following algorithm, involving propagation of the coherent part of the solution followed by an iterative refinement, was found to work well:

(1) Find the element $P_{nn}(x_j)=\langle p_n(x_j)p_n^*(x_j)\rangle$ on the diagonal of $\mathbf{P}(x_j)$ having the largest absolute value (i.e., the point on the vertical grid with the most energy). Extract the column vector $\mathbf{s}(x_j)$ containing $P_{nn}(x_j)$. Normalize $\mathbf{s}(x_j)$ by $|p_n(x_j)|$. The vector $\mathbf{s}(x_j)$ represents the part of the pressure field coherent between $p_n(x_j)$ and the other vertical grid points.

(2) Advance the coherent solution forward in range, while neglecting scattering from the random medium, by solving the following tridiagonal system for $\mathbf{s}(x_{j+1})$:

$$\left(\mathbf{I}_N-\frac{\Delta x}{2}\mathbf{E}\right)\mathbf{s}(x_{j+1})=\left(\mathbf{I}_N+\frac{\Delta x}{2}\mathbf{E}\right)\mathbf{s}(x_j). \tag{35}$$

(3) Form an initial estimate for $\mathbf{P}(x_{j+1})$ based on the coherent calculation, $\mathbf{P}^0(x_{j+1})\simeq\mathbf{s}(x_{j+1})\tilde{\mathbf{s}}(x_{j+1})$.

(4) For $m=1,2,\ldots,M$, refine the estimate using

$$\mathbf{P}^m(x_{j+1})=\mathbf{P}(x_j)+\frac{\Delta x}{2}\{(\mathbf{E}-\mathbf{H}\odot)\times[\mathbf{P}^{m-1}(x_{j+1})+\mathbf{P}(x_j)]+[\mathbf{P}^{m-1}(x_{j+1})+\mathbf{P}(x_j)]\tilde{\mathbf{E}}\}. \tag{36}$$

(5) Set $\mathbf{P}(x_{j+1})=\mathbf{P}^M(x_{j+1})$.

(6) Increment j and repeat steps 1–5 until the desired range is reached.

The second step is essentially the same as the standard $N\times 1$ PE with refraction only. The result from this step is subsequently refined to include random scattering effects. Equation (36) in step 4 follows from Eq. (33) and the assumption that the terms involving the range step Δx are small. Through trial and error, we found that range steps had to be smaller than about 1/50 of a wavelength to assure convergence of the iterative refinement. This value, significantly smaller than the typical PE range step of 1/10 wavelength, makes the second-moment equation proportionately slower than the first-moment one. Despite this shortcoming, the iterative algorithm was the fastest we devised that provided acceptable accuracy. We found that $M=5$ iterations in step 4 yielded excellent results for the example scenarios to be discussed in the following section. [Instead of fixing M to a constant value, one might implement some type of stopping criterion based, for example, on convergence of the diagonal elements in $\mathbf{P}^m(x_{j+1})$.] More specifics on computational times will be provided with the examples in the next section.

An alternative technique for propagating the second-moment matrix, based on Cholesky factorization, is described in Appendix C. Although we found this technique somewhat slower than the iterative one, it does lead to an interesting reduction of the second-moment system to a linear tridiagonal one. This reduction may eventually prove beneficial to development of computationally efficient second-moment PEs.

IV. EXAMPLES

A. Description of the calculations

In this section we compare numerical solutions of the first- and second-moment equations to the standard PE method based on synthesis of random fields and averaging. The frequency for all our calculations is $f=40$ Hz. Higher frequencies were impractical for the second-moment equation, because of the demanding computational requirements.

The standard PE is based on the Crank–Nicholson formulation described in West *et al.*[14] The first-moment equation is essentially the same, except for the inclusion of the diagonal matrix representing the turbulence correlation function (Sec. III A). The computational grid spacing in both the vertical and horizontal directions was one-tenth wavelength. The algorithm for the second-moment calculations is that described in Sec. III B, with grid spacing 1/10 wavelength in the vertical and 1/50 in the horizontal. For all calculations, an upper absorbing layer of 30 wavelengths was appended to the top of the computational domain, with the attenuation

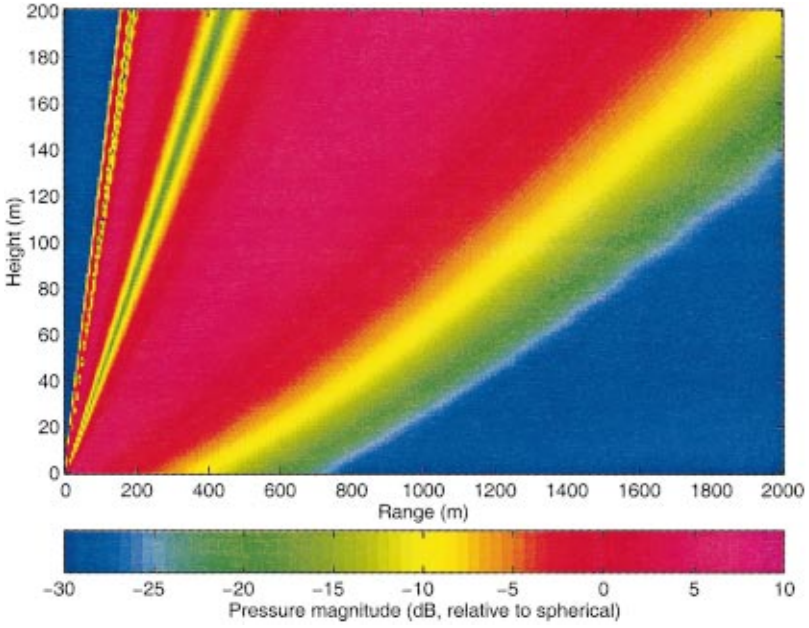


FIG. 1. Relative sound pressure level for upward refraction in the absence of turbulent scattering.

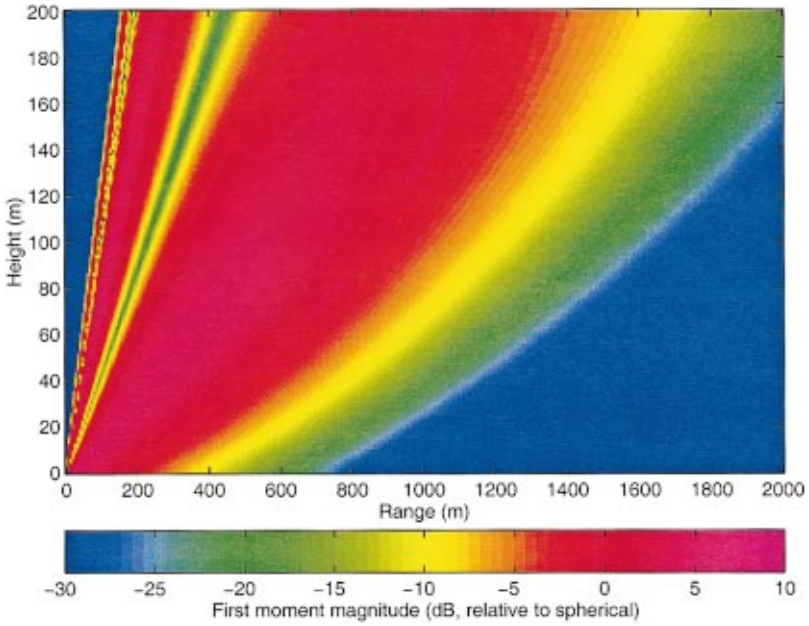


FIG. 2. First moment of the sound pressure for upward refraction as calculated with Eq. (16). Plotted is the magnitude, compensated for spherical spreading loss.

parameter μ set to 0.43, as recommended by Salomons.[18] The Gaussian starting function discussed by Salomons was used for $p_0(z)$ in Eq. (9).

The source height (5 m) and ground parameters in the calculations were chosen for consistency with the NATO RSG-11 benchmark problems,[19] although the acoustic frequency and model atmosphere in this article are different. The surface impedance was determined from a relaxation model,[20] with parameters chosen to mimic Attenborough's microstructural model.[21] Using the parameters for the flow resistivity, porosity, tortuosity, and shape factor given in Table I of Ref. 18, the resulting normalized impedance at $f=40$ Hz is $Z_s/(\rho_0 c_0)=20.8+i19.2$. Air absorption is included in the calculations although it is negligible at this frequency.

The model atmosphere is based on turbulence scaling for a neutral atmospheric surface layer. Two parameters are involved: the friction velocity, u_*, and the roughness length, z_0. For the simulations in this article we used $u_*=0.6$ m/s and $z_0=0.01$ m, values typical of windy conditions over long grass. Since conditions are neutral, the adiabatic sound speed profile is taken as constant, $c(z)=c_0$. The mean wind speed in the propagation direction is

$$V_x(z)=\frac{u_*\cos\alpha}{0.4}\ln\frac{z}{z_0}, \tag{37}$$

where α is the angle between the wind vector and propagation direction.

We use the von Kármán model for the turbulent velocity fluctuations, v_x. (Due to the assumption of neutral turbulence, the fluctuations in temperature and hence the adiabatic sound speed are zero.) The von Kármán model implies the two-dimensional correlation function[22–24]

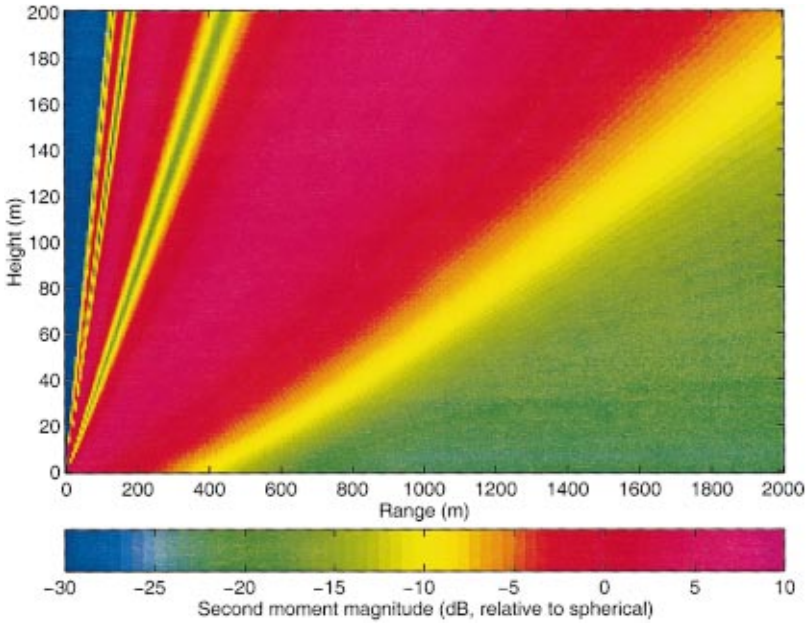


FIG. 3. Second moment of the sound pressure (mean-square pressure) for upward refraction as calculated with Eq. (17). The plotted values are compensated for spherical spreading loss.

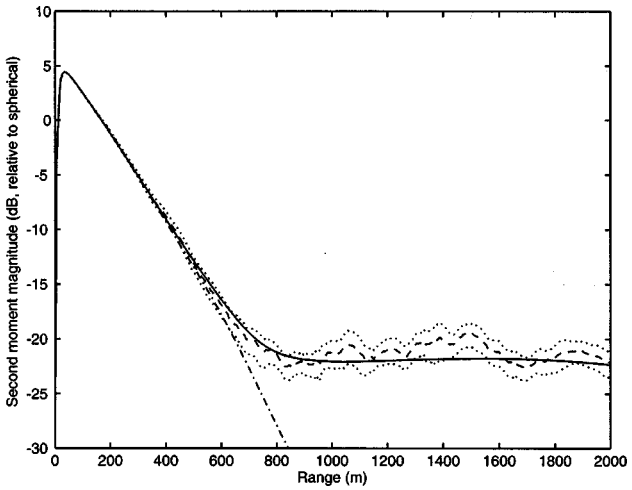


FIG. 4. Comparison of second moments (mean-square pressures) for upward refraction calculated using various methods. The receiver height is 2.6 m. Dash–dotted line is the calculation without turbulence. Solid line is the second-moment equation with turbulence. Dashed line is the average of 40 runs of the standard PE using random snapshots of the turbulence. Dotted lines enclose the 90% confidence interval for the 40 runs.

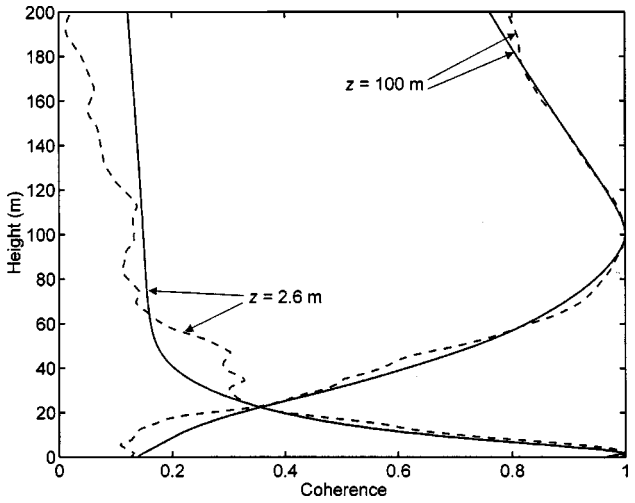


FIG. 5. Vertical normalized coherence for upward refraction at 1000 m range. Solid lines are the second-moment PE, and dashed lines are calculations from 40 random trials. The coherence centered at two different heights, 2.6 and 100.5 m, is shown.

$$b_{\text{eff}}(z;z') = \frac{16\pi^{1/2}\sigma^2 l}{\Gamma(1/3)}\left(\frac{|z-z'|}{2l}\right)^{5/6} \times\left[K_{5/6}\left(\frac{|z-z'|}{l}\right) - \frac{|z-z'|}{2l}K_{1/6}\left(\frac{|z-z'|}{l}\right)\right], \tag{38}$$

where $K_\nu(z)$ is the modified Bessel function. The variance σ^2 is set to $3.0u_*^2$ and the length scale l to $1.8[(z+z')/2]$ for pure shear turbulence.[25]

For our example calculations involving random synthesized turbulence fields, the method described in the Appendix of Ref. 25 was used so that the height dependence of the turbulence length scale could be accommodated. The synthesized fields were two-dimensional, spanning 250 m in the vertical direction and 2000 m in the horizontal. The turbulence grid contained 200 points in the vertical and 300 points in the horizontal.

Each random trial with the deterministic PE required about 190 s on a 500-MHz Pentium III PC. (The code, written in MATLAB, runs more slowly than comparable Crank–Nicholson PEs written in FORTRAN or C.) The first and second moments were calculated from an ensemble of 40 random trials, leading to a total run time of 2.1 h for the ensemble. The second-moment equation with iterative refinement required roughly 12 times this long to complete a single run. The reader is reminded that this single run, although quite long, directly produces the ensemble average statistics of interest.

B. Upward refraction

When $\alpha=\pi$ in Eq. (37), propagation is upwind and a refractive shadow zone forms. Scattering of sound energy into such refractive shadows has been an active research topic during the past decade.[1–3] The sound pressure level relative to spherical spreading, without turbulence included in the calculation, is shown in Fig. 1. Calculations from the first- and second-moment equations with turbulence are shown in Figs. 2 and 3, respectively. (The moment calculations without turbulence would be identical to Fig. 1.) The first-moment calculation exhibits slightly diminished levels in comparison to the calculation without turbulence, particularly at longer ranges. This result is in qualitative agreement with the analytical solution, Eq. (18), obtained for the case of isotropic and homogeneous turbulence. The primary reason for the observed behavior is that the signal phase becomes progressively more randomized as the distance from the source increases. (The first moment vanishes when the phase is completely randomized.) In the second-moment calculations, the levels for elevation angles larger than about 10° are nearly the same as without turbulence, but at lower elevation angles, within the refractive shadow, they are sharply increased due to scattering of sound energy by turbulence. The main difference between Fig. 3 and previous calculations of this phenomenon[1,2] is that the second-moment equation produces a very smooth distribution of scattered energy without averaging an ensemble of deterministic calculations.

Predictions from the second-moment equation at a receiver height $z=2.6$ m are compared to ensemble averages in Fig. 4. There is generally a good agreement, although the ensemble average curves have a rougher appearance due to their random nature. The second-moment equation and ensemble average calculation both follow the result without scattering out to several hundred meters. At this distance the relative sound pressure level from the second-moment equation and ensemble average start to rise as scattering becomes significant. There appears a tendency for the ensemble average predictions to be 1–2 dB higher deep in the shadow, although a larger ensemble would be necessary to assert the statistical significance of this behavior.

Next, we compare calculations of the vertical coherence from the second-moment equation and ensemble-average methods. The normalized coherence function at range x, between two observation heights z and z', is defined as

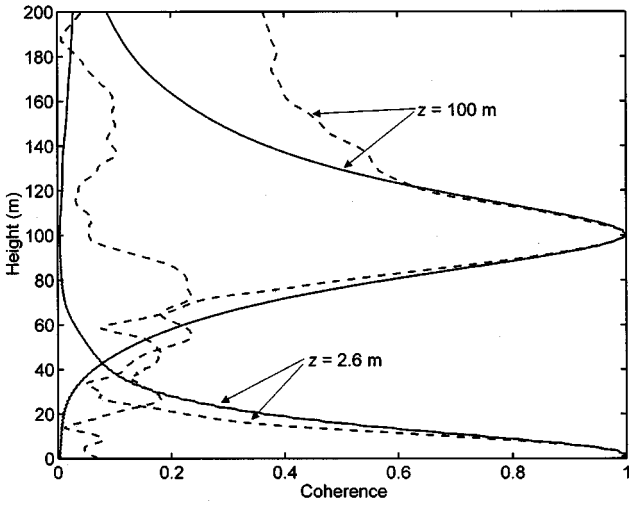


FIG. 6. Vertical normalized coherence for upward refraction at 2000 m range. Legend same as Fig. 5.

$$\gamma(x;z;z') = \frac{|P(x;z;z')|}{\sqrt{|P(x;z;z)|}\sqrt{|P(x;z';z')|}}. \tag{39}$$

Figures 5 and 6 are the normalized coherence at ranges of 1000 and 2000 m, respectively. Curves for two values of z, 2.6 and 100 m, are shown on each plot. The vertical axis represents z'. For $z=2.6$ m at 1000 m range, the normalized coherence decreases rapidly with increasing height (increasing z') near the ground. At roughly 50 m (the vicinity of the shadow boundary), the normalized coherence falls to just 0.15, but is nearly constant thereafter. The behavior for $z=2.6$ m at 2000 m range is similar, although the coherence drops nearly to zero at the shadow boundary (occurring at $z' \simeq 100$ m), and then recovers slightly. For $z=100$ m at 1000 m range, the normalized coherence decreases rapidly for z' decreasing downward from 100 m, but more slowly for z' increasing upward. Because this behavior is observed outside the shadow zone, it is likely attributable to the height-proportional length scale of the turbulence. Overall agreement between the second-moment equation and ensemble average calculation is good, although the former yields a much smoother curve.

Regarding the variable length scale of the turbulence, it should be kept in mind that the parabolic approximation requires the wavelength to be smaller than the scale of the largest inhomogeneities. Strictly speaking, this approximation cannot be valid at all heights for our turbulence model, in which the length scale decreases with height until it vanishes at ground level. In fact, for the examples in this article, where the wavelength is 8.5 m, the parabolic approximation is tenuous for heights below 20 m or so. We ran additional upward-refraction cases with l set to a fixed value of 50 m, to test whether agreement between second-moment equation and ensemble averages could be improved. No systematic improvements were observed, however, probably because most of the scattering in this example originates from elevated regions where the parabolic approximation is valid.

C. Downward refraction

When $\alpha=0$ in Eq. (37), propagation is downwind and a surface-based duct forms. This duct results in a weak modal intereference pattern. Interference minima are evident in Fig. 7 as dark spots at roughly 600, 1300, and 2000 m, which is the calculation without turbulence. First- and second-moment calculations with turbulence are shown in Figs. 8 and 9, respectively. Compared to the calculation without turbulence, the first moment is highly attenuated at long horizontal distances and large elevation angles from the source. The second-moment calculation demonstrates that scattering by turbulence disrupts the modal interference pattern, thereby increasing signal levels within the interference minima. Except for this behavior, the effect of scattering on the mean signal energy levels is subtle.

Comparisons for a receiver height $z=2.6$ m (Fig. 10) again show a close correspondence between the second-moment equation and the ensemble average predictions. For an interference mimimum between 1500 and 1800 m, the

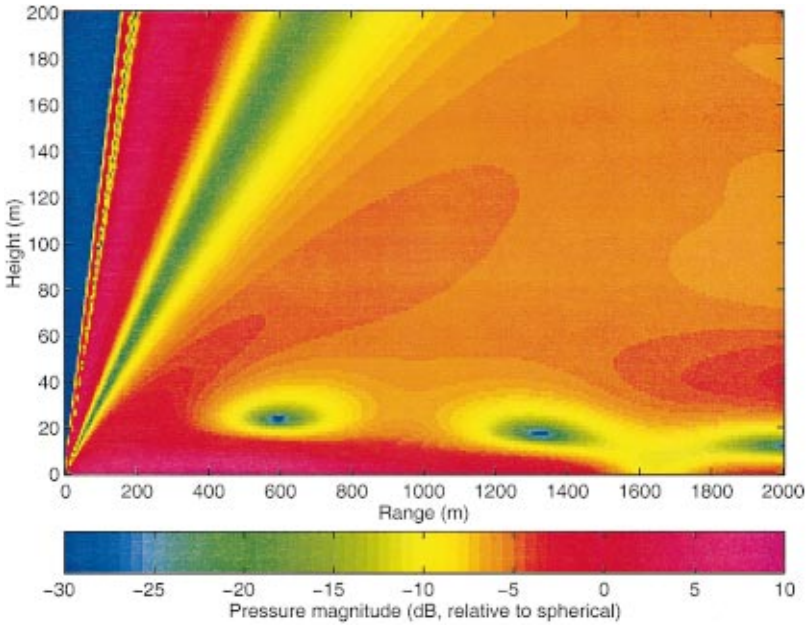


FIG. 7. Relative sound pressure level for downward refraction in the absence of turbulent scattering.

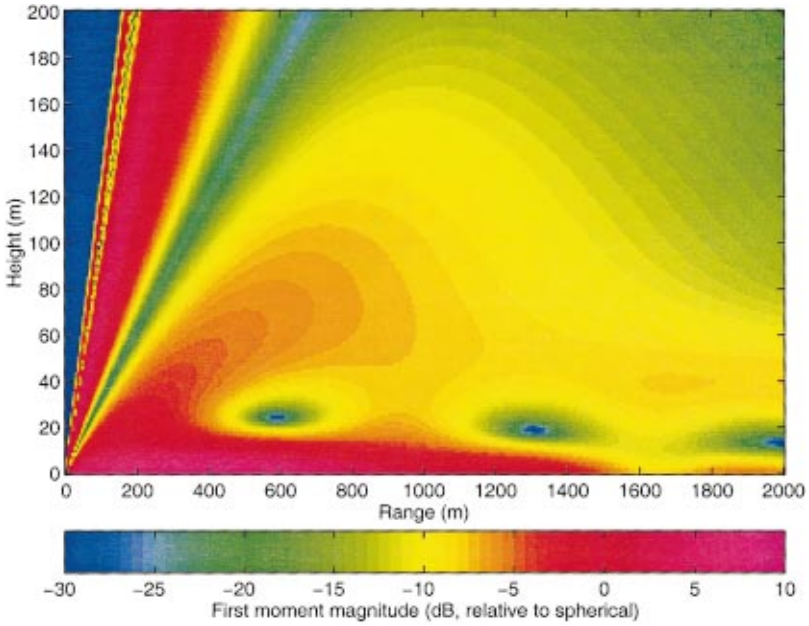


FIG. 8. First moment of the sound pressure for downward refraction as calculated with Eq. (16). Plotted is the magnitude, compensated for spherical spreading loss.

second-moment equation and ensemble averages both predict slightly higher levels than the calculation without turbulence. The situation is reversed for an interference maximum between 1800 and 2100 m.

Normalized coherence calculations for this case are shown in Figs. 11 and 12. The behavior of the coherence is significantly different for upward refraction compared to downward, exhibiting only a gradual decay for both $z=2.6$ m and $z=100$ m at 1000 m range. This behavior is similar at 2000 m range, although the coherence decay is somewhat more rapid. Furthermore, a sharp drop in the coherence is observed at $z' \simeq 10$ m for both values of z. The cause of this feature is clear from Figs. 7 and 9: the dropout occurs in the vicinity of an interference minimum, into which most of the received sound energy is randomly scattered by turbulence. This random energy has low coherence.

V. CONCLUSIONS

Starting from the parabolic equation and using the Markov approximation, a closed equation for the statistical moments of a sound field propagating in a refractive and turbulent atmosphere near an impedance boundary was derived. This equation was solved analytically for the first moment for the case of isotropic and homogeneous turbulence and numerically for the first and second moments for isotropic and inhomogeneous turbulence.

Numerical solution of the first-moment equation is very similar to previous PEs, requiring solution of an $N_z \times N_z$ tridiagonal matrix equation (where N_z is the number of vertical grid points, usually spaced at 10 per wavelength) for a column vector of $N_z \times 1$ unknown pressures. Numerical solution of the second-moment equation is quite different from

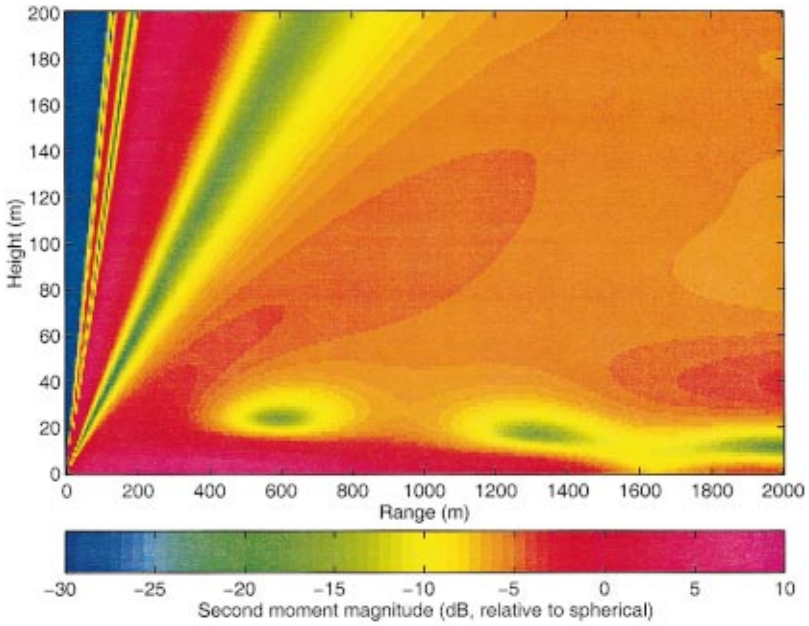


FIG. 9. Second moment of the sound pressure (mean-square pressure) for downward refraction as calculated with Eq. (17). The plotted values are compensated for spherical spreading loss.

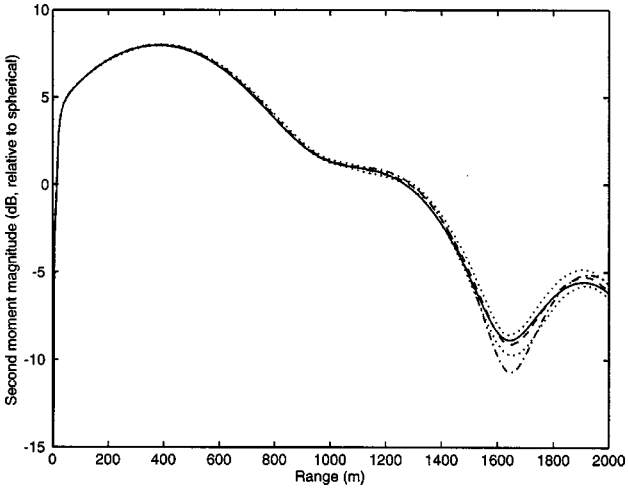


FIG. 10. Comparison of second moments (mean-square pressures) for downward refraction calculated using various methods. The receiver height is 2.6 m. Dash–dotted line is the calculation without turbulence. Solid line is the second-moment equation with turbulence. Dashed line is the average of 40 runs of the standard PE using random snapshots of the turbulence. Dotted lines enclose the 90% confidence interval for the 40 runs.

previous PEs, however, being analogous to a two-dimensional diffusion problem. Because an $N_z \times N_z$ matrix of second moments must be propagated, the second-moment solution is quite demanding on computational resources. The most straightforward numerical method for formulating the second-moment solution involves rearranging the $N_z \times N_z$ as an $N_z^2 \times 1$ column vector, in which case a linear system involving a very large $(N_z^2 \times N_z^2)$ tridiagonal matrix with fringes must be solved at each range step. We found this method impractical to implement, however, due to the demands on computational resources. Alternative methods based on iterative refinement and factorization of the second-moment matrix were therefore developed. These methods proved feasible at low frequencies. In our example case, at a frequency 40 Hz, the second-moment equation required about 12 times the computational time of 40 random trials with a standard Crank–Nicholson PE. The second-moment equation has the advantage of producing much smoother results than the random trials. It also does not require complicated techniques for synthesizing inhomogeneous random fields and testing for statistical convergence.

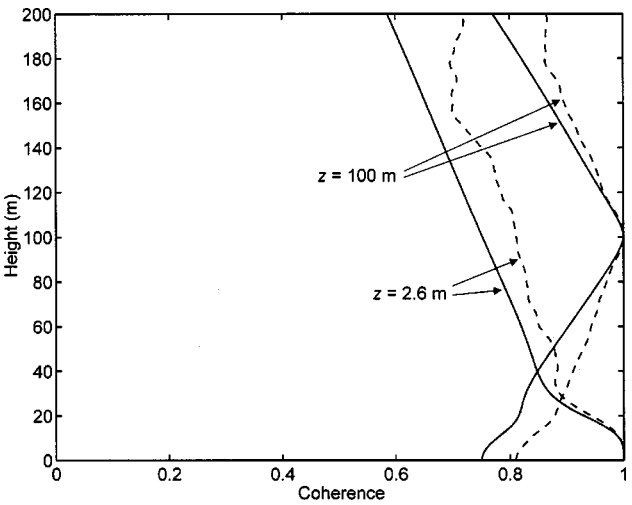


FIG. 11. Vertical normalized coherence for downward refraction at 1000 m range. Legend same as Fig. 5.

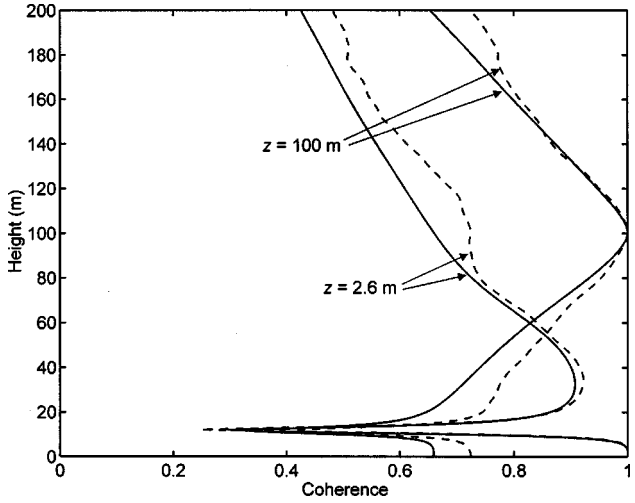


FIG. 12. Vertical normalized coherence for downward refraction at 2000 m range. Legend same as Fig. 5.

Unlike the current generation of PEs that perform deterministic, or ''snapshot'' calculations, an efficient second-moment PE would provide straightforward calculation of the mean received sound energy. The second-moment PE could be made into a very practical and valuable tool for propagation calculations if a transformation could be found that efficiently compresses the second-moment matrix. Cholesky factorization as considered in this article may serve as a guideline.

ACKNOWLEDGMENTS

D. K. Wilson was supported in part through the U.S. Army Research Laboratory Director's Research Initiative Program. V. E. Ostashev was supported by the U.S. Army Research Office under contract number DAAG55-98-1-0463 and by an NRC-ETL Research Associateship.

APPENDIX A: DERIVATION OF A CLOSED EQUATION FOR THE MOMENTS OF A SOUND FIELD

An equation for the arbitrary statistical moment of the wavefield, $\Gamma_{n,m}$, is derived in this Appendix. The derivation is similar to that of Rytov *et al.*,[8] who considered line-of-sight electromagnetic wave propagation through homogeneous and isotropic turbulence. Here we generalize their derivation to the case of sound propagation through a refractive medium with inhomogeneous and anisotropic turbulence statistics.

First, we will derive an equation for the function g given by

$$g(x;\mathbf{r}_1,\ldots,\mathbf{r}_n;\mathbf{r}_1',\ldots,\mathbf{r}_m') = p(x,\mathbf{r}_1)\cdots p(x,\mathbf{r}_n)p^*(x,\mathbf{r}_1')\cdots p^*(x,\mathbf{r}_m'). \quad \text{(A1)}$$

The initial term on the right-hand side of this equation, $p(x,\mathbf{r}_1)$, satisfies Eq. (1):

$$2ik_0\frac{\partial p(x,\mathbf{r}_1)}{\partial x}+\hat{L}_{\mathbf{r}_1}p(x,\mathbf{r}_1)+k_0^2\varepsilon_{\mathrm{mov}}(x,\mathbf{r}_1)p(x,\mathbf{r}_1)=0. \tag{A2}$$

Next we multiply both sides of this equation by the function $p(x,\mathbf{r}_2)\cdots p(x,\mathbf{r}_n)p^*(x,\mathbf{r}_1')\cdots p^*(x,\mathbf{r}_m')$. Taking into account that this function does not depend on $\mathbf{r}_1$, we can write the resulting equation as

$$2ik_0\frac{\partial p(x,\mathbf{r}_1)}{\partial x}p(x,\mathbf{r}_2)\cdots p(x,\mathbf{r}_n)p^*(x,\mathbf{r}_1')\cdots p^*(x,\mathbf{r}_m')$$
$$+\hat{L}_{\mathbf{r}_1}g+k_0^2\varepsilon_{\mathrm{mov}}(x,\mathbf{r}_1)g=0. \tag{A3}$$

Equations for $\partial p(x,\mathbf{r}_i)/\partial x$, where $i=2,\ldots,n$, can be derived analogously. For example, the equation for $\partial p(x,\mathbf{r}_2)/\partial x$ has the form

$$2ik_0p(x,\mathbf{r}_1)\frac{\partial p(x,\mathbf{r}_2)}{\partial x}p(x,\mathbf{r}_3)\cdots p(x,\mathbf{r}_n)p^*(x,\mathbf{r}_1')\cdots$$
$$p^*(x,\mathbf{r}_m')+\hat{L}_{\mathbf{r}_2}g+k_0^2\varepsilon_{\mathrm{mov}}(x,\mathbf{r}_2)g=0. \tag{A4}$$

The equation for $p^*(x,\mathbf{r}_1')$ can be obtained by complex conjugation of Eq. (1):

$$2ik_0\frac{\partial p^*(x,\mathbf{r}_1')}{\partial x}-\hat{L}^*_{\mathbf{r}_1'}p^*(x,\mathbf{r}_1')-k_0^2\varepsilon_{\mathrm{mov}}(x,\mathbf{r}_1')p^*(x,\mathbf{r}_1')=0. \tag{A5}$$

We multiply this equation by the function $p(x,\mathbf{r}_1)\cdots p(x,\mathbf{r}_n)p^*(x,\mathbf{r}_2')\cdots p^*(x,\mathbf{r}_m')$. Because this function does not depend on $\mathbf{r}_1'$, the resulting equation has the form

$$2ik_0p(x,\mathbf{r}_1)\cdots p(x,\mathbf{r}_n)\frac{\partial p^*(x,\mathbf{r}_1')}{\partial x}p^*(x,\mathbf{r}_2')\cdots p^*(x,\mathbf{r}_m')$$
$$-\hat{L}^*_{\mathbf{r}_1'}g-k_0^2\varepsilon_{\mathrm{mov}}(x,\mathbf{r}_1')g=0. \tag{A6}$$

Equations for $\partial p^*(x,\mathbf{r}_j')/\partial x$, where $j=2,\ldots,n$, can be derived analogously.

Adding together all equations for $\partial p(x,\mathbf{r}_i)/\partial x$, where $i=1,\ldots,n$, and all equations for $\partial p^*(x,\mathbf{r}_j')/\partial x$, where $j=1,\ldots,m$, we obtain the following result for g:

$$2ik_0\frac{\partial g(x)}{\partial x}+\hat{M}(x)g(x)+k_0^2\nu(x)g(x)=0, \tag{A7}$$

where the operator $\hat{M}$ is given by

$$\hat{M}(x;\mathbf{r}_1,\ldots,\mathbf{r}_n;\mathbf{r}_1',\ldots,\mathbf{r}_m')$$
$$=\hat{L}_{\mathbf{r}_1}+\cdots+\hat{L}_{\mathbf{r}_n}-\hat{L}^*_{\mathbf{r}_1'}-\cdots-\hat{L}^*_{\mathbf{r}_m'}, \tag{A8}$$

and the random variable ν is

$$\nu(x;\mathbf{r}_1,\ldots,\mathbf{r}_n;\mathbf{r}_1',\ldots,\mathbf{r}_m')$$
$$=\varepsilon_{\mathrm{mov}}(x,\mathbf{r}_1)+\cdots+\varepsilon_{\mathrm{mov}}(x,\mathbf{r}_n)-\varepsilon_{\mathrm{mov}}(x,\mathbf{r}_1')$$
$$-\cdots-\varepsilon_{\mathrm{mov}}(x,\mathbf{r}_m'). \tag{A9}$$

For simplicity, in Eq. (A7) and some others to follow, the functional arguments $\mathbf{r}_i$ and $\mathbf{r}_j'$ are omitted.

Equation (A7) can be written in the following equivalent integral form:

$$2ik_0g(x)-2ik_0\exp\left(\frac{ik_0}{2}\int_0^x\nu(\xi)d\xi\right)g(x=0)$$
$$=-\int_0^x\exp\left(\frac{ik_0}{2}\int_{x'}^x\nu(\xi)d\xi\right)\hat{M}(x')g(x')dx'. \tag{A10}$$

Indeed, multiplying both sides of Eq. (A10) by $\exp[-(ik_0/2)\int_0^x\nu(\xi)d\xi]$ and then differentiating with respect to x, one obtains Eq. (A7). In the integrand of Eq. (A10), $\exp[(ik_0/2)\int_{x'}^x\nu(\xi)d\xi]$ and $g(x')$ do not correlate, since $\varepsilon_{\mathrm{mov}}$ is a δ-correlated random field along the x-axis [see Eq. (5)]. Therefore, averaging both sides of Eq. (A10) yields:

$$2ik_0\langle g(x)\rangle-2ik_0K(x,0)g(x=0)$$
$$=-\int_0^xK(x,x')\hat{M}(x')\langle g(x')\rangle dx'. \tag{A11}$$

Here, $K(x,x')=\langle\exp[(ik_0/2)\int_{x'}^x\nu(\xi)d\xi]\rangle$. Assuming $\varepsilon_{\mathrm{mov}}$ is a random Gaussian field, the expression in the exponent of the formula $K(x,x')$ is a Gaussian random variable. Using the formula $\langle e^{\eta}\rangle=e^{\langle\eta\rangle^2/2}$ (valid for a zero-mean, Gaussian random variable η), one obtains

$$K(x,x')=\exp\left[-\frac{k_0^2}{8}\int_{x'}^x\int_{x'}^x\langle\nu(\xi_1)\nu(\xi_2)\rangle d\xi_1d\xi_2\right]. \tag{A12}$$

From Eqs. (A9) and (5), it can be shown that the statistical moment $\langle\nu(\xi_1)\nu(\xi_2)\rangle=\delta(\xi_1-\xi_2)Q_{n,m}(\xi_1)$, where

$$Q_{n,m}(x;\mathbf{r}_1,\ldots,\mathbf{r}_n;\mathbf{r}_1',\ldots,\mathbf{r}_m')$$
$$=\sum_{i=1}^n\sum_{j=1}^n b_{\mathrm{eff}}(x;\mathbf{r}_i;\mathbf{r}_j)+\sum_{i=1}^m\sum_{j=1}^m b_{\mathrm{eff}}(x;\mathbf{r}_i';\mathbf{r}_j')$$
$$-2\sum_{i=1}^n\sum_{j=1}^m b_{\mathrm{eff}}(x;\mathbf{r}_i;\mathbf{r}_j'). \tag{A13}$$

We now substitute the value of $\langle\nu(\xi_1)\nu(\xi_2)\rangle$ into Eq. (A12), and then substitute the value of $K(x,x')$ into Eq. (A11). We also take into account that, as follows from Eqs. (7) and (A1), $\langle g(x)\rangle=\Gamma_{n,m}(x)$. As a result, Eq. (A11) takes the form

$$2ik_0\Gamma_{n,m}(x)-2ik_0\exp\left[-\frac{k_0^2}{8}\int_0^xQ_{n,m}(\xi)d\xi\right]\Gamma_{n,m}(x=0)$$
$$=-\int_0^x\exp\left[-\frac{k_0^2}{8}\int_{x'}^xQ_{n,m}(\xi)d\xi\right]\hat{M}(x')\Gamma_{n,m}(x')dx'. \tag{A14}$$

Multiplying both sides of this equation by $\exp[(k_0^2/8)\int_0^xQ_{n,m}(\xi)d\xi]$ and differentiating with respect to x, one obtains

$$\frac{\partial \Gamma_{n,m}(x)}{\partial x}+\frac{k_0^2}{8}Q_{n,m}(x)\Gamma_{n,m}(x)-\frac{i}{2k_0}\hat{M}(x)\Gamma_{n,m}(x)=0. \tag{A15}$$

Substitution of the values of $\hat{M}$ and $Q_{n,m}$ into Eq. (A15) results in Eq. (8) for $\partial\Gamma_{n,m}/\partial x$.

APPENDIX B: DERIVATION OF BOUNDARY CONDITIONS FOR THE MOMENTS

Boundary conditions for $\Gamma_{n,m}$ are derived in this Appendix. We begin by writing Eq. (4) in the form

$$\left.\frac{\partial p(x,y_1,z_1)}{\partial z_1}\right|_{z_1=0}=-ik_0\beta p(x,y_1,z_1)|_{z_1=0}. \tag{B1}$$

Multiplying both sides of this equation by $p(x,\mathbf{r}_2)\cdots p(x,\mathbf{r}_n)p^*(x,\mathbf{r}_1')\cdots p^*(x,\mathbf{r}_m')$ and averaging the resulting equation, one obtains

$$\left.\frac{\partial \Gamma_{n,m}(x;\mathbf{r}_1,\ldots,\mathbf{r}_n;\mathbf{r}_1',\ldots,\mathbf{r}_m')}{\partial z_1}\right|_{z_1=0}$$

$$=-ik_0\beta\Gamma_{n,m}|_{z_1=0}. \tag{B2}$$

This equation coincides with Eq. (10), where $i=1$. Analogously, we can obtain equations for the boundary conditions at $z_2=0,\ldots,z_n=0$. These equations coincide with Eq. (10) for $i=2,\ldots,n$.

Equation (4) can also be written in the following form:

$$\left.\frac{\partial p^*(x,y_1',z_1')}{\partial z_1'}\right|_{z_1'=0}=ik_0\beta^* p^*(x,y_1',z_1')|_{z_1'=0}. \tag{B3}$$

We multiply both sides of Eq. (B3) by $p(x,\mathbf{r}_1)\cdots p(x,\mathbf{r}_n)p^*(x,\mathbf{r}_2')\cdots p^*(x,\mathbf{r}_m')$ and then average the resulting equation, arriving at

$$\left.\frac{\partial \Gamma_{n,m}(x;\mathbf{r}_1,\ldots,\mathbf{r}_n;\mathbf{r}_1',\ldots,\mathbf{r}_m')}{\partial z_1'}\right|_{z_1'=0}$$

$$=ik_0\beta^*\Gamma_{n,m}|_{z_1'=0}. \tag{B4}$$

This equation is essentially the same as Eq. (11), with $j=1$. The equations for the boundary conditions at $z_2'=0,\ldots,z_m'=0$ can be obtained analogously. They coincide with Eq. (11) for $j=2,\ldots,m$.

Our numerical implementation of the boundary conditions (B2) and (B4) is analogous to the second-order scheme of West *et al.*[14] For example, the finite-difference approximation of the derivative in Eq. (B2) is

$$\left.\frac{\partial \Gamma_{n,m}(x;\mathbf{r}_1,\ldots,\mathbf{r}_n;\mathbf{r}_1',\ldots,\mathbf{r}_m')}{\partial z_1}\right|_{z_1=0}$$

$$=\frac{3}{2\Delta z}(\Gamma_{n,m}|_{z_1=\Delta z}-\Gamma_{n,m}|_{z_1=0})$$

$$-\frac{1}{2\Delta z}(\Gamma_{n,m}|_{z_1=2\Delta z}-\Gamma_{n,m}|_{z_1=\Delta z}). \tag{B5}$$

The following boundary condition results:

$$\Gamma_{n,m}|_{z_1=0}=\sigma_1\Gamma_{n,m}|_{z_1=\Delta z}+\sigma_2\Gamma_{n,m}|_{z_1=2\Delta z}, \tag{B6}$$

where

$$\sigma_1=\frac{4}{3-2ik_0\beta\Delta z}\quad\text{and}\quad\sigma_2=\frac{1}{3-2ik_0\beta\Delta z}. \tag{B7}$$

APPENDIX C: SOLUTION OF THE SECOND-MOMENT EQUATION BASED ON MATRIX FACTORIZATION

An alternative method for solving the second-moment equation, based on Cholesky factorization of the second-moment matrix, is described in this Appendix. Hermitian matrices such as $\mathbf{P}$ in Sec. III B possesses a Cholesky factorization, such that $\mathbf{P}=\tilde{\mathbf{C}}\mathbf{C}$, where $\mathbf{C}$ is an upper triangular matrix.[16] Substituting the factorization into Eq. (32), we have

$$\tilde{\mathbf{C}}\frac{\partial\mathbf{C}}{\partial x}+\frac{\partial\tilde{\mathbf{C}}}{\partial x}\mathbf{C}=\mathbf{E}\tilde{\mathbf{C}}\mathbf{C}+\tilde{\mathbf{C}}\mathbf{C}\tilde{\mathbf{E}}-\mathbf{H}\odot(\tilde{\mathbf{C}}\mathbf{C}). \tag{C1}$$

The two terms on the left side, as well as the first two terms on the right, are conjugate/transposes of one another. Since $\mathbf{H}$ is a real, symmetric matrix, the third term on the right is Hermitian. Therefore in lieu of Eq. (C1) we can solve

$$\frac{\partial\tilde{\mathbf{C}}}{\partial x}\mathbf{C}=\mathbf{E}\tilde{\mathbf{C}}\mathbf{C}-\frac{1}{2}\mathbf{H}\odot(\tilde{\mathbf{C}}\mathbf{C}). \tag{C2}$$

Post-multiplying by $\mathbf{C}^{-1}$, we have

$$\frac{\partial\tilde{\mathbf{C}}}{\partial x}=\mathbf{E}\tilde{\mathbf{C}}-\frac{1}{2}[\mathbf{H}\odot(\tilde{\mathbf{C}}\mathbf{C})]\mathbf{C}^{-1}. \tag{C3}$$

If the second term on the right were not present, this equation would be a linear tridiagonal system and could be solved in a manner entirely analogous to the first-moment equation. Unfortunately, the second term on the right, being nonlinear in $\tilde{\mathbf{C}}$, complicates matters considerably. To circumvent this problem, we can calculate the effects of the first (refractive) and second (random scattering) terms separately. The result is the following algorithm, performed at each range step:

(1) Find the Cholesky factorization $\mathbf{C}(x_j)$ of $\mathbf{P}(x_j)$ using a standard algorithm.
(2) Apply a tridiagonal solver to

$$\left(\mathbf{I}_N-\frac{\Delta x}{2}\mathbf{E}\right)\tilde{\mathbf{C}}(x_{j+1})=\left(\mathbf{I}_N+\frac{\Delta x}{2}\mathbf{E}\right)\tilde{\mathbf{C}}(x_j). \tag{C4}$$

$\mathbf{C}(x_{j+1})$ at this stage includes the effects of refraction and diffusion only. (It is also no longer an upper triangular matrix.)
(3) Re-form the second-moment matrix: $\hat{\mathbf{P}}(x_{j+1})=\tilde{\mathbf{C}}(x_{j+1})\mathbf{C}(x_{j+1})$.
(4) Now incorporate the coherence reduction from random scattering (i.e., the phase screen), by solving Eq. (32) without the refractive terms:

$$\mathbf{P}(x_{j+1})=\hat{\mathbf{P}}(x_{j+1})\odot\exp(-\mathbf{H}\Delta x). \tag{C5}$$

The preceding algorithm was found to yield results virtually identical to the one with iterative refinement in Sec.

III B. It has the advantage of not depending on an iterative convergence. It also works with a range step of 1/10 wavelength, five times longer than the iterative algorithm. Nonetheless, the preceding algorithm was several times slower for the example problems in Sec. IV than the iterative one. The main reason is that both the Cholesky factorization (step 1) and the $N \times N$ matrix multiplication (step 3) involve $O(N^3)$ floating point multiplications, and hence are very time consuming when N is large. The steps in the iterative algorithm are all $O(N^2)$ or less.

Other factorizations besides Cholesky can be used to produce the linear tridiagonal system. For example, setting $\mathbf{P}=\tilde{\mathbf{V}}\mathbf{V}$, where $\mathbf{V}$ is a matrix whose columns are the eigenvectors of $\mathbf{P}$ scaled by the square roots of the eigenvalues (also known as the empirical orthogonal functions), will accomplish the same goal. The Cholesky factorization has the advantage of computational efficiency, though, in comparison to the eigenvalue computation.

One possible approach to solving the second-moment equation involves a reduced approximate factorization limited to a few dominant eigenvalues. Initially, the matrix $\mathbf{P}$ has just one nonzero eigenvalue. We found that even at long horizontal distances from the source most of the energy ($\gtrsim$99%) is stored in just the first five or so dominant eigenvalues. We attempted a solution based on propagating only a few eigenvectors corresponding to the largest eigenvalues, calculated using the basic power method.[17] The main shortcoming was that these eigenvectors did not incorporate most of the energy scattered into the shadow zone. This can be understood from the fact that the energy scattered into the shadow is small in comparison to the directly illuminated field, having low mutual coherence with the direct field.

[1] K. E. Gilbert, R. Raspet, and X. Di, ''Calculation of turbulence effects in an upward refracting atmosphere,'' J. Acoust. Soc. Am. **87**, 2428–2437 (1990).

[2] P. Chevret, P. Blanc-Benon, and D. Juvé, ''A numerical model for sound propagation through a turbulent atmosphere near the ground,'' J. Acoust. Soc. Am. **100**, 3587–3599 (1996).

[3] X. Di and K. E. Gilbert, ''Wave propagation in a 3-D turbulent atmosphere: horizontal coherence,'' in *Eighth International Symposium on Long Range Sound Propagation* (Pennsylvania State University, State College, PA, 1998), pp. 169–180.

[4] J. X. Zhou, X. Z. Zhang, and P. H. Rogers, ''Resonant interaction of sound wave with internal solitons in the coastal zone,'' J. Acoust. Soc. Am. **90**, 2042–2054 (1991).

[5] D. Tielbürger, S. Finette, and S. Wolf, ''Acoustic propagation through an internal wave field in a shallow water waveguide,'' J. Acoust. Soc. Am. **101**, 789–808 (1997).

[6] V. I. Tatarskii, *The Effects of the Turbulent Atmosphere on Wave Propagation* (Keter, Jerusalem, 1971).

[7] A. Ishimaru, *Wave Propagation and Scattering in Random Media* (Academic, New York, 1978).

[8] S. M. Rytov, Y. A. Kravtzov, and V. I. Tatarskii, *Principles of Statistical Radio Physics, Part 4: Wave Propagation Through Random Media* (Springer, Berlin, 1989).

[9] V. E. Ostashev, *Acoustics in Moving Inhomogeneous Media* (E&FN Spon, London, 1997).

[10] A. A. Puzenko and E. V. Chaevskii, ''Mutual coherence function for glancing propagation of waves above an interface in a statistically inhomogeneous medium,'' Radiophys. Quantum Electron. **19**, 158–166 (1977).

[11] R. J. Hill, ''A stochastic parabolic wave equation and field-moment equations for random media having spatial variation of mean refractive index,'' J. Acoust. Soc. Am. **77**, 1742–1753 (1985).

[12] B. J. Bates and S. M. Bates, ''Stochastic simulation and first-order multiple scatter solutions for acoustic propagation through oceanic internal waves,'' J. Acoust. Soc. Am. **82**, 2042–2050 (1987).

[13] E. Mandine and M. C. Pelissier, ''Direct stochastic treatment of electromagnetic propagation problems in a random range-depending medium, using a PE model,'' in *Proceedings of the 1996 Battlespace Atmospherics Conference*, edited by J. H. Richter and K. D. Anderson (NCCOSC, San Diego, CA, 1996), pp. 113–121.

[14] M. West, K. Gilbert, and R. A. Sack, ''A tutorial on the parabolic equation (PE) model used for long range sound propagation in the atmosphere,'' Appl. Acoust. **37**, 31–49 (1992).

[15] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes: The Art of Scientific Computing* (Cambridge U. P., Cambridge, 1986).

[16] The MathWorks, Inc., *Using MATLAB* (MathWorks, Natick, MA, 1999).

[17] R. L. Burden, J. D. Faires, and A. C. Reynolds, *Numerical Analysis*, 2nd ed. (Prindle, Weber & Schmidt, Boston, 1981).

[18] E. M. Salomons, ''Improved Green's function parabolic equation method for atmospheric sound propagation,'' J. Acoust. Soc. Am. **104**, 100–111 (1998).

[19] 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 Hoof, ''Benchmark cases for outdoor sound propagation models,'' J. Acoust. Soc. Am. **97**, 173–191 (1995).

[20] D. K. Wilson, ''Relaxation models for the acoustical properties of sound-absorbing porous materials,'' in *Proceedings of NOISE-CON 97*, edited by S. D. Sommerfeldt (The Pennsylvania State University, University Park, PA, 1997), pp. 85–90.

[21] K. Attenborough, ''Acoustical impedance models for outdoor ground surfaces,'' J. Sound Vib. **99**, 521–544 (1985).

[22] The function b_{eff} is 8π times the two-dimensional correlation function $b_{\|}$ appearing in earlier papers by Wilson.[22,24] This factor is strictly a matter of convention, arising from the placement of $1/2\pi$ in front of the integral in Eq. (6), and from using the effective sound speed fluctuation μ to define the correlation function instead of $\varepsilon_{\text{mov}}=2\mu$.

[23] D. K. Wilson, ''Performance bounds for acoustic direction-of-arrival arrays operating in atmospheric turbulence,'' J. Acoust. Soc. Am. **103**, 1306–1319 (1998).

[24] V. E. Ostashev, B. Brähler, V. Mellert, and G. H. Goedecke, ''Coherence functions of plane and spherical waves in a turbulent medium with the von Karman spectrum of medium inhomogeneities,'' J. Acoust. Soc. Am. **104**, 727–737 (1998).

[25] D. K. Wilson, ''A turbulence spectral model for sound propagation in the atmosphere that incorporates shear and buoyancy forcings,'' J. Acoust. Soc. Am. **108**, 2021–2038 (2000).

Spectral broadening of sound scattered by advecting atmospheric turbulence

G. H. Goedecke and Roy C. Wood[a)]
Department of Physics, New Mexico State University, Las Cruces, New Mexico 88003-0001

Harry J. Auvermann
U.S. Army Research Laboratory, Adelphi, Maryland 20783-1155

V. E. Ostashev
NOAA/Environmental Technology Laboratory, 325 Broadway, Boulder, Colorado 80303 and
Department of Physics, New Mexico State University, Las Cruces, New Mexico 88003-0001

David I. Havelock
Institute for Microstructural Sciences, National Research Council, Ottawa, Ontario K1A OR6, Canada

Chueh Ting
Klipsch School of Electrical and Computer Engineering, New Mexico State University, Las Cruces, New Mexico 88003-0001

(Received 19 August 1999; revised 14 February 2001; accepted 15 February 2001)

Scattering and spectral broadening of a monochromatic sound wave by atmospheric turbulence that is flowing with a uniform constant horizontal wind is considered. The acoustic source and a detector are at rest and at different positions in a ground-fixed frame. Analytic expressions are derived for the sound pressure scattered to the detector by a single eddy. Since distances and the scattering angle change with time as the eddy flows through the scattering volume, the detector signal has time-dependent amplitude and frequency, for which general formulas are derived. A computer code is developed that calculates the scattered signal and its Fourier transform from a single eddy, or from a steady-state collection of eddies of many different scale lengths that represents isotropic homogeneous turbulence flowing with the wind. The code utilizes a time-shift algorithm that reduces the calculation time substantially. Several numerical results from this code are presented, including simulations of a recent experiment. The predicted spectral shape, including peak width and jaggedness, are in good agreement with experiment.
[DOI: 10.1121/1.1364489]

PACS numbers: 43.28.Py, 43.20.Fn [LCS]

I. INTRODUCTION

Spectral broadening of sound scattered by atmospheric turbulence has been studied both theoretically[1–8] and experimentally[9–11] for many years. It is important in several applications, including remote sounding of the atmosphere and other Doppler sodar and sonar applications,[3–6,12] and in understanding jet noise.[7]

The shape of the frequency spectrum of sound scattered by turbulence depends on several factors.[6] These factors include the length scale L of the scattering volume and its distances (r_S, r_D) from source and detector, the mean wind speed w, the intrinsic time dependence of the turbulence in coordinate frames moving with the mean wind, the scattering properties of the turbulence, the directivity patterns of the source and detector, and the duration and time dependence of the source pulse. In this work, we consider a continuous monochromatic source wave, so the last factor is irrelevant.

Most authors treat the limiting case $L \ll r_S,\ r_D$,[1–8] which usually results from assuming narrow directivity patterns of the source and detector. In this limit, the spectral broadening due to the flow of the turbulence with a mean wind is negligible compared to that due to the intrinsic time dependence.[4] Also, many of these authors consider only monostatic (coincident) source/detector locations, or only backward and forward scattering.[3–6]

In this paper we treat a different limiting case, namely, bistatic (noncoincident) source–detector locations, and $L \gg r_S,\ r_D$, with wide-angle directivity patterns for the source and detector. In this limit, the spectral broadening due to turbulent eddies flowing with the mean wind across the scattering volume is much greater than that due to their intrinsic time dependence,[4,8] and we neglect the latter. Our goal is to compare our calculated frequency spectrum with an experimental one for an experiment that fits the limiting case quite well.[11]

Neglect of intrinsic time dependence means that we are using Taylor's frozen turbulence model. We construct the model turbulent temperature and velocity fluctuation fields via superposition of self-similar localized eddies or quasi-wavelets of many sizes, time independent but randomly located and oriented in the frame comoving with the mean wind. In that frame, this is the eddy or turbule model of isotropic homogeneous turbulence developed by Goedecke

[a)]Current address: Department of Chemistry and Physics, Armstrong Atlantic State University, Savannah, GA 31419-1997.

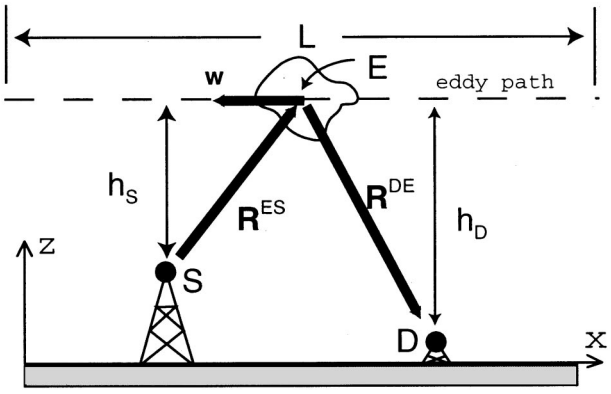


FIG. 1. Model geometry.

and Auvermann.[13] This eddy model allows us to calculate and understand easily the general shape, width, and jaggedness of the above-mentioned experimental frequency spectrum.

In Sec. II, we present our model. In Sec. III, we present several results of our calculations and show that our model predictions agree well with some experimental data. In Sec. IV, we provide a summary and discussion in which we suggest other experiments that are needed to ascertain the conditions under which a frozen turbulence model is adequate to predict observed spectral broadening in sound waves scattered by atmospheric turbulence. We also indicate how our model could be modified to account for anisotropy, inhomogeneity, and intermittency of atmospheric turbulence.

II. MODEL AND THEORY

A. Geometry

The geometry of our model is defined by Fig. 1, which is a projection onto the $(x-z)$ plane. The letters (S,D,E) stand for (source, detector, eddy), respectively. Coordinate z is vertical, $z=0$ is the ground plane, and coordinates x and y are horizontal, ground fixed; y is into the paper. The uniform constant horizontal wind velocity is $\mathbf{w}$, given by

$$\mathbf{w}=w(\mathbf{e}_x\cos\varphi_w+\mathbf{e}_y\sin\varphi_w), \tag{1}$$

where $(\mathbf{e}_x,\mathbf{e}_y,\mathbf{e}_z)$ are the Cartesian unit basis vectors, $w=|\mathbf{w}|$ is the wind speed, and φ_w is the azimuthal angle of the wind; w and φ_w are constants. The source S and detector D are at locations given by position vectors $(\mathbf{s},\mathbf{d})$, respectively, fixed with respect to the ground-fixed reference frame. The location of the center of an eddy is given by

$$\mathbf{R}^E(t)=\overline{\mathbf{b}}+\mathbf{w}t, \tag{2}$$

so $\overline{\mathbf{b}}$ is the location of the eddy at $t=0$ in the ground-fixed frame.

The vectors $\mathbf{R}^{ES}$ and $\mathbf{R}^{DE}$ are the time-dependent vectors from source to eddy and eddy to detector, respectively, defined by

$$\mathbf{R}^{ES}(t)\equiv\mathbf{R}^E(t)-\mathbf{s},\quad \mathbf{R}^{DE}(t)=\mathbf{d}-\mathbf{R}^E(t). \tag{3}$$

The comoving (barred) frame is defined by the Galilean coordinate transformation

$$\overline{x}=x-w_xt,\quad \overline{y}=y-w_yt,\quad \overline{z}=z, \tag{4}$$

so that the origins coincide at $t=0$. In this frame, the eddy, source, and detector locations are given, respectively, by

$$\overline{\mathbf{R}}^E=\overline{\mathbf{b}},\quad \overline{s}(t)=\mathbf{s}-\mathbf{w}(t),\quad \overline{\mathbf{d}}(t)=\mathbf{d}-\mathbf{w}t, \tag{5}$$

so that

$$\overline{\mathbf{R}}^{ES}(t)=\overline{\mathbf{R}}^E-\overline{s}(t)=\overline{\mathbf{b}}-(\mathbf{s}-\mathbf{w}t)=\mathbf{R}^{ES}(t), \tag{6}$$

$$\overline{\mathbf{R}}^{DE}(t)=\overline{\mathbf{d}}(t)-\overline{\mathbf{R}}^E=\mathbf{d}-\mathbf{w}t-\overline{\mathbf{b}}=\mathbf{R}^{DE}(t). \tag{7}$$

The reference temperature and the mean wind are chosen to be independent of z, with the wind horizontal, for two reasons. First, these were the conditions in the experiments to which we refer for comparisons, except for the region within 10 meters of the ground. Second, restriction to frozen turbulence and uniformity of the background allows analytic solutions that are valid for all wind Mach numbers less than 1.

B. Acoustic equations

In the comoving (barred) frame, defined by Eq. (4), the relevant acoustic equations are the same as those given by Tatarskii[14] for the ground-fixed frame with no wind

$$\partial_t\overline{v}_i+\rho_0^{-1}\overline{\partial}_i\overline{p}=-(\overline{T}_F/\rho_0T_0)\overline{\partial}_i\overline{p}-\overline{v}_j\overline{\partial}_j\overline{v}_{Fi}-\overline{v}_{Fj}\overline{\partial}_j\overline{v}_i, \tag{8}$$

$$\partial_t\overline{p}+\gamma p_0\overline{\partial}_i\overline{v}_i=-\overline{v}_{Fi}\overline{\partial}_i\overline{p}+\overline{S}. \tag{9}$$

In these equations $(\overline{p},\overline{v}_i)$ are the acoustic (pressure, velocity) fields; $(\overline{T}_F,\overline{v}_{Fi})$ are the turbulent (temperature, flow velocity) fluctuations; and (ρ_0,T_0,p_0) are the uniform constant reference (mass density, temperature, pressure), related by the ideal gas equation of state. Also, γ is the specific heat ratio c_p/c_v, which is assumed constant and uniform, equal to 1.4 for usual atmospheric conditions. We use summation notation, with Latin indices ranging and summing from 1 to 3 representing Cartesian coordinates, and the notation ∂_t, which stands for ∂/∂_t. Also, we have included a scalar acoustic source function $\overline{S}$, to be chosen.

As discussed in the Introduction, the turbulent temperature and velocity fluctuations are regarded as time independent (frozen) in the comoving frame, i.e., $\overline{T}_F(\overline{\mathbf{r}})$, $\overline{v}_{Fi}(\overline{\mathbf{r}})$ only, with no explicit time dependence. They may be stochastic functions. Thus, the description of both the turbulence and the acoustics in the comoving frame is the same as the description used by Goedecke and Auvermann[13] in a reference frame with no wind, which is equivalent to the standard description.

We write the acoustic quantities as the sum of the waves created by the source and the waves scattered by the turbulence

$$\overline{p}=\overline{p}_e+\overline{p}_s,\quad \overline{v}_i=\overline{v}_{ei}+\overline{v}_{si}, \tag{10}$$

where the subscript e stands for ''external,'' i.e., for the waves created by the source that act as externally applied fields on the turbulence, and the subscript s stands for ''scattered,'' representing waves scattered by the turbulence.

The external acoustic pressure and velocity waves in the barred frame, due to a scalar source, satisfy Eqs. (8) and (9) with $(\bar{v}_{Fi}, \bar{T}_F) = 0$

$$\partial_t \bar{p}_e + \gamma p_0 \bar{\partial}_i \bar{v}_{ei} = \bar{S}(\bar{\mathbf{r}}, t), \tag{11}$$

$$\partial_t \bar{v}_{ei} + \rho_0^{-1} \bar{\partial}_i \bar{p}_e = 0. \tag{12}$$

According to Eq. (12), we may write

$$\bar{v}_{ei} = \bar{\partial}_i \bar{\Phi} \rightarrow \bar{p}_e = -\rho_0 \partial_t \bar{\Phi}, \tag{13}$$

where $\bar{\Phi}$ is a ''velocity potential.'' Then from Eq. (11), we get

$$(\nabla^2 - c^{-2}\partial_t^2)\bar{\Phi} = \rho_0^{-1} c^{-2} \bar{S}, \tag{14}$$

where $c = (\gamma p_0/\rho_0)^{1/2}$ is the adiabatic sound speed.

In the unbarred frame, the monochromatic isotropic point source is located at $\mathbf{r} = \mathbf{s}$ and has frequency $\nu_s = \omega_s/2\pi$, so we may write

$$S(\mathbf{r}, t) = A' \delta(\mathbf{r} - \mathbf{s}) e^{-i\omega_s t} + \text{c.c.}$$

$$= A' \delta(\bar{\mathbf{r}} - \mathbf{s} + \mathbf{w}t) e^{-i\omega_s t} + \text{c.c.} = \bar{S}(\bar{\mathbf{r}}, t), \tag{15}$$

where A' is a complex amplitude to be interpreted. Here, and in what follows, ''c.c.'' stands for ''complex conjugate of the preceding term.'' The causal Green function solutions of Eqs. (13)–(15) for $\bar{\Phi}$, and thus for $\bar{p}_e$ and $\bar{v}_{ei}$, valid to all orders of the mean wind Mach number $\beta = w/c$, are obtained in Appendix A.

In the Born approximation, the scattered fields $(\bar{p}_s, \bar{v}_{si})$ satisfy Eqs. (8) and (9) with $\bar{S} \equiv 0$ and $(\bar{p}_e, \bar{v}_{ei})$ on the right-hand sides, and with $(\bar{p}_s, \bar{v}_{si})$ on the left-hand sides. Combining these equations to eliminate $\bar{\partial}_i \bar{v}_{si}$, and applying the solenoidal condition $\bar{\partial}_i \bar{v}_{Fi} = 0$ to the turbulent velocity fluctuation, we get

$$(\bar{\partial}_i \bar{\partial}_i - c^{-2}\partial_t^2)\bar{p}_s = -T_0^{-1} \bar{\partial}_i (\bar{T}_F \bar{\partial}_i \bar{p}_e) - 2\rho_0 \bar{\partial}_i \bar{\partial}_j (\bar{v}_{Fj} \bar{v}_{ei}) \tag{16}$$

as the wave equation for the scattered acoustic pressure. In the derivation we used Eq. (11), setting $\bar{S} \equiv 0$ where the turbulent velocity field $\bar{v}_{Fi}$ is nonzero. Equation (16) is equivalent to the wave equation derived by Monin[15] as given by Tatarskii.[14]

The causal Green function solution of Eq. (16) valid to all orders of the Mach number β is also obtained in Appendix A. The solution is written as the sum

$$\bar{p}_s(\bar{\mathbf{r}}, t) = \bar{p}_s^T(\bar{\mathbf{r}}, t) + \bar{p}_s^v(\bar{\mathbf{r}}, t), \tag{17}$$

where $(\bar{p}_s^T, \bar{p}_s^v)$ satisfy Eq. (16) with only the first and second terms on the right-hand side, respectively, so they are the acoustic pressure waves scattered by the turbulent (temperature, velocity) fluctuations.

C. Scattered acoustic pressure at detector

We seek expressions for the acoustic pressure waves $(p_s^T(t), p_s^v(t))$ at the detector, scattered by the (temperature, velocity) fields of a localized turbulent eddy or ''turbule.'' For a turbulent eddy flowing with the wind, so that its given $(\bar{T}_F, \bar{v}_{Fj})$ are stationary and localized in the barred frame, $p_s^T(t)$ is given by the solution $\bar{p}_s^T(\bar{\mathbf{r}}, t)$ of Eq. (17), evaluated at the detector location $\bar{\mathbf{r}} = \bar{\mathbf{d}}(t) = \mathbf{d} - \mathbf{w}t$, and similarly for $p_s^v(t)$. These expressions are derived in some detail in Appendix A for any $\beta < 1$. Here, we quote only the results valid through first order in β. Reflections from the ground are ignored. The scattered wave $p_s^T(t)$ at the detector has the form

$$p_s^T(t) = \mathrm{A}^{\mathrm{T}}(t) e^{-i\omega_s t} + \text{c.c.}, \tag{18}$$

where $\mathrm{A}^{\mathrm{T}}(t)$ is a slowly varying complex-valued amplitude, given by Eq. (A17). To first order in β this is

$$\mathrm{A}^{\mathrm{T}}(t) = (\mathcal{P}_e a')(R^{ES} R^{DE})^{-1} [-(k_s^2/4\pi T_0) \times (\hat{\mathbf{R}}^{ES} \cdot \hat{\mathbf{R}}^{DE}) \bar{\mathcal{T}}(\mathbf{K})] e^{ik_s(R^{ES} + R^{DE})}, \tag{19}$$

where a hat over a vector indicates that it is a unit vector. Here, $k_s = \omega_s/c$, $\mathcal{P}_e$ is the pressure of the source wave at a distance a' from the source in a nonmoving atmosphere, and $\bar{\mathcal{T}}(\mathbf{K})$ is the spatial Fourier transform of $\bar{T}(\bar{\mathbf{r}} - \bar{\mathbf{b}})$, the stationary temperature variation of the turbule in the comoving frame, given by

$$\bar{\mathcal{T}}(\mathbf{K}) \equiv \int \bar{T}(\boldsymbol{\xi}) e^{-i\mathbf{K}\cdot\boldsymbol{\xi}} d^3\xi, \tag{20}$$

where $d^3\xi$ is the three-dimensional volume element. $(\mathbf{R}^{ES}(t), \mathbf{R}^{DE}(t))$ are given by Eqs. (6) and (7). The vector $\mathbf{K}$ is the scattering vector defined by Eq. (A15). For $\beta \ll 1$, it is given by

$$\mathbf{K} = k_s(\hat{\mathbf{R}}^{DE}(t) - \hat{\mathbf{R}}^{ES}(t)), \quad K = |\mathbf{K}| = 2k_s \sin(\theta(t)/2), \tag{21}$$

where $\theta(t)$ is the time-dependent scattering angle, defined here by

$$\cos\theta = \hat{\mathbf{R}}^{DE} \cdot \hat{\mathbf{R}}^{ES}. \tag{22}$$

The square bracket in Eq. (19) has exactly the form of the Monin[15] scattering amplitude for temperature scattering from stationary turbulence (see Goedecke and Auvermann[13]), as it should for $\beta \ll 1$, but A^{T} still retains time dependence in all its factors unless $\beta \equiv 0$.

Similarly, the contribution $p_s^v(t)$ scattered by the solenoidal turbulent velocity field of the turbule may be written

$$p_s^v(t) = \mathrm{A}^{\mathrm{v}}(t) e^{-i\omega_s t} + \text{c.c.}, \tag{23}$$

where the complex amplitude $\mathrm{A}^{\mathrm{v}}(t)$ is given by Eq. (A20). For $\beta \ll 1$, this reduces to

$$\mathrm{A}^{\mathrm{v}}(t) \equiv (\mathcal{P}_e a') \times \left\{ \frac{-k_s^2 (\hat{\mathbf{R}}^{ES} \cdot \hat{\mathbf{R}}^{DE})(\hat{\mathbf{R}}^{DE} \cdot \bar{\boldsymbol{\mathcal{V}}}(\mathbf{K}))}{(2\pi c)(R^{ES} R^{DE})} \right\} e^{ik_s(R^{ES} + R^{DE})}, \tag{24}$$

where $\bar{\boldsymbol{\mathcal{V}}}(\mathbf{K})$ is the spatial Fourier transform of the turbule velocity field $\bar{\mathbf{v}}(\bar{\mathbf{r}} - \bar{\mathbf{b}})$, stationary in the comoving (barred) frame, given by

$$\bar{\boldsymbol{\mathcal{V}}}(\mathbf{K}) \equiv \int e^{-i\mathbf{K}\cdot\boldsymbol{\xi}} \bar{\mathbf{v}}(\boldsymbol{\xi}) d^3\xi. \tag{25}$$

The expression in the braces in Eq. (24) is also the Monin result for the scattering amplitude due to turbulent velocity fluctuations, but with time-dependent factors, just as for $A^T(t)$. In order to obtain expressions for the spatial Fourier transforms $\bar{\mathcal{T}}(\mathbf{K})$ and $\bar{\boldsymbol{\mathcal{V}}}(\mathbf{K})$, we must use a specific model for turbulent eddies, described below.

D. Spectral broadening

Both p_s^T and p_s^v contain the factor

$$e^{-i\omega_s[t-(R^{ES}(t)+R^{DE}(t))/c]} = e^{-i\Phi(t)}, \tag{26}$$

where the equality defines the phase Φ. Consider an arbitrary instant of time t_0. Then, write for t near t_0

$$\Phi(t) \approx \Phi(t_0) + \dot{\Phi}(t_0)(t-t_0), \tag{27}$$

where we can neglect terms of higher order in $(t-t_0)$ if $R^{ES}(t)$ and $R^{DE}(t)$ are ''slowly varying,'' which is the case for $\beta \ll 1$. Thus, $\dot{\Phi}(t_0)$ is the effective angular frequency $\omega(t_0)$ at time t_0. Since t_0 is arbitrary, we have

$$\omega(t) = \omega_s[1 - c^{-1}(\dot{R}^{ES}(t) + \dot{R}^{DE}(t))]. \tag{28}$$

Using Eqs. (6) and (7), we get

$$\dot{R}^{ES}(t) = d_t(\mathbf{R}^{ES}\cdot\mathbf{R}^{ES})^{1/2} = \hat{\mathbf{R}}^{ES}\cdot\dot{\mathbf{R}}^{ES} = \mathbf{w}\cdot\hat{\mathbf{R}}^{ES}, \tag{29}$$

$$\dot{R}^{DE}(t) = d_t(\mathbf{R}^{DE}\cdot\mathbf{R}^{DE})^{1/2} = \hat{\mathbf{R}}^{DE}\cdot\dot{\mathbf{R}}^{DE} = -\mathbf{w}\cdot\hat{\mathbf{R}}^{DE}, \tag{30}$$

where $d_t \equiv d/dt$. Therefore, to order β,

$$\omega(t) \approx \omega_s[1 + \boldsymbol{\beta}\cdot(\hat{\mathbf{R}}^{DE}(t) - \hat{\mathbf{R}}^{ES}(t))]. \tag{31}$$

Note that this can be written in terms of the approximate scattering vector $\mathbf{K}(t)$ of Eq. (21) as

$$\omega(t) = \omega_s + \mathbf{w}\cdot\mathbf{K}(t), \tag{32}$$

which is actually valid for the exact $\mathbf{K}(t)$ given in Eq. (A15). Equation (31) is the same as that given in Ostashev,[16] to order β. It reveals that $\omega(t)$ can vary from $\omega_s(1+2\beta)$, when an eddy is very far horizontally from the source and detector and approaching, to $\omega_s(1-2\beta)$, when the eddy is far away and receding. Thus, the maximum Doppler broadening is $\pm 2\beta\omega_s$ around ω_s.

E. Turbule model

In this subsection, for clarity, we present some of the results and relations developed in Goedecke and Auvermann,[13] in a slightly different notation that is more suitable to this investigation. For convenience we choose self-similar localized turbulent eddies or ''turbules'' with Gaussian envelopes and spherical symmetry. The resulting model turbulence will be isotropic and homogeneous and will exhibit the Kolmogorov spectrum in the inertial range for these Gaussian envelopes *or for any others*. A temperature turbule of scale length or size a_α, centered at the fixed point $\bar{\mathbf{b}}$ in the comoving (barred) frame, is written similarly to a wavelet[17] form,

$$\bar{T}(\bar{\mathbf{r}}-\bar{\mathbf{b}}) = \delta T_\alpha \exp[-|\bar{\mathbf{r}}-\bar{\mathbf{b}}|^2/a_\alpha^2], \tag{33}$$

and a corresponding velocity turbule is written as

$$\bar{\mathbf{v}}(\bar{\mathbf{r}}-\bar{\mathbf{b}}) = [\Omega\times(\bar{\mathbf{r}}-\bar{\mathbf{b}})]\exp[-|\bar{\mathbf{r}}-\bar{\mathbf{b}}|^2/a_\alpha^2], \tag{34}$$

where α is a running scale length index. Note that $\boldsymbol{\Omega}$ is an angular velocity, and $\bar{\mathbf{v}}$ is solenoidal. For all turbules of scale length a_α, the amplitude δT_α is the same (within a $\pm$ sign, as discussed later), and the $\boldsymbol{\Omega}$ have the same magnitude, Ω_α. The amplitudes $(\delta T_\alpha, \Omega_\alpha)$ are scaled with a_α. Specifically, we designate a_1 as the size of the largest turbule to be used, the outer scale length for this model. Then, in accordance with conventional models,[18] we choose for any size a_α

$$\delta T_\alpha = (a_\alpha/a_1)^{1/3}\delta T_1, \quad \Omega_\alpha a_\alpha = (a_\alpha/a_1)^{1/3}\Omega_1 a_1. \tag{35}$$

This relates all the turbule amplitudes to the amplitude of the largest ones, which is chosen as discussed below. We must also choose the direction $\hat{\boldsymbol{\Omega}}$, as discussed below. From the definitions (20) and (25) and from Eqs. (33) and (34), we obtain the Fourier transforms of the temperature and velocity fluctuations for turbules of size α_α

$$\bar{\mathcal{T}}(\mathbf{K}) = \pi^{3/2}a_\alpha^3\delta T_\alpha \exp[-K^2a_\alpha^2/4], \tag{36}$$

$$\bar{\boldsymbol{\mathcal{V}}}(\mathbf{K}) = \frac{-i}{2}\pi^{3/2}a_\alpha^5\Omega_\alpha\hat{\boldsymbol{\Omega}}\times\mathbf{K}\exp[-K^2a_\alpha^2/4]. \tag{37}$$

Note that there is no velocity scattering if $\hat{\boldsymbol{\Omega}} \| \mathbf{K}$. From these expressions and Eqs. (18), (19), (21), and (22), we see that if $k_s a \gg 1$, then $\bar{\mathcal{T}}(\mathbf{K})$ and $\bar{\boldsymbol{\mathcal{V}}}(\mathbf{K})$ are extremely small unless the scattering angle θ is near 0°, so the scattering is negligible at large angles. But, for $k_s a \ll 1$ the Gaussian envelope is essentially unity for all scattering angles, so backward scattering is comparable to near-forward scattering.

For Kolmogorov spectra of isotropic homogeneous turbulence, the temperature and velocity structure parameters (C_T^2, C_v^2) are related to the turbule parameters by[13]

$$\begin{aligned} C_T^2 &= (1.89)(\phi/\mu)(\delta T_1/a_1^{1/3})^2\Gamma(11/6), \\ C_v^2 &= (0.34)(\phi/\mu)(\Omega_1 a_1^{2/3})^2\Gamma(17/6), \end{aligned} \tag{38}$$

and successively smaller turbule sizes a_α are related by

$$a_{\alpha+1}/a_\alpha = e^{-\mu} \equiv N_d^{-1}, \tag{39}$$

and $\phi = n_1 a_1 = n_\alpha a_\alpha$ is the packing fraction, the same for all sizes, n_α is the number density of turbules of size α_α, and Γ is the gamma function. The parameter $\mu = \ln N_d$ in Eqs. (38) and (39) is calculated after a_{N_s}, the smallest turbule size, and the number of sizes N_s in the cascade, are chosen:

$$a_{N_s}/a_1 = e^{-\mu(N_s-1)} \rightarrow \mu = -(N_s-1)^{-1}\ln(aN_s/a_1). \tag{40}$$

We must also choose values for the packing fraction ϕ and the parameters $(\delta T_1/T_0, \Omega_1 a_1/c)$.

For a model of homogeneous turbulence, we put one turbule of size a_α somewhere in each cubical cell of size d_α, where

$$d_\alpha = n_\alpha^{-1/3} = a_\alpha/\phi^{1/3}. \tag{41}$$

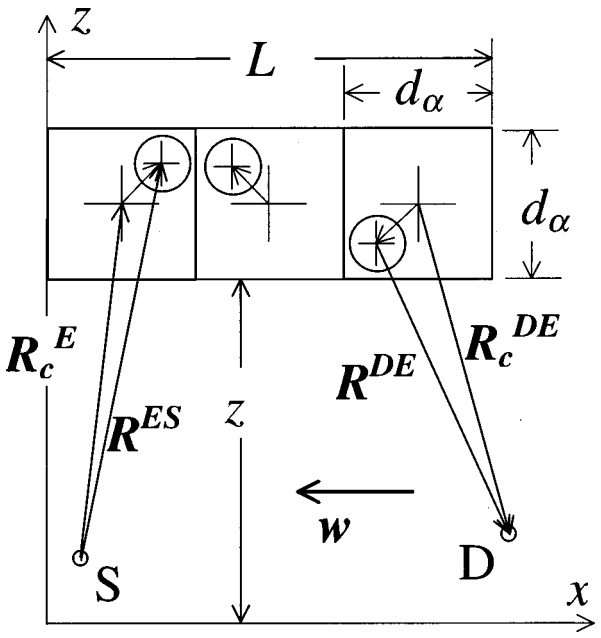


FIG. 2. Eddy tube at height z_b above ground, for wind $\mathbf{w}$ in $(-x)$ direction. The circles represent eddies, and the little arrows represent the random displacements $\boldsymbol{\eta}$ of eddy centers from cell centers.

For a typical $\phi \approx 10^{-3}$, $d_\alpha = 10a_\alpha$. If we had unlimited computer resources, we could populate a defined scattering volume V_s with turbules of different sizes, one randomly located in each cell for each size, and simply total the $(A^T(t), A^v(t))$ from each. This is actually what we do for the larger turbules with $k_s a \gtrsim 1$. But, if there are, say, 100 turbules with $a = 2$ m in V_s, then for $a = 2$ cm, according to Eq. (41) there will be $(100)(2/0.02)^3 = 10^8$ turbules, a truly formidable number.

Fortunately, we found a good approximation that we call the time-shift algorithm that reduces the computation time significantly for the smaller turbules. The algorithm is discussed briefly in Appendix B. Whether or not this algorithm is used for our model, the designated scattering volume has length L along the wind direction and, in the complete model, is filled with square tubes of side d_α and length L along that direction. Each tube contains one turbule per cell of side length d_α. Each cell and its turbule can be visualized as flowing with the wind, with a new cell entering the tube as an old one is exiting, to ensure steady state. This is done for each of the N_s scale lengths used. Figure 2 illustrates the turbule placements for one such tube at one scale length.

III. RESULTS

In this section we discuss results of numerical simulations based on the theory presented in the previous section. We also show how the theory easily predicts the major features of the computational results, and we compare simulation results to experimental results.

The values of the physical parameters were chosen to be good approximations to some of those in a recent experiment.[11] In particular, the source frequency ν_s was chosen to be 500 Hz, and the uniform adiabatic sound speed c to be 340 m/s, corresponding to a reference wavelength $\lambda = c/\nu_s = 0.68$ m. The detector was placed at $(-150,0,0)$ m, the source at $(150,0,10)$ m, and the wind velocity $\mathbf{w}$ was chosen with components (3.1,0,0) m/s (upwind propagation). The structure parameters were chosen to be $C_v^2 = 0.088\ \mathrm{m}^{4/3}/\mathrm{s}^2$, and $C_T^2 = 0.044(K)^2/\mathrm{m}^{2/3}$. These are within the ranges given in Ostashev[16] and Brown and Clifford.[19]

For multiturbule computations, each turbule angular velocity unit vector was generally chosen to have a uniformly random distribution in 4π solid angle, independently of other turbules, in accordance with the model of isotropic turbulence discussed in Goedecke and Auvermann.[13] The δT_a in the scattering volume V_s were generally all chosen to have the same sign, positive, because the temperature turbules are pockets of air warmer than the ambient temperatures that have risen from the ground due to buoyancy forces, such as those over a paved runway which was used in the relevant experiment. Over a larger volume, the mean temperature fluctuation should be zero. One calculation was made with the δT of each turbule either >0 or <0 with equal probability, in order to contrast the results of the different conditions. Also, some calculations were made with the angular velocity vectors of the larger turbules all in the $+y$ direction, for the same reason; this also introduces some reasonable anisotropy. Neither of these changes influenced the results significantly, probably because each was equivalent to introducing more randomness in turbule locations.

The other parameter values were chosen to minimize computation time while still maintaining adequate spectral bandwidth and resolution as well as providing a reasonable approximate model of isotropic turbulence as a superposition of self-similar turbules of several different sizes. In particular, for simulation #1, the minimum appropriate time interval $T = 244$ s was used, corresponding to a tube length $L = wT = 756$ m centered around the origin. The number of time points in this simulation was chosen to be $N = 6048$, corresponding to a time interval $\Delta t = 0.0403$ s. For simulation #2, T was chosen to be 900 s, corresponding to the duration of the relevant experimental measurement, so $L = 2793$ m, and $N = 23\,334$, for the same Δt.

The ratio of cell size $d = n^{-1/3}$ to turbule size a was chosen as $d/a = 8$, yielding a packing fraction $\varphi = na^3 = 0.001\,95$. The total number of sizes used was $N_s = 6$, and the largest size used was $a_1 = 1.5$ m. The actual outer scale size in the relevant experiment must have been much greater than this, but eddies larger than a_1 contribute very little to the Born approximation scattering for the 0.68-m wavelength. Each successively smaller turbule size was chosen to be half the previous size, so $N_d = 2$, $\mu = \ln 2$ [see Eq. (39)]. A smaller value of N_d and a larger value of N_s would be more realistic, but much more demanding of computation time. These values yield the size of the smallest turbule used, $a_6 = 2^{-5} a_1 = 4.68$ cm. The actual smallest size should probably be of the order of millimeters, but this is computationally intractable on the PC platform used. Finally, utilizing the values given above for $(T_0, c, \phi, \mu, C_T^2, C_v^2)$, Eqs. (38) yield values $\delta T_1/T_0 = 0.013$ and $\Omega_1 a_1/c = 0.027$ for the temperature and velocity fluctuation amplitudes respectively, of the largest turbule used.

Using these values of the input parameters and the

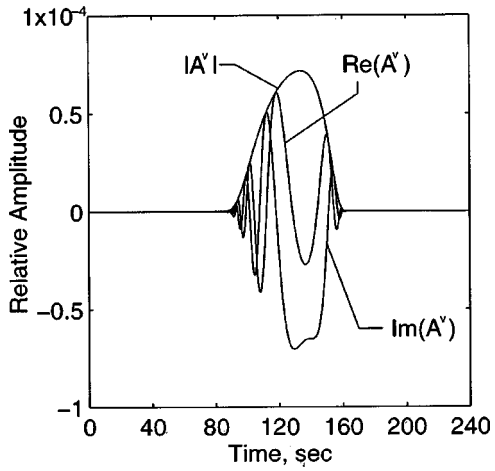


FIG. 3. Real and imaginary parts and magnitude of the complex amplitude $A^{v}(t)$ of the detector signal due to velocity scattering by a moving eddy of scale length 1.5 m.

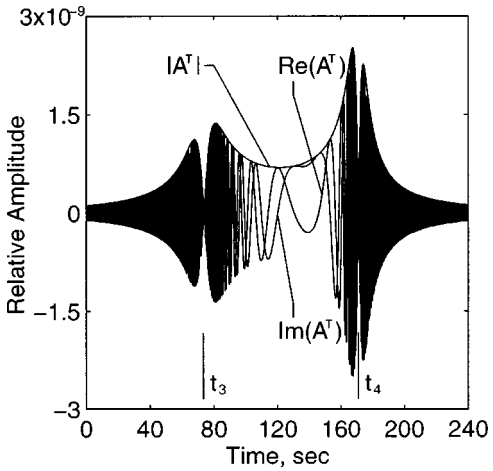


FIG. 5. Real and imaginary parts and magnitude of the complex amplitude $A^{T}(t)$ of the detector signal due to temperature scattering by a moving eddy of scale length 4.7 cm.

theory and algorithms described above and in Appendixes A and B, the total time-dependent complex acoustic pressure amplitudes $(A^{T}(t),A^{v}(t))$ due to individual turbules, tubes of turbules of one size, and all tubes of turbules of all sizes used were calculated numerically, starting from the exact equations (A17) and (A20) for one turbule. Then, they were Fourier transformed to get the temperature and velocity spectra $(\mathcal{A}^{T}(\omega),\mathcal{A}^{v}(\omega))$. For convenience, each (A^{T},A^{v}) was normalized by dividing by the factor $\mathcal{P}_e$, and setting $a'=1$ m. Thus, $\mathcal{P}_e$ is now the pressure amplitude of the source wave at a distance $a'=1$ m from the source in a still atmosphere. Equations (38) reveal that many of the results should not be sensitive to ϕ and μ, as long as μ is small enough to allow several large and small eddy sizes in the population, e.g., doubling ϕ/μ is equivalent to reducing δT_1 and Ω_1 by a factor of $\sqrt{2}$, or reducing the source amplitude $\mathcal{P}_e$ by $\sqrt{2}$, etc.

In Fig. 3, we plot the real and imaginary parts and the magnitude of $A^{v}(t)$ relative to $\mathcal{P}_e a'$ from Eq. (A20) for a single turbule with length scale $a=a_1=1.5$ m, with $\overline{b_x}=-L/2$, $\overline{b_y}=0$, $\overline{b_x}=20$ m. Here, $\overline{b_z}=20$ m corresponds to an above-ground base height $z_b=14$ m (see Fig. 2) plus $d_1/2=6$ m, half the side of a cubical cell having $d_1=8a_1$. In Fig. 4, we plot the normalized magnitude of the discrete Fourier transform $\mathcal{A}^{v}(\omega)$ of the $A^{v}(t)$ in Fig. 3. In this and all subsequent spectral plots, zero frequency corresponds to the source frequency. From Eqs. (21) and (37) it is easy to understand why the pulse has a fairly narrow width in time: the Gaussian envelope kills it rapidly as the scattering angle increases, since for Fig. 3, $k_s a_\alpha = k_s a_1 \approx 13.9$. The frequency width can be obtained easily from Eq. (31). That is, calculate or obtain from Fig. 3 the two times (t_1,t_2) at which the time-dependent magnitudes $|A(t)|$ are, say, 5% of their maximum values, and find $\hat{\mathbf{R}}^{DE}$ and $\hat{\mathbf{R}}^{ES}$ at these times. Then, verify from Fig. 3 that $|\omega(t_1)-\omega(t_2)|$ is a good estimate of the (angular) frequency width of the $\mathcal{A}^{v}(\omega)$. This works because the complete frequency range, $\nu_s(1-2\beta)$ to $\nu_s(1+2\beta)$, is *scanned* by any turbule, but only a small part of it is manifest in the scattered waves from large turbules.

In Figs. 5 and 6 we repeat the presentations of Figs. 3 and 4 but for $(A^{T}(t),|\mathcal{A}^{T}(\omega)|)$ starting from Eq. (A17) for a single small turbule of size $a_6=4.6875$ cm, with the same $(\overline{b_x},\overline{b_y},\overline{b_x})$. Here $(k_s a_6)^2=0.188$, so the Gaussian envelope of the $A^{T}(t)$ is almost unity for all scattering angles. Therefore, the $A(t)$ are nonzero over the total time duration, and

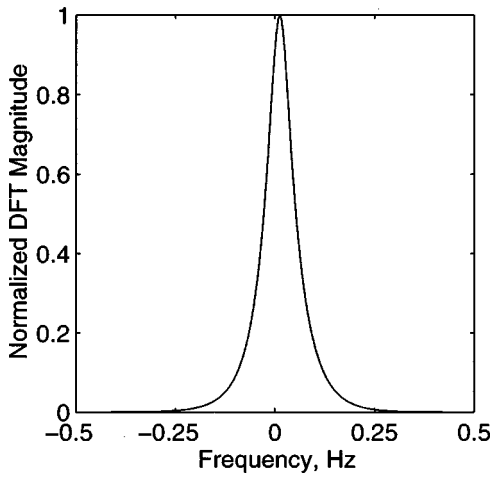


FIG. 4. Magnitude of the Fourier transform of the $A^{v}(t)$ in Fig. 3.

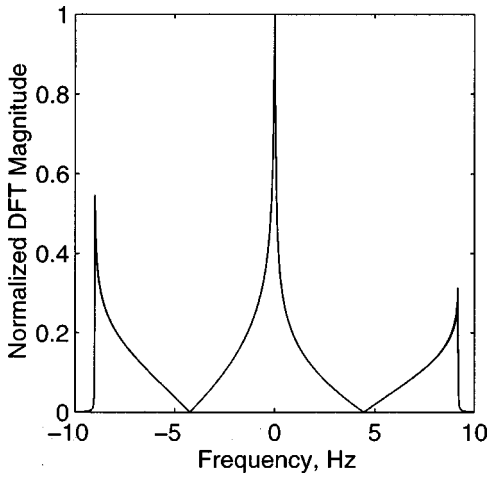


FIG. 6. Magnitude of the Fourier transform of the $A^{T}(t)$ in Fig. 5.

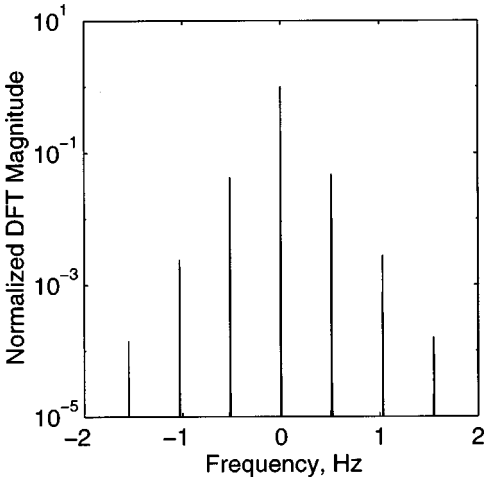


FIG. 7. Magnitude of the Fourier transform of the complex amplitude of the detector signal, due to velocity scattering by a tube of equally spaced moving eddies of scale length 0.75 m.

the frequency width extends to almost the maximum possible, $\pm 2\beta\nu_s$, obtainable due to backscattering from Eq. (31). Note also that $\mathcal{A}^T(\omega)$ goes to zero at frequencies $\nu_+ = 4.3\,\mathrm{Hz}$, $\nu_- = -4.2\,\mathrm{Hz}$. This is easily understandable from Eqs. (19), (24), and (31): The time-dependent amplitudes $\mathrm{A}^{\mathrm{T}}(t)$ are approximately zero when $\hat{\mathbf{R}}^{DE} \cdot \hat{\mathbf{R}}^{ES} = \cos\theta = 0$ (to lowest order in β). From Fig. 5, obtain the times t_3 and t_4 at which this is true, and then use Eq. (31) to get $\omega(t_3)$ and $\omega(t_4)$; these agree very well with the frequencies $\nu_\pm$. Also, note the asymmetry of the $\mathrm{A}(t)$ and the $\mathcal{A}(\omega)$ in Figs. 3–6; this is due to the source height being different from the detector height.

Figures 7 and 8 contrast the effects of randomness versus regularity in turbule center locations for many turbules. In Fig. 7, we present $|\mathcal{A}^v(\omega)|$ for a tube of turbules with $a = 0.75\,\mathrm{m}$, with all angular velocities in the y direction, and each turbule center at the center of its comoving cell of side length $d_a = 8a = 6\,\mathrm{m}$ (see Fig. 2). This tube forms a moving diffraction grating. The tube centerline is $(\overline{b_y} = 0,\ \overline{b_z} = 20\,\mathrm{m})$, and for simulation #1, with $T = 244\,\mathrm{s}$, there are

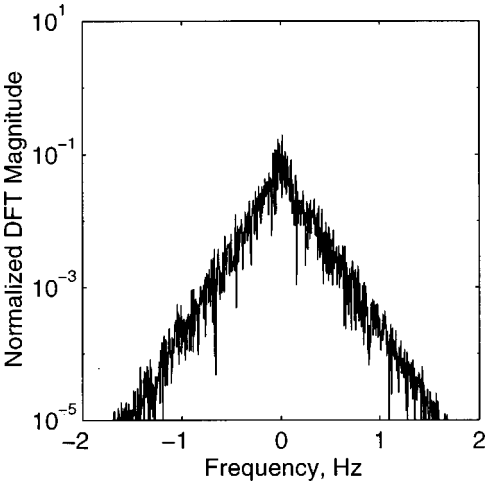


FIG. 8. Magnitude of the Fourier transform of the complex amplitude of the detector signal, due to velocity scattering by a tube of randomly placed moving eddies of scale length 0.75 m.

$M_a = L/d_a = 126$ turbules contributing at any time. The principal maxima in Fig. 7 are spaced just as they should be, i.e., at frequencies $\nu_q = qM_a/T = qw/d_a \approx (q)(0.516)\,\mathrm{Hz}$, where q is an integer. For a discrete Fourier transform (DFT), there should be exactly zero contribution for other frequencies, i.e., no subsidiary maxima, because the zeros of the interference pattern occur at just the discrete transform frequencies $\nu_l = l/T$, $l = \text{integer} \neq 0,\ \pm M_a,\ \pm 2M_a$, etc. The magnitudes of the principal maxima diminish rapidly as $|\nu|$ increases, because of the Gaussian envelope of a single turbule. (The situation is analogous to a multiple-slit interference pattern as modified by the single-slit diffraction pattern in optics.)

Figure 8 is for the same conditions as Fig. 7, except that each turbule center is randomly placed inside its comoving cell, with no correlations among placements of different turbules, as in Fig. 2. The randomness consists of uniform uncorrelated probabilities of a turbule center's (x, y, z) locations being between $-d/4$ and $d/4$ from its comoving cell center. This randomness in turbule placement yields the jagged spectrum of Fig. 8. Both Figs. 7 and 8 display a remarkable roof-like envelope, which corresponds on these semilog plots to an exponential spectral envelope, $\exp(-|\nu|\tau_a)$, where τ_a is a constant with the dimension of time that very likely depends on the scale length a. We have not yet been able to predict this behavior theoretically.

Figure 9 is the sum over contributions of many turbules of all sizes used, i.e., $a_\alpha = (1.5, 0.75, 0.375, 0.1875, 0.093\,75, 0.046\,875)\,\mathrm{m}$ for $\alpha = (1,\ldots,N_s = 6)$, respectively. This is for simulation #1, with $T = 244\,\mathrm{s}$. The amplitudes $(\Omega_\alpha, |\delta T_\alpha|)$ were scaled as discussed earlier, so these calculations are an attempt to represent realistic homogeneous fully developed turbulence numerically by a turbule model, following the formulas in Goedecke and Auvermann.[13] Figure 9 plots $|\mathcal{A}^T(\omega) + \mathcal{A}^v(\omega)|$ for four big tubes with centerlines $b_y = \pm 6\,\mathrm{m}$, $b_z = (20, 32)\,\mathrm{m}$. The semilog plots clearly show the contributions of the small turbules all the way out to $\pm 2\beta\nu_s \approx \pm 9.2\,\mathrm{Hz}$. In Figs. 9–11, the spectra are normalized such that the largest spike has unit value.

Figures 10 and 11 contrast experimental results with our model results. The source–detector geometry and wind velocity of the experimental scenario are similar to what we modeled. The experiment has previously been described;[11,20] a summary is provided here.

The experimental data were obtained on 29 October 1996 at an airstrip near Ottawa. The mean wind speed was 3.1 m/s directed essentially from detector to source, so the detector was in the source shadow zone and the data represent scattered sound. The sky was clear and the mean temperature was about 10 °C.

One microphone in a linear array was used as the receiver. The array was deployed along the centerline of the airport runway and the microphones, with wind screens, were positioned about 5 cm from the surface. The microphone of interest was about 300 m from the source, which was a circular exponential horn 61 cm in diameter, mounted 10 m above the surface of the runway and about 10 m off the runway centerline. The source was directed away from the receiver and at an angle of approximately 15° downward from horizontal. Measurements outdoors and in an anechoic

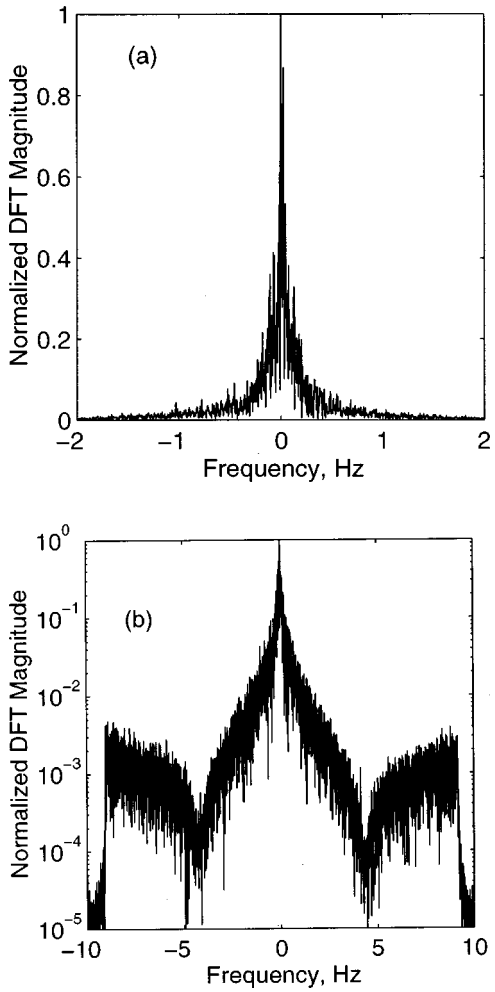


FIG. 9. Magnitude of the Fourier transform of the complex amplitude of the detector signal due to both velocity and temperature scattering by the collection of randomly moving eddies of different scale lengths chosen to represent isotropic homogeneous turbulence. (a) Linear scale; (b) log scale.

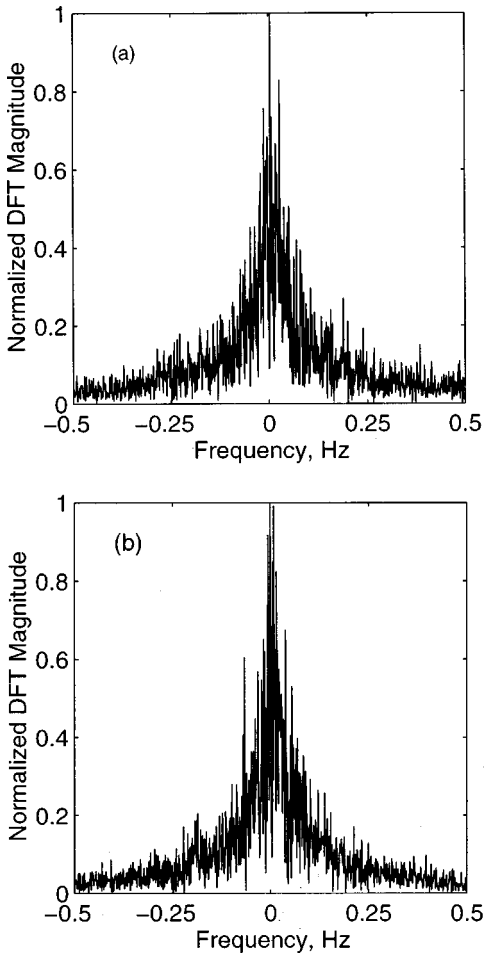


FIG. 10. Normalized Fourier transform magnitudes of a detector signal of duration 900 s. (a) Experiment; (b) model.

chamber indicate that the SPL of the sound field generated by the horn in the rearward direction is about 12 dB less than on the forward axis.

Six tones (380, 500, 600, 700, 820, and 940 Hz) were prerecorded on digital audiotape and reproduced by the source. The signal at the receiver was digitized at 8000 samples per second and subsequently processed to extract only the 500-Hz tone. This was done by applying a bandpass filter, bandshifting so that the 500-Hz tone is in the center of the passband, and then decimating 20:1. Details of the digital signal processing used can be found in an article by Havelock *et al.*[21]

Figure 10(a) is the normalized magnitude of the DFT of the experimental data for the full 900 s, using $N=23\,334$ points plotted versus frequency between -0.5 and $+0.5$ Hz. We arbitrarily limited the plots to this range because the signal-to-noise ratio of the experimental data seemed to be unity or less for frequency shifts much beyond $\pm.5$ Hz. Figure 10(b) is the normalized $|\mathcal{A}^T(\omega)+\mathcal{A}^v(\omega)|$ for our model simulation #2 for $T=900$ s, using the parameters, sizes, scalings, and the four large tubes mentioned above. The two figures clearly have the same general shape. In particular, they exhibit the same peak width and jaggedness.

Figure 11(a) was obtained by subdividing the experimental data into three disjoint segments, each 244 s long, taking the DFT using $N=6048$ points for each segment, and then averaging the DFT magnitudes and normalizing as above. This averages out variations having time scales between about 244 and 900 s. Figure 11(b) is for simulation #1, for $T=244$ s, normalized as above; it is just a stretch of Fig. 9(a) along the abscissa, to match the -0.5 to $+0.5$-Hz width of Fig. 11(a). Again, the experimental and simulation graphs have the same general shape, including the same peak width and jaggedness. We feel that this agreement between Figs. 10(a) and (b) and between Figs. 11(a) and (b) is important evidence for the efficacy of the turbule model.

Figures 10 and 11 also reveal that the rapidity of the jagged fluctuations in the frequency spectra changes with T. To investigate this, we expanded the horizontal scale of these figures and counted the number of local maxima. In all cases we found that the mean frequency spacing between adjacent maxima is essentially $3/T$. We have done a theoretical derivation and further computer simulation which show that a mean spacing of $3/T$ is an exact result for frequency spectra of a great many stochastic systems. We hope to publish this work later.

Finally, Fig. 12 compares smoothed versions of the experimental and simulation spectra from Figs. 10(a) and (b), respectively. The smoothing was done using a 23-point run-

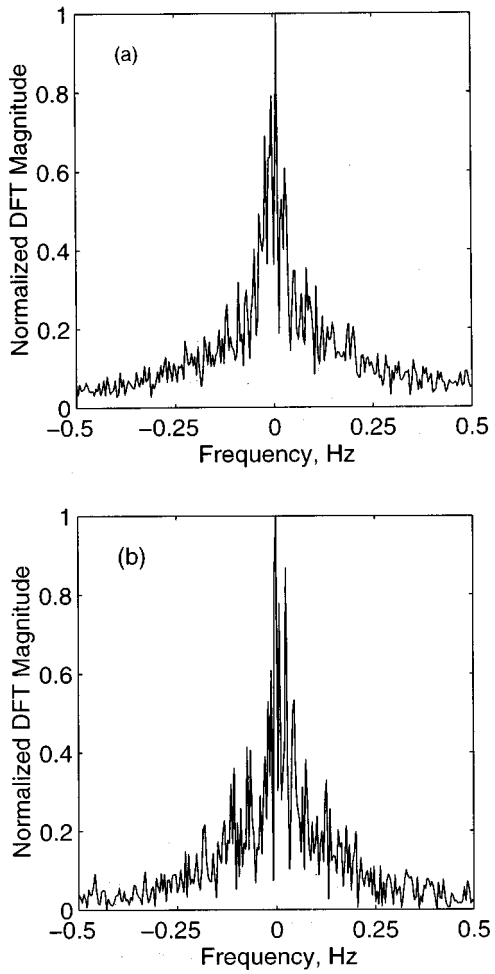


FIG. 11. Normalized Fourier transform magnitudes of a detector signal of duration 244 s. (a) Experiment; (b) model.

ning average for both these curves. The two smoothed spectra are quite similar indeed.

IV. SUMMARY AND DISCUSSION

For wide-angle acoustic sources and detectors located close to the scattering volume, the theory developed in this work allows understanding of frequency spectra observed in acoustic waves scattered from advecting turbulence on the basis of three simple general physical principles. The first principle is that, even if the turbulence is time independent (frozen) in a frame comoving with the wind, substantial spectral broadening occurs because the scattering angles and the distances from source to turbulent eddies and from the eddies to the detector are time dependent. The second principle is that large objects, in comparison to the wavelength, scatter strongly but mainly in near-forward directions, while small objects scatter weakly, but nearly isotropically. Thus, large eddies produce only a small spectral width $\Delta\nu_w$ due to flow with the wind, while small eddies produce essentially the maximum possible, $\Delta\nu_w \approx 4\beta\nu_s$ (β is the Mach number of the wind, ν_s is the source frequency), but lesser amplitude. This predicts the general shape of the frequency spectrum, i.e., a sharp peak around the source frequency due to the large eddies, plus long tails due to the small ones. The third principle is that the observed jaggedness of the frequency spectra is due to randomness in the locations of eddies.

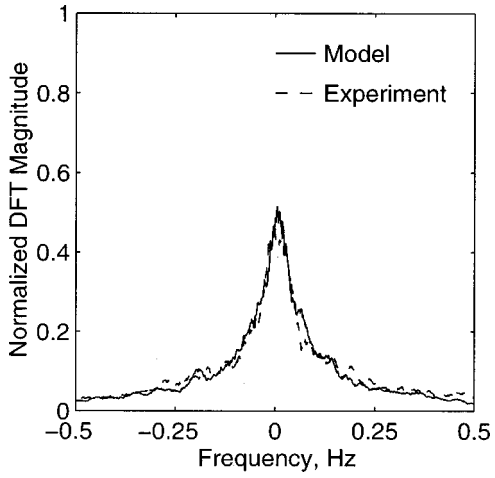


FIG. 12. Running averages of Figs. 10(a) and (b) over 23 points.

It seems worth emphasizing that in this work we have not added any parameters to the turbule model of isotropic homogeneous turbulence developed by Goedecke and Auvermann.[13] The model in the frame moving with the wind is the same as it was in the ground frame with no wind. The model does not require many parameters to determine the microstructure of the turbulence, and some of these are irrelevant to the normalized spectra since they are overall amplitude factors. We made no effort to fine-tune the microstructure parameters to achieve a best fit. Instead, as discussed in the previous section, we chose several such parameters primarily on the basis of computational tractability. Still, we got very good agreement with the experimental frequency spectrum. We believe that the crucial factors that produce this agreement are the scaling relations in Eqs. (35), (39)–(41).

Of course, the turbulent flow in a frame comoving with the mean wind is also intrinsically time dependent. The usual estimates of spectral broadening attribute all of it to this time dependence. For example, Galindo and Havelock[11] analyzed their results on the basis of the fluctuating vertical wind components, using the conventional treatment, in which the source and detector are assumed to be far from the whole turbulent scattering volume, i.e., $h_S, h_D \gg L$ (see Fig. 1). However, as discussed by Tatarskii,[8] this intrinsic time dependence generally produces a spectral broadening $\Delta\nu_1 \ll \beta\nu_s$ for typical atmospheric conditions.

One of the approximations made in the model developed was the assumption of a uniform mean wind and a uniform ambient temperature. In actuality, these ambient fields are vertically stratified. We expect that this would not change the general features of our results appreciably, at least for weak winds.

Another approximation was that the source and detector are in the far-field region of any one turbule at all times. This is not true at all times for all turbules in realistic scenarios, in which a few of the turbules may actually intercept the source and/or detector, and in which a substantial fraction of the large turbules spend some time close to the source and detector. The far-field expressions were used as a compromise, partly because they yield the same results as the conventional

stationary stochastic theory in the limit of no wind. We expect that corrections to this approximation would not alter the general character of the spectra.

Other facets of an experimental scenario are (i) source waves reflected from the ground will be incident on the turbulence; (ii) the source and/or the detector may be strongly anisotropic; and (iii) the turbulence may be partially anisotropic and /or inhomogeneous. The general character of scattered wave spectra may be altered appreciably by all these effects.

Overall, the frequency spectra predicted by this model of frozen turbulence flowing with a horizontal wind seem to have the same general characteristics as some observed spectra. Whether this model applies well to many different sets of experimental results needs to be tested. In particular, experiments using sodars with narrow directivity patterns need to be done, and their results compared with corresponding predictions of this model. The spectral width of the scattered wave due to frozen turbulence should be much less than $4\beta\nu_s$ in these cases. Also, a monostatic sodar should receive less than a bistatic one, because the former should record very little scattering from the larger turbules. A monostatic but wide-angle source–detector should record appreciable broadening due to the scattering from the smaller turbules, but no strong central peak from the larger turbules. If the flowing frozen turbulence model agrees with these experiments fairly well, then quite likely it could be used to analyze acoustical remote-sensing data with a view toward identification of sources, or inferring detailed information about atmospheric turbulence and/or wind speeds.

Finally, we remark that the use of localized turbulent eddies (''quasiwavelets''or turbules) has allowed us to develop a relatively simple theory and numerical model to describe frequency broadening of a monochromatic wave due to advecting turbulence. This approach looks ''natural'' for the considered problem and probably has some advantages over other possible approaches such as the statistical one in Tatarskii.[8] In particular, anisotropy could be modeled via nonspherical turbules and/or turbule rotation axes with anisotropic orientations, and inhomogeneity and intermittency could be modeled by comoving regions having different packing fractions and/or different turbule fluctuation amplitudes. We plan to investigate these models in the future.

ACKNOWLEDGMENTS

This paper is partly based upon work that was supported by the U.S. Army Research Office under Contract Numbers DAAG5-98-1-0463 and DAAG55-97-1-0178, by the U.S. Army Research Laboratory, and by an NRC-ETL Research Associateship.

APPENDIX A: SCATTERING BY A MOVING TURBULENT EDDY

In this Appendix, we outline the derivation of the expressions for sound waves emitted from an isotropic point source and scattered into a detector by a turbule moving with a uniform constant wind **w**.

After some algebra, we get the causal (retarded) Green function solution of Eqs. (14) and (15) as

$$\bar{\Phi}(\bar{\mathbf{r}},t)=-\frac{A'}{4\pi\rho_0 c^2}\frac{e^{-i\omega_s(t-\bar{R}_0/c)}}{\bar{D}_s}+\text{c.c.}, \tag{A1}$$

where

$$\bar{\mathbf{R}}_0\equiv\bar{\mathbf{R}}_s-\boldsymbol{\beta}\bar{R}_0, \tag{A2}$$

$$\bar{R}_0=\{-\boldsymbol{\beta}\cdot\bar{\mathbf{R}}_s+\bar{D}_s\}/(1-\beta^2), \tag{A3}$$

$$\bar{D}_s\equiv\sqrt{(1-\beta^2)(\bar{R}_s)^2+(\boldsymbol{\beta}\cdot\bar{\mathbf{R}}_s)^2}, \tag{A4}$$

$$\bar{\mathbf{R}}_s\equiv\bar{\mathbf{r}}-\mathbf{s}+\mathbf{w}t. \tag{A5}$$

Here, $\bar{\mathbf{R}}_s$ is the instantaneous vector from the source to any point $\bar{\mathbf{r}}$ fixed in the barred (comoving) frame, and $\boldsymbol{\beta}\equiv\mathbf{w}/c$.

From Eqs. (13), (A1)–(A5), after more algebra, we get the source wave

$$\begin{aligned}&\bar{p}_e(\bar{\mathbf{r}},t)=(\mathcal{P}_e)(a'\bar{R}_0/\bar{D}_s^2)e^{-i\omega_s(t-\bar{R}_0/c)}+\text{c.c.},\\&\bar{\mathbf{v}}_e(\bar{\mathbf{r}},t)=\bar{\mathbf{R}}_0\bar{p}_e(\bar{\mathbf{r}},t)/\rho_0 c\bar{R}_0.\end{aligned} \tag{A6}$$

Here, we have put $A'=i4\pi c\mathcal{P}_e a'/k_s$ for convenience; $\mathcal{P}_e$ may be chosen real with no loss of generality. Then, it is clear from Eqs. (A4)–(A6) that $\mathcal{P}_e$ is half the source wave pressure amplitude at distance a' from the source in a still atmosphere, i.e., if $\beta=0$, for which $\bar{R}_s=\bar{D}_s=|\bar{\mathbf{r}}-\mathbf{s}|=|\mathbf{r}-\mathbf{s}|$. In obtaining Eqs. (A6), we assumed that the point $\bar{\mathbf{r}}$ is always in the far field of the source, so $k_s\bar{R}_0\gg1$, where $k_s=\omega_s/c$. So, we neglected all terms like $\partial\bar{D}_s^{-1}/\partial s_i$ in comparison to terms like $ik_s\bar{D}_s^{-1}\partial\bar{R}_0/\partial s_i$.

The causal Green function solutions of Eqs. (16) and (17) are

$$\begin{aligned}\bar{p}_s^T(\bar{\mathbf{r}},t)=&\frac{\partial}{\partial\bar{r}_i}\left(\frac{1}{4\pi T_0}\right)\\&\times\int d^3\bar{r}'\frac{\bar{T}_F(\bar{\mathbf{r}}')}{|\bar{\mathbf{r}}-\bar{\mathbf{r}}'|}\left[\frac{\partial\bar{p}_e(\bar{\mathbf{r}},t')}{\partial\bar{\mathbf{r}}_i'}\right]_{t'=t-|\bar{\mathbf{r}}-\bar{\mathbf{r}}'|/c}\end{aligned} \tag{A7}$$

$$\bar{p}_s^v(\bar{\mathbf{r}},t)=\frac{\partial^2}{\partial\bar{r}_i\partial\bar{r}_j}\frac{2\rho_0}{4\pi}\int d^3\bar{r}'\frac{\bar{v}_{Fj}(\bar{r}')}{|\bar{\mathbf{r}}-\bar{\mathbf{r}}'|}\bar{v}_{ei}(\bar{\mathbf{r}}',t-|\bar{\mathbf{r}}-\bar{\mathbf{r}}'|/c), \tag{A8}$$

where $\bar{\mathbf{r}}$ is now the observation point. After doing the indicated derivatives, we will set $\bar{\mathbf{r}}=\mathbf{d}-\mathbf{w}t$, the detector location in the barred frame.

From Eq. (A6), we get

$$\begin{aligned}&\left[\frac{\partial\bar{p}_e(\bar{\mathbf{r}}',t')}{\partial\bar{\mathbf{r}}_i'}\right]_{t'=t-|\bar{\mathbf{r}}-\bar{\mathbf{r}}'|/c}\\&\quad=(ik_s\mathcal{P}_e)(a_0'\bar{R}\bar{R}_{0i}/\bar{D}_s^3)e^{-i\omega_s[t-(\bar{R}_0+\bar{R}_d)/c]}+\text{c.c.},\end{aligned} \tag{A9}$$

where $\bar{R}_{0i}$, $\bar{R}_0$, and $\bar{D}_s$ are still given by Eqs. (A2)–(A4), and

$$\bar{\mathbf{R}}_d=\bar{\mathbf{r}}-\bar{\mathbf{r}}', \tag{A10}$$

but now

$$\overline{R}_{si}=\overline{r}_i'-s_i+w_it-\beta_i\overline{R}_d. \tag{A11}$$

In obtaining Eq. (A9), we neglect terms like $\partial\overline{R}_d^{-1}/\partial\overline{r}_i$ compared to $k_s\overline{R}_d^{-1}\partial(\overline{R}_0+\overline{R}_d)/\partial\overline{r}_i$, since we assume the far-field limit $k_s\overline{R}_d\gg 1$.

Using Eqs. (A7), (A9), and (A10), after some algebra we get

$$\overline{p}_s^T(\overline{\mathbf{r}},t)=\frac{-k_s^2\mathcal{P}_e a'}{4\pi T_0}\int d^3\overline{r}'\frac{\overline{R}_0\overline{R}_{0i}}{(\overline{D}_s)^3}\frac{\overline{R}_{di}}{\overline{R}_d^2}$$
$$\times\overline{T}_F(\overline{\mathbf{r}}')e^{ik_s(\overline{R}_0+\overline{R}_d)}e^{-i\omega_s t}+\text{c.c.} \tag{A12}$$

The turbule is localized around $\overline{\mathbf{r}}'=\overline{\mathbf{b}}$, as in Eq. (2)

$$\overline{T}_F(\overline{\mathbf{r}}')=\overline{T}(\overline{\mathbf{r}}'-\overline{\mathbf{b}})=\overline{T}(\boldsymbol{\xi}),\quad \boldsymbol{\xi}\equiv\overline{\mathbf{r}}'-\overline{\mathbf{b}}, \tag{A13}$$

where $\overline{T}$ is some function of $\overline{\mathbf{r}}'-\overline{\mathbf{b}}$. We change integration variables from $\overline{\mathbf{r}}'$ to $\boldsymbol{\xi}$ in the integrand of Eq. (A12) and discard terms in $\boldsymbol{\xi}$ in the $\overline{R}_0$, $\overline{R}_{0i}$, etc., in the numerator and denominator, because $\overline{R}_d\gg\xi$, $\overline{R}_0\gg\xi$. But, we must keep terms of order ξ in the exponent. That is, we write the exponent as a power series in $\boldsymbol{\xi}$, keeping only the first two terms. We also put the observation point $\overline{\mathbf{r}}$ at the detector, $\overline{\mathbf{r}}=\mathbf{d}-\mathbf{w}t$. After some algebra, we get

$$e^{ik_s(\overline{R}_0+\overline{R}_d)}\rightarrow e^{ik_s(\overline{R}_0+\overline{R}^{DE})}e^{-i\mathbf{K}\cdot\boldsymbol{\xi}}, \tag{A14}$$

where

$$K_i=k_s\{\hat{R}_i^{DE}-(\delta_{ij}+\hat{R}_i^{DE}\beta_j)(\overline{R}_{0j}/\overline{D}_s)\}, \tag{A15}$$

with $\overline{\mathbf{R}}_0$, $\overline{R}_0$, and $\overline{D}_s$ still given by Eqs. (A2)–(A4), but now with

$$\overline{\mathbf{R}}_s=\mathbf{R}^{ES}-\boldsymbol{\beta}R^{DE}, \tag{A16}$$

where $(\mathbf{R}^{ES},\mathbf{R}^{DE})$ are defined by Eqs. (6) and (7).

Then, from Eq. (A12) we get $p_s^T(t)$ in the form of Eq. (18) of the text, with

$$\mathrm{A}^{\mathrm{T}}(t)=-\frac{k_s^2\mathcal{P}_e a'}{4\pi T_0}\frac{\overline{R}_0\overline{\mathbf{R}}_0\cdot\hat{\mathbf{R}}^{DE}}{(\overline{D}_s)^3R^{DE}}e^{ik_s(\overline{R}_0+R^{DE})}\overline{\mathcal{T}}(\mathbf{K}), \tag{A17}$$

where $\overline{\mathcal{T}}(\mathbf{K})$ is defined by Eq. (20). This expression is accurate if the source and detector are in the far-field region of the turbule. It is also valid to all orders of $\boldsymbol{\beta}$. We used it in our calculations. To lowest order in β, we have $\overline{R}_s=R^{ES}=\overline{R}_0=\overline{D}_s$, so $\mathbf{K}=k_s(\hat{\mathbf{R}}^{DE}-\hat{\mathbf{R}}^{ES})$ as in Eq. (21) of the text, and Eq. (A17) reduces to Eq. (19).

In Eq. (A8), we need the second derivative of the factor $\exp[ik_s(\overline{R}_0+\overline{R}_d)]$. Using the far-field limit, we get

$$\frac{\partial^2}{\partial\overline{r}_i\partial\overline{r}_j}e^{ik_s(\overline{R}_0+\overline{R}_d)}\approx -k_s^2\frac{\overline{R}_{di}\overline{R}_{dj}}{\overline{R}_d^2}\frac{\overline{R}_0^2}{(\overline{D}_s)^2}e^{ik_s(\overline{R}_0+\overline{R}_d)}. \tag{A18}$$

We also write

$$\overline{v}_{Fj}(\overline{\mathbf{r}}')\equiv\overline{v}_j(\overline{\mathbf{r}}'-\overline{\mathbf{b}})\equiv\overline{v}_j(\boldsymbol{\xi}), \tag{A19}$$

where $\overline{v}_j$ are localized functions.

Therefore, for $\overline{\mathbf{r}}=\overline{\mathbf{d}}-\mathbf{w}t$, using the same method as for p_s^T, we get the form of Eq. (23) for $p_s^v(t)$, in which $\mathrm{A}^{\mathrm{v}}(t)$ is given by

$$\mathrm{A}^{\mathrm{v}}(t)=\left\{-\frac{k_s^2\mathcal{P}_e a'}{2\pi c}\frac{(\overline{R}_0)^2}{(\overline{D}_s)^4}\frac{(\overline{\mathbf{R}}_0\cdot\hat{\mathbf{R}}^{DE})(\hat{\mathbf{R}}^{DE}\cdot\overline{\boldsymbol{\mathcal{V}}}(\mathbf{K}))}{R^{DE}}\right\}$$
$$\times e^{ik_s(\overline{R}_0+\overline{R}^{DE})}, \tag{A20}$$

when $\overline{\boldsymbol{\mathcal{V}}}(\mathbf{K})$ is defined by Eq. (25). To lowest order in β, Eq. (A20) reduces to Eq. (24) of the text.

APPENDIX B: TIME-SHIFT ALGORITHM

Imagine two identical turbules, #1 and #2, more or less in line along the wind direction, chosen to be the x direction for convenience in this discussion. That is, let the centers of their two cells have coordinates $(\overline{b}_{cx}^{(1)},\overline{b}_{cy}^{(1)},\overline{b}_{cz}^{(1)})$ and $(\overline{b}_{cx}^{(2)},\overline{b}_{cy}^{(1)},\overline{b}_{cz}^{(1)})$, i.e., their cells are in line with the wind (see Fig. 2 for illustration).

Now, consider Eq. (A17) for $\mathrm{A}_{(1)}^{\mathrm{T}}(t)$, from turbule #1, and assume that the source and detector are always in the far-field region of a comoving cell of side length $d=a/\varphi^{1/3}$ in which the turbule is located. This implies $R_{(1)}^{DE}(t)$ and $R_{(1)}^{ES}(t)\gg(\lambda,d^2/\lambda)$, whichever is greater, at all (or most) times. For such cases, vectors such as $(\mathbf{R}^{ES(1)},\ldots)$ can be replaced by $(\mathbf{R}_c^{ES(1)},\ldots)$, the vectors to the center of the cell, except in the factor $\exp[ik_s(\overline{R}_0^{(1)}+R^{DE(1)})]$. In this factor, we may write

$$e^{ik_s(\overline{R}_0^{(1)}+R^{DE(1)})}\approx e^{ik_s(\overline{R}_{0c}^{(1)}+R_c^{DE(1)})}e^{-i\mathbf{K}_c^{(1)}(t)\cdot\boldsymbol{\eta}_1}, \tag{B1}$$

where $\boldsymbol{\eta}_1$ is the displacement of the center of turbule #1 from its cell center (see Fig. 2), and $\mathbf{K}_c^{(1)}$ is given by Eq. (A15) but with vectors to the cell center. So, we get

$$\mathrm{A}_{(1)}^{\mathrm{T}}(t)=\mathrm{A}_{(1)c}^{T}(t)e^{-i\mathbf{K}_c^{(1)}(t)\cdot\boldsymbol{\eta}_1}, \tag{B2}$$

that is, $\mathrm{A}_{(1)c}^{\mathrm{T}}(t)$ is the old $\mathrm{A}_{(1)}^{\mathrm{T}}(t)$, but calculated with vectors to the cell center.

Now, for turbule #2, it is clear that $(\mathrm{A}_{(2)c}^{\mathrm{T}}(t),\mathbf{K}_c^{(2)}(t))$ are simply time shifted from $(\mathrm{A}_{(1)c}^{\mathrm{T}}(t),\mathbf{K}_c^{(1)}(t))$. That is, we have

$$\mathrm{A}_{(2)c}^{\mathrm{T}}(t)=\mathrm{A}_{(1)c}^{\mathrm{T}}(t+\tau_2),\quad \mathbf{K}_c^{(2)}(t)=\mathbf{K}_c^{(1)}(t+\tau_2), \tag{B3}$$

where $\tau_2\equiv w^{-1}(\overline{b}_{cx}^{(2)}-\overline{b}_{cx}^{(1)})$. For example, suppose #2 is one cell ahead of #1, so that $b_{cx}^{(2)}-b_{cx}^{(1)}=d$. Then, $\tau_2=d/w$.

Thus, the time-shift algorithm for a whole tube of M turbules of a given size, with cell centers equally spaced by d along the wind direction, is

$$\mathrm{A}_{\mathrm{tube}}^{\mathrm{T}}(t)=\sum_{l=1}^{M}\mathrm{A}_{(1)c}^{\mathrm{T}}(t+\tau_l)e^{-i\mathbf{K}_c^{(1)}(t+\tau_l)\cdot\boldsymbol{\eta}_l}, \tag{B4}$$

where

$$\tau_l=w^{-1}(\overline{b}_{cx}^{(l)}-\overline{b}_{cx}^{(1)})=(l-1)(d/w), \tag{B5}$$

and the intracell displacement $\boldsymbol{\eta}_l$ is chosen ''randomly'' for each l, say, $-d/4\leqslant(\eta_{lx},\eta_{ly},\eta_{lz})\leqslant d/4$ with uniform independent probabilities (again, refer to Fig. 2). So, for a given

tube, we need to compute the functions $(A^{T}_{(1)c}(t), \mathbf{K}^{(1)}_{c}(t))$ only once.

By comparison with direct computation of the $A^{T}_{tube}(t)$ for the same tube, using Eq. (A17) for each turbule to *its* center, with the same set of random locations as above, we found that, for source frequency $\nu_s = 500\,\mathrm{Hz}$, the time-shift algorithm is a good approximation for $d \lesssim 4\,\mathrm{m}$, but not as good for $d \geqslant 8\,\mathrm{m}$, for the case of interest, with R_c^{DE} and $R_c^{ES} \gtrsim 10\,\mathrm{m}$ minimum but actually $\gtrsim 100\,\mathrm{m}$ during the times of largest signal production. Since a direct computation involves ~ 100 operations per turbule, and a time shift only ~ 10 operations, the computing time advantage of the time-shift algorithm is substantial.

The same considerations as above apply for the $A^{v}(t)$ of Eq. (A20), except that we need to specify $\hat{\mathbf{\Omega}}$ for each turbule. For any turbule, we may write in general

$$A^{v}(t) = \hat{\mathbf{\Omega}} \cdot \mathbf{B}^{v}(t), \tag{B6}$$

where, from Eqs. (34) and (A20), we have

$$\mathbf{B}^{v}(t) = -\left(\frac{i\pi^{1/2} a^5 \Omega \mathcal{P}_e a'}{4c}\right) \frac{(\overline{R}_0)^2 (\overline{\mathbf{R}}_0 \cdot \hat{\mathbf{R}}^{DE})}{(\overline{D}_s)^4 R^{DE}} \times e^{-K^2 a^2/4} e^{ik_s(\overline{R}_0 + R^{DE})} (\hat{\mathbf{R}}^{DE} \times \mathbf{K}). \tag{B7}$$

The time-shift algorithm for a tube of M turbules along the wind is now

$$A^{v}_{tube}(t) = \sum_{l=1}^{M} \hat{\mathbf{\Omega}}_l \cdot \mathbf{B}^{v(1)}_{(c)}(t+\tau_l) e^{-i\mathbf{K}^{(1)}_c (t+\tau_l) \cdot \boldsymbol{\eta}_l}, \tag{B8}$$

where $\mathbf{B}^{v(1)}_{(c)}$ is calculated using vectors to the center of the cell containing and comoving with turbule #1. We calculate in the same way as for $A^{T}_{tube}(t)$, except that we must choose an $\hat{\mathbf{\Omega}}_l$ as well as an $\boldsymbol{\eta}_l$ for each turbule.

[1] M. J. Lighthill, ''On the energy scattered from the interaction of turbulence with sound or shock waves,'' Proc. Cambridge Philos. Soc. **49**, 531–551 (1953).

[2] G. W. Ford and W. C. Meecham, ''Scattering of sound by isotropic turbulence of large Reynolds numbers,'' J. Acoust. Soc. Am. **32**, 1668–1672 (1960).

[3] E. H. Brown, ''Acoustic Doppler radar scattering equation and general solution,'' J. Acoust. Soc. Am. Suppl. 1 **52**, 1391–1396 (1972).

[4] E. H. Brown, ''Turbulent spectral broadening of backscattered acoustic pulses,'' J. Acoust. Soc. Am. **56**, 1398–1406 (1974).

[5] A. Spizzichino, ''Discussion of the operating conditions of a Doppler sodar,'' J. Geophys. Res. **79**, 5585–5591 (1974).

[6] A. Spizzichino, ''Spectral broadening of acoustic and radio waves scattered by atmospheric turbulence in the case of radar and sodar experiments,'' Ann. Geophys. (C.N.R.S.) **31**, 433–445 (1975).

[7] L. M. B. C. Campos, ''The spectral broadening of sound by turbulent shear layers. I. The transmission of sound through turbulent shear layers,'' J. Fluid Mech. **89**, 723–749 (1978); ''II. The spectral broadening of sound and aircraft noise,'' **89**, 751–783 (1978).

[8] V. I. Tatarskii, *The Effects of the Turbulent Atmosphere on Wave Propagation* (Keter, Jerusalem, 1971), Chap. 2.

[9] P. Blanc-Benon and D. Juvé, ''Elargissement spectral d'un faiseau ultrasonore par traversée d'un champ turbulent,'' C. R. Seances Acad. Sci., Ser. 2 **292**, 551–554 (1981); in French.

[10] G. Y. Patrushev and A. P. Rostov, ''Anisotropy of the temporal amplitude fluctuations of a sound wave propagating over the ground surface in a turbulent atmosphere,'' Acoust. Phys. **42**, 76–78 (1996).

[11] M. Galindo and D. I. Havelock, ''Temporal coherence of a sound field in the turbulent atmosphere near the ground,'' Proceedings of the 7th International Symposium on Long-range Sound Propagation, Ecully, France, July 1996, pp. 149–159.

[12] E. H. Brown and F. F. Hall, Jr., ''Advances in atmospheric acoustics,'' Rev. Geophys. Space Phys. **16**, 47–109 (1978).

[13] G. H. Goedecke and H. J. Auvermann, ''Acoustic scattering by atmospheric turbules,'' J. Acoust. Soc. Am. **102**, 759–771 (1997).

[14] Reference 8, Sec. 34.

[15] A. S. Monin, ''Characteristics of the scattering of sound in a turbulent atmosphere,'' Sov. Phys. Acoust. **7**, 130 (1962).

[16] V. E. Ostashev, *Acoustics in Moving Inhomogeneous Media* (E & FN SPON, London, 1997), Chap. 6.

[17] D. Astruc, L. Plantié, R. Murenzi, Y. Lebret, and D. Vandrome, ''On the use of 3D wavelet transform for the analysis of computational fluid dynamics results,'' in *Progress in Wavelet Analysis and Applications*, edited by Y. Meyer and S. Rogues (Editions Frontières, France, 1993), pp. 463–470.

[18] Reference 8, Chap. 1, Sec. 12.

[19] E. H. Brown and S. F. Clifford, ''On the attenuation of sound by turbulence,'' J. Acoust. Soc. Am. **60**, 788–793 (1976).

[20] H. J. Auvermann and G. H. Goedecke, ''Influence of the Doppler effect on the bandwidth of acoustic signals scattered from atmospheric turbulence,'' Proceedings of the 8th International Symposium on Long-range Sound Propagation, University Park, PA, 9–11 Sept. 1998, pp. 375–388.

[21] D. I. Havelock, X. Di, G. A. Daigle, and M. R. Stinson, ''Spatial coherence of a sound field in a refractive shadow: Comparison of simulation and experiment,'' J. Acoust. Soc. Am. **98**, 2289–2302 (1995).

Long range source localization from single hydrophone spectrograms

W. A. Kuperman and G. L. D'Spain
Marine Physical Laboratory, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093-0701

K. D. Heaney
SAIC, 888 Prospect Street, La Jolla, California 92037

(Received 18 April 2000; revised 20 September 2000; accepted 14 November 2000)

A source near the deep sound channel axis excites mode groups (or paths) that involve both deep sound channel and boundary interacting propagation. Dispersion from a broadband source as measured on a single hydrophone can be used to estimate source range. Furthermore, modal group speeds have a functional transition when passing through purely refractive to boundary reflecting phase speed regions which, under certain conditions, provides additional arrival structure to aid in source localization. This additional arrival structure is in the form of a focal region in a spectrogram. Indeed, different data sets from the Acoustic Thermometry of the Ocean Climate (ATOC) Program [ATOC Consortium, Science **281**, 1327–1332 (1998)] show that localization can be accomplished using this focal region and/or the overall dispersion properties as originally suggested fifty years ago [M. Ewing and J. L. Worzel, Geo. Soc. Am., Memoir **27** (1948)]. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1339829]

PACS numbers: 43.30.Bp, 43.30.Wi, 43.30.Pc [DLB]

I. INTRODUCTION

Fifty years ago Ewing and Worzel[1] suggested that a source could be localized in the ocean with an estimated accuracy of 30 miles at 1000 miles using waveguide dispersion physics. More recently, there has been a rekindled interest in very long range acoustic propagation in the ocean.[2,3] It is well known that quantitative and qualitative information about sound propagation in a waveguide can be obtained by studying the modal (or ray) group speed (or cycle distance) and phase speed (or launch angle). We can consider three manifestations of these quantities:

(1) Total dispersion of a pulse;
(2) focusing with respect to frequency of a fixed mode;
(3) focusing of a group of modes at a fixed frequency.

A. Total dispersion

The first case is simply the Ewing and Worzel suggestion that one can obtain the range from the total dispersion of a pulse if the total span of group speeds are known from the environment. It turns out that one can also perform source localization by considering a combination of cases 2 and 3. The localization we are addressing is based on the relationship between group and phase velocities and is expected to be more environmentally robust than multihydrophone matched field processing (MFP) methods.[4] Alternative single phone localization algorithms related to MFP, and hence, based on precise Greens function replicas have already been addressed and evaluated using simulations.[5]

B. Airy phase

The physics of the second case is explained in original work of Pekeris. As cited in Ref. 6, there is a detailed discussion of the Airy phase referring to waves associated with a stationary value of the group velocity. For an isospeed, lossless air–water, bottom–fluid waveguide, now commonly referred to as the Pekeris waveguide, Pekeris plotted dimensionless group speeds, u_n/c_w versus a dimensionless waveguide parameter Hf/c_w where u_n is the group speed of the nth mode, c_w is the sound speed in the water layer, H is the depth of the water layer, and f is the acoustic frequency. An example is shown in Fig. 1. The Airy phase of each mode are the stationary minima; modal cutoff is to the left of the Airy phase, and the group speed to the right of the cutoff approaches the sound speed in the water column. In the arrival structure of a pulse wavetrain, the Airy phases only undergoes a geometrical spreading loss at range r of $r^{-5/6}$ as opposed to the larger loss of r^{-1} for other parts of the wavetrain. (Note that the r^{-1} factor is a combination of cylindrical spreading and dispersion but can also be seen from the point of view of separate ray arrivals with spherical spreading.) A simple physical explanation of the higher amplitude of the Airy phase is that only in the region of the stationary points in the group velocity curves is there *no* intramodal dispersion. Therefore, over a small frequency interval, all contributing frequency components arrive at the same time, a kind of frequency focusing. Each mode curve has a minimum group speed as a function of the waveguide parameter. For lower waveguide parameters, the group speed rises rapidly to the bottom sound speed indicating modal cutoff. In this region to the left of the Airy phase, modal cutoff combined with finite bottom attenuation results in the mode rapidly decaying with range. For larger waveguide parameters than where the Airy phase occurs, the group speed slowly rises asymptotically to the water column sound speed. A related phenomenon in geophysics is the *G* wave (a Love wave) showing a frequency region of near constant group

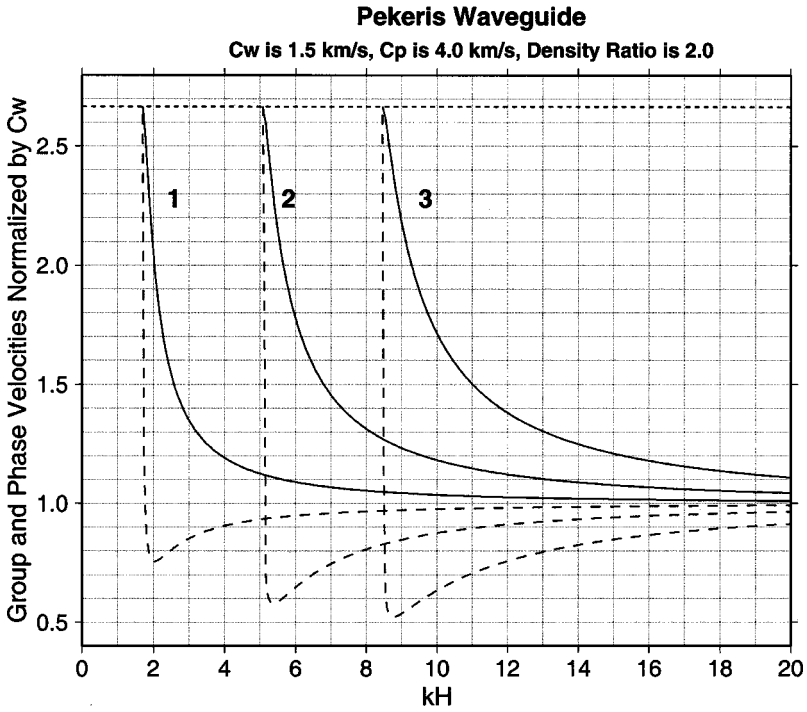


FIG. 1. Group and phase velocity plots for a Pekeris waveguide for the first three modes (labeled 1, 2, 3). The phase velocity curves are plotted as solid lines and the group velocity curves as dashed lines. The abscissa is a dimensionless parameter proportional to frequency and waveguide depth.

speed. Thus, for example, there exist Love waves that exhibit an impulse like arrival structure from distant earthquakes making the two-and-a-half hour roundtrip of the earth over a frequency interval corresponding to periods from 40–300 s traveling at a nearly constant group speed of 4400 m/s.[7,8]

C. Nondispersive mode group

The physics of the third case is contained in the group speed dependence on phase speed. More recent efforts[9–14] characterize the dependence of group speed on phase speed through a waveguide ''invariant'' denoted β. Though completely related to what follows, the full ''β'' invariant formalism discussed in these papers is not necessary to explain the phenomenon under discussion in this paper. On the other hand, a partial motivation of this study is the concept of a weakly divergent bundle of rays (WDBR).[15] In this latter case, regions of phase speeds exist in which the modes have the same group speed. This phenomenon is equivalent to the existence of launch angle intervals over which the cycle distances are constant. It is different than the Airy phase where a stationary region of group speed versus frequency exists for a single mode. For the WDBR case, there convergence zone ray bundles exist which do not diverge with range. Experimental confirmation was obtained by measuring the vertical distribution of the acoustic field at various ranges, i.e., the range-depth plane.[15]

D. Spectrogram focal regions

In this paper we search for a combination of cases 2 and 3 where many modes have a nondispersive stationary region (analogous to the Airy phase) but at the same phase speed interval (analogous to WDBR). For this case, we expect a bunching of energy in the frequency-time plane (spectrogram) caused by the group speed being approximately constant in an interval of phase speeds and frequencies. This phenomenon occurs, among other places, in a deep water scenario in the transition region from purely refracted paths (RR) to refracted-surface-reflected (or bottom-reflected-refracted) paths (RSR). This phenomenon should therefore occur for a more general class of sound speed profiles than that for the existence of WDBR. The surface-reflected/refracted transition was identified, for example, in the SLICE89 experiment[16] where a 3000 m vertical array was deployed. For one of the cases under study in this paper, we consider single phone arrival structure of a pulse which shows a concentration of energy at a specific time across a frequency interval at a single point in space.

After a discussion of the dependence of group speed on phase speed, we will compute the expected group speed dependence on phase speed for the environments associated with some of the data taken in the deep water ATOC series of experiments.[17–19] We show that a focal region in spectrograms corresponding to the above discussion is present in the ATOC acoustic data as displayed in a single phone spectrogram. Furthermore, we show a ''control'' example (case 1, above) of the ATOC data in which the total dispersion as suggested by Ewing and Worzel yields the correct range of the source. For this case the transition is from a refracted to bottom reflected path and no significant bottom interacting energy propagates to the long ranges under consideration. Clearly, we are able to explain the spectrogram structure in terms of the ocean environment and we demonstrate the use of these time markers when combined with the last axial arrival. Finally, we use our results to predict the utility of two single-phone receiving stations.

II. GROUP SPEED VERSUS PHASE SPEED FOR SIMPLE DEEP WATER PROPAGATION

Consider two families of modes or ray paths, RR and RSR, in a simple deep ocean environment as shown in the

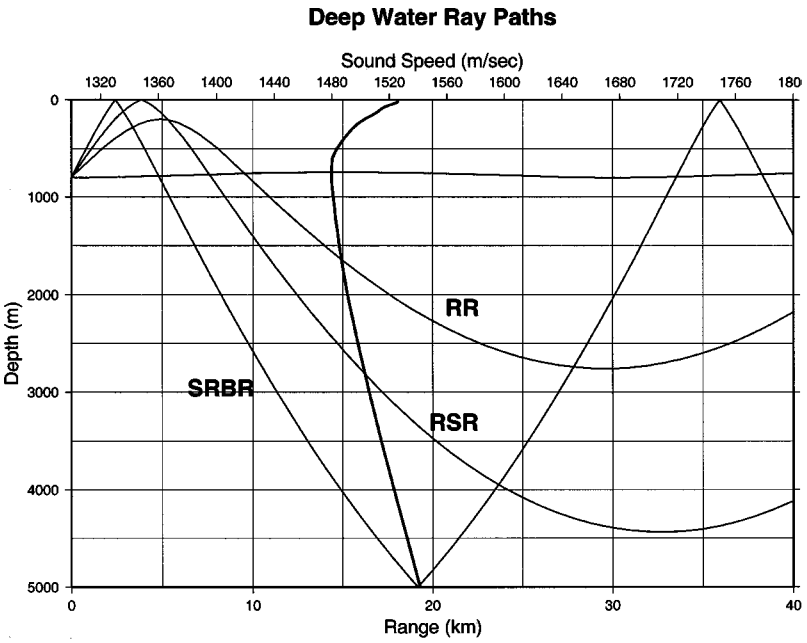


FIG. 2. Deep water ray trace examples for the sound speed profile (ssp) shown. The ssp scale is on the upper horizontal axis whereas the ray trace range scale is on lower horizontal axis. The discussion concerns the transition between the refracted refracted (RR) and the refracted-surface-reflected (RSR) paths. Also shown is the surface-reflected-bottom reflected (SRBR) path. Note that there is a positive critical depth meaning that there exists a RSR path; otherwise RR would transition RBR and then to SRBR.

ray schematic of Fig. 2. The sound speed profile used in this example is a range averaged profile for one of the environments where we will compare data and theory. For the purely refracted RR case the group speeds increase with increasing phase speed. (Recall that phase speed increases with increasing launch angle with respect to the horizontal.) That is, the up and down going rays are refracted in regions of higher medium speeds and this phenomenon dominates the group speed dependence on phase speed. For the RSR path, the upper surface reflects rays before they can be refracted at higher speed regions and hence this ray family has a different group speed dependence on phase speed.

We transition from RR to RSR by increasing phase speed of the modes (or launch angles in the ray picture). Clearly, as we transition from RR to RSR or vice versa, the group speeds (or cycle distances) must approach each other in the transition region. This means that contributions of a point source from modes or rays in this transition region should arrive at a receiver more or less at the same time. If, over the finite bandwidth of a pulse, the modes are only moderately dispersive, then it follows that the frequency-time arrival structure of a pulse should show a local maximum, or focus, in spectral level at an arrival time for all frequencies corresponding to the source-to-receiver range divided by the group velocity of the transition region. Lowering the frequency of a RR mode causes it to transition to a RSR mode. This is confirmed using ATOC data.

III. RANGE-DEPENDENT ATOC ENVIRONMENT

The data we examine in this paper comes from two different paths in the ATOC series of experiments.[17–19] The two paths are shown in Fig. 3 and the climatology derived sound

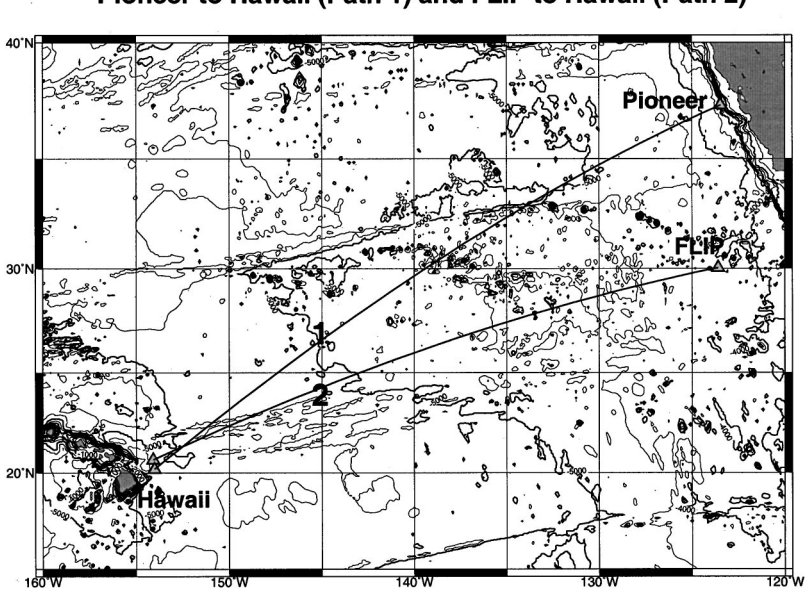
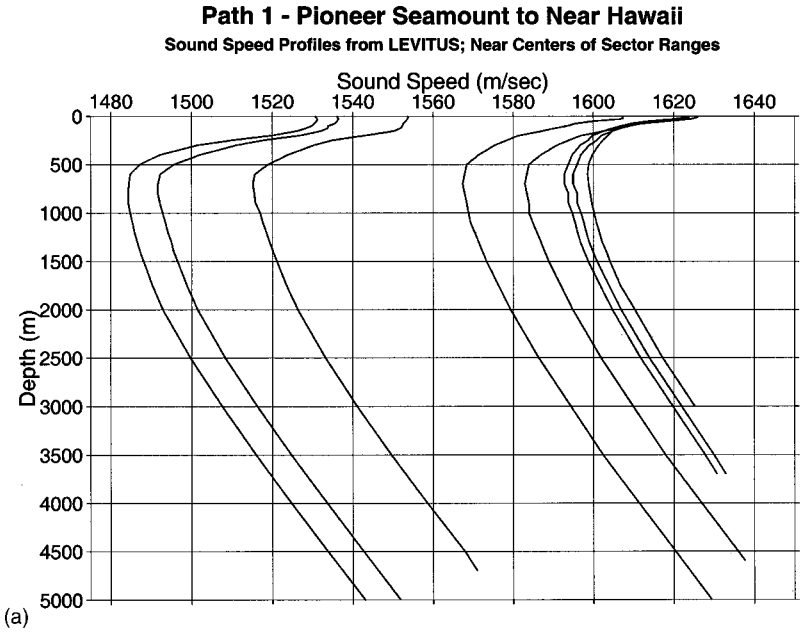


FIG. 3. Geographic specification of the two long range propagation paths under consideration.



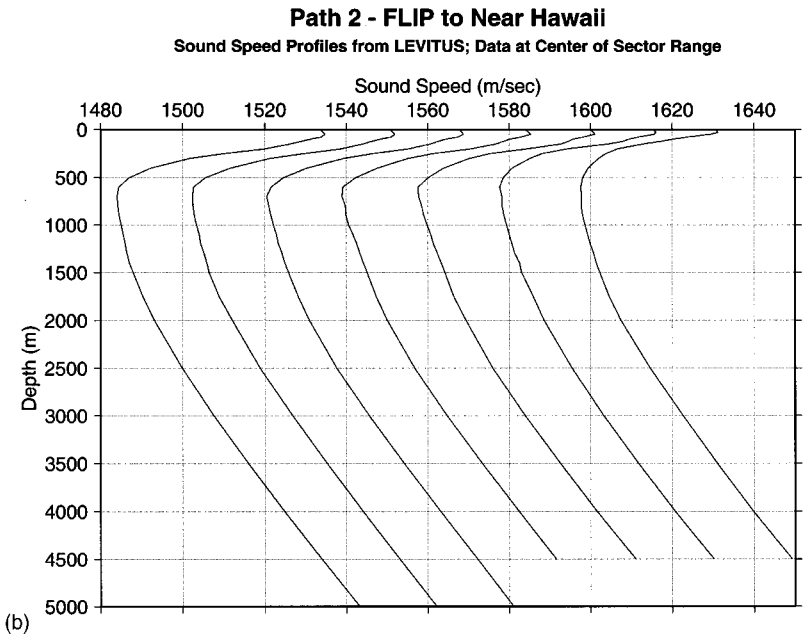


FIG. 4. Climatology derived sound speed profiles. (a) SSPs for path 1; (b) SSPs for path 2.

speed profiles are shown in Fig. 4. We calculate the effective range-dependent group speeds over these paths as described below.

We obtain the curves in Fig. 5 from an adiabatic normal mode computation[20] and approximate the phase speed to be a continuous variable whereas in reality it takes on discrete modal values. This is a good approximation because there are of order 100–150 modes in the phases speed interval shown on the abscissa in Fig. 5. The local phase speed of the nth mode at the source location (which corresponds to a launch angle) is related to the modal wave number k_n:

$$c_{pn}(\omega)=\frac{\omega}{k_n}, \tag{1}$$

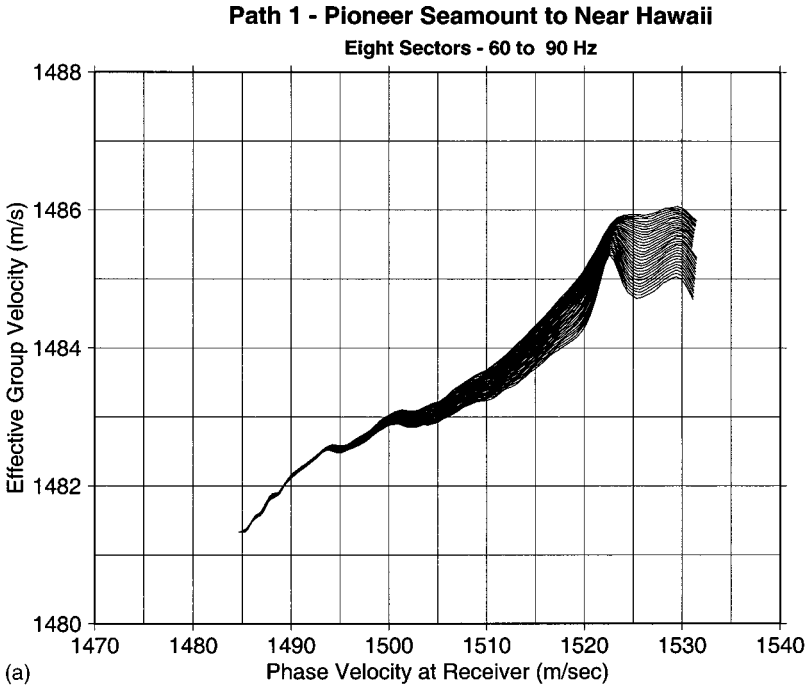
where ω is the angular frequency of the propagating acoustic mode. In a range-dependent environment, we must derive an effective average modal group speed, u_n, from the range averaged slowness. That is, we note that the pulse observables at range r are arrival times which are given by r/u_n. Hence, for a range-dependent environment we are concerned with $\overline{1/u_n}\equiv\overline{s_{gn}}$ where the overbar denotes range averaging and s_{gn} is referred to as the (local) modal group slowness. Since the adiabatic mode approximation has a range averaged modal wave number associated with each term, it follows that[20]

$$\overline{s_{gn}(r,\omega)}=\frac{\partial}{\partial\omega}\left(\frac{1}{r}\int_0^r k_n(r',\omega)dr'\right) \tag{2}$$

and the range-dependent effective group speed is

$$\langle u_n(r,\omega)\rangle=\frac{1}{\overline{s_{gn}(r,\omega)}}. \tag{3}$$

Hence, the range-dependent effective group speed is obtained from the harmonic average:



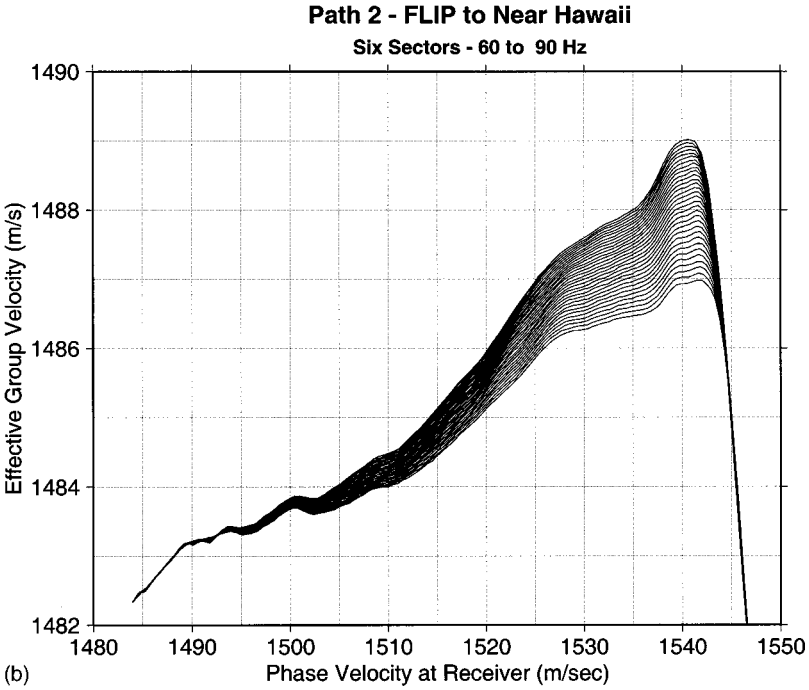


FIG. 5. Group velocity as a function of phase velocity curves as parameterized by frequency; (a) path 1; (b) path 2.

$$\frac{1}{\langle u_n(r,\omega)\rangle} = \frac{1}{r}\sum_{\Delta r_n'} \frac{\Delta r_n'}{u_n(r',\omega)}, \tag{4}$$

where the sum is taken over the approximate n range-independent subintervals, $\Delta r_n'$, of the propagation path. Finally, we note that we can interchange the partial with the integral sign in Eq. (2) and can use the modal formula for the (local or range independent) modal group speed:[21]

$$\frac{\partial k_n}{\partial \omega} = \frac{1}{u_n(\omega)} = \frac{\omega}{k_n(\omega)} \int_0^\infty \frac{p_n^2(\omega;z)}{\rho(z)c^2(z)}\, dz, \tag{5}$$

where $p_n(\omega;z)$ are the normal mode eigenfunctions of the pressure field in the ocean waveguide and $\rho(z)$ and $c(z)$ are the density and sound speed profiles as a function of depth.

IV. SPECTROGRAMS OF THE ATOC ACOUSTIC DATA

Figure 6 shows the spectrograms for the two paths displayed in Fig. 3. These spectrograms were obtained by adding 32 successive pulses (actually, m sequences were used and the pulse compression was obtained from the matched filter output) over a time interval of 20 min (path 1) and 13.8 min (path 2).

In both cases we see the well-known last arrival which travels with the slowest group velocity u_s representative of

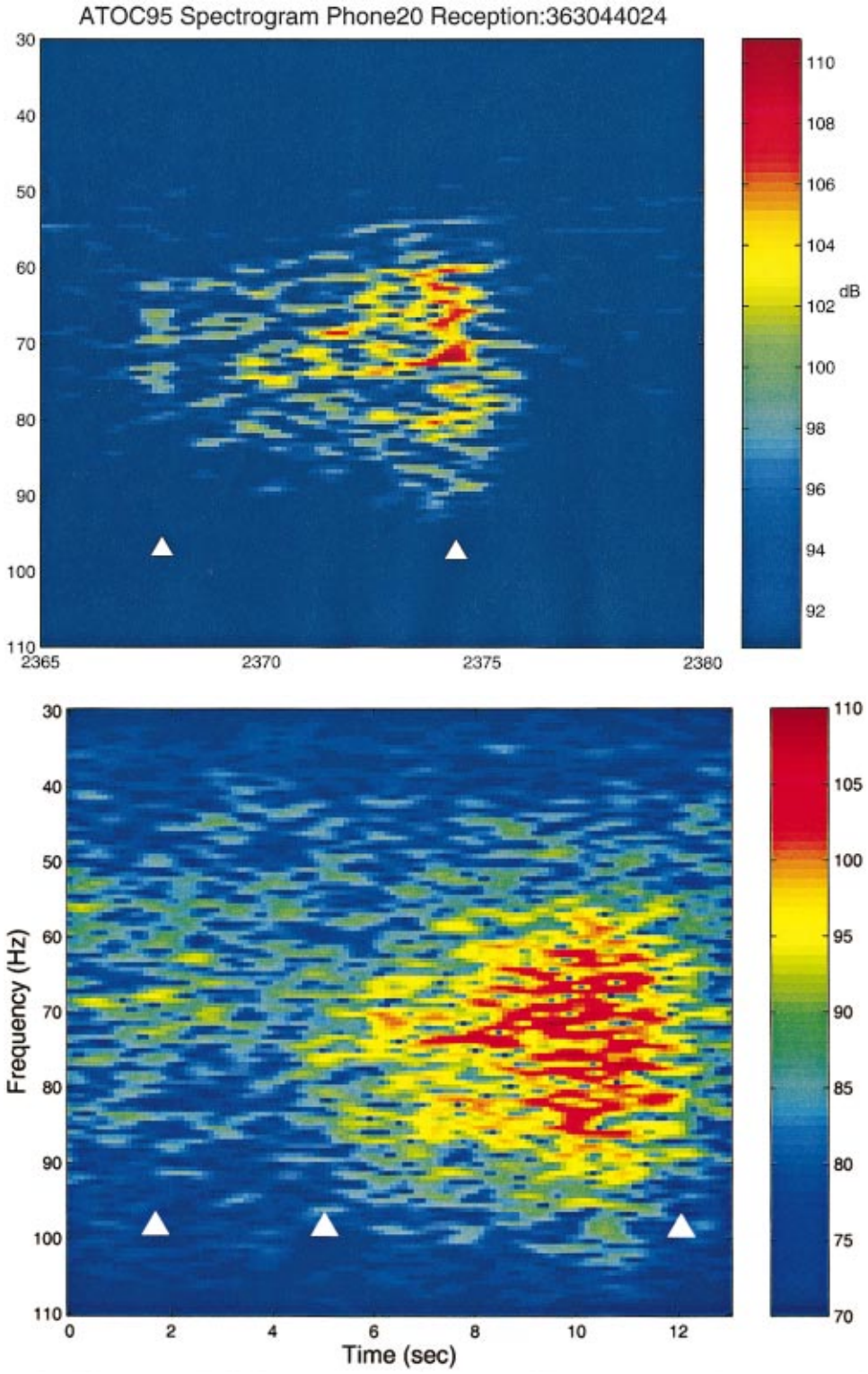


FIG. 6. Spectrograms for paths 1 and 2. (Top) The path 1 spectrogram for a source and receiver depth of 939 and 991 m, respectively, shows an early focusing corresponding to the RR–RSR transition (bottom). For path 2 with a source and receiver depth of 650 and 935 m, respectively, the initial arrival comes from the path with the highest phase speed not interacting with the ocean bottom.

the axial sound speed. Further, we see the earlier arrivals traveling with the faster group velocity u_f. For both cases, u_s and u_f can be taken from the corresponding plot in Fig. 5. We obtain the range from the simple relation (time of travel is the difference of distance divided by speed for the two group speeds):

$$r = \frac{u_s \Delta t}{1 - \frac{u_s}{u_f}}, \tag{6}$$

where Δt, the travel time difference, is read off the spectrograms. The times for each focal region are indicated on the spectrograms with white arrows. An intermediate focal region which provides a second range estimate for path 2 is identified in the lower panel of Fig. 6. The results are tabulated in Table I.

We have therefore confirmed that the transition region between the RR and RSR paths corresponded to a focal region in a spectrogram. In the path 2 case, this transition is confined to a fairly narrow frequency band, as shown by the

TABLE I. Results for Paths 1 and 2.

Path	u_s	u_f	Δt	Estimated range	Actual range
1	1481.3 m/s	1485.6 m/s	6.3 s	3429±100 km	3514 km
2	1482.3 m/s	1488.7 m/s	10.1 s	3482±250 km	3252 km
3	1482.3 m/s	1487.2 m/s	6.7 s	3032±250 km	3252 km

first arrivals in the lower panel in Fig. 6. Their narrowband nature is explained by the significant spread in the maximum values of the individual frequency effective group velocity curves in the lower panel in Fig. 5. In any case, this focal region combined with the axial arrival and climatological data permits localization of a source over a very long range. Intermediate focal regions may exist and can be used to provide additional range information.

This single element spectrogram technique for estimating source range is valid only when both the source and receiver are located near the sound channel axis. For the data just presented, the source and receiver depths for path 1 are 939 and 991 m, respectively, and the corresponding depths for path 2 are 650 and 935 m. These depths are sufficiently close to the sound channel axis that accurate source ranges are obtained. In contrast, Fig. 7 shows a path 1 spectrogram for a receiving element at 326 m depth. Comparison with Fig. 6 (top) shows that the earliest focal region occurs at exactly the same arrival time in both figures. However, the latest arriving focus comes in about 1.5 s earlier for the receiver at 326 m than the one at 991 m, due to the greater medium sound speed at the shallower depth. This 1.5 s difference would result in an underestimate of the source range by 23% if the same group velocities are used in the calculation.

A constructive/destructive interference pattern as a function of frequency is also evident in the spectrograms, particularly in the later-arriving focal region, in Figs. 6 and 7. This pattern arises from the mutual interaction between modes arriving at the same time (i.e., same group velocity), but having different phase velocities. If the difference in frequency between two peaks in the interference pattern is denoted Ω_c, then

$$\Omega_c = \frac{1}{r\Delta s_p^{mn}}, \tag{7}$$

where Δs_p^{mn} is the largest difference in mode phase slowness for those modes that contribute significantly to the field. The structure of the interference patterns in Figs. 6 and 7 suggest a difference in phase velocity on the order of 1 part in 10^4 th, consistent with the curves in Fig. 5. When the modes' phase slowness is a function of frequency, Δs_p^{mn} is replaced by the group slowness difference; Eq. (8) of Ref. 14.

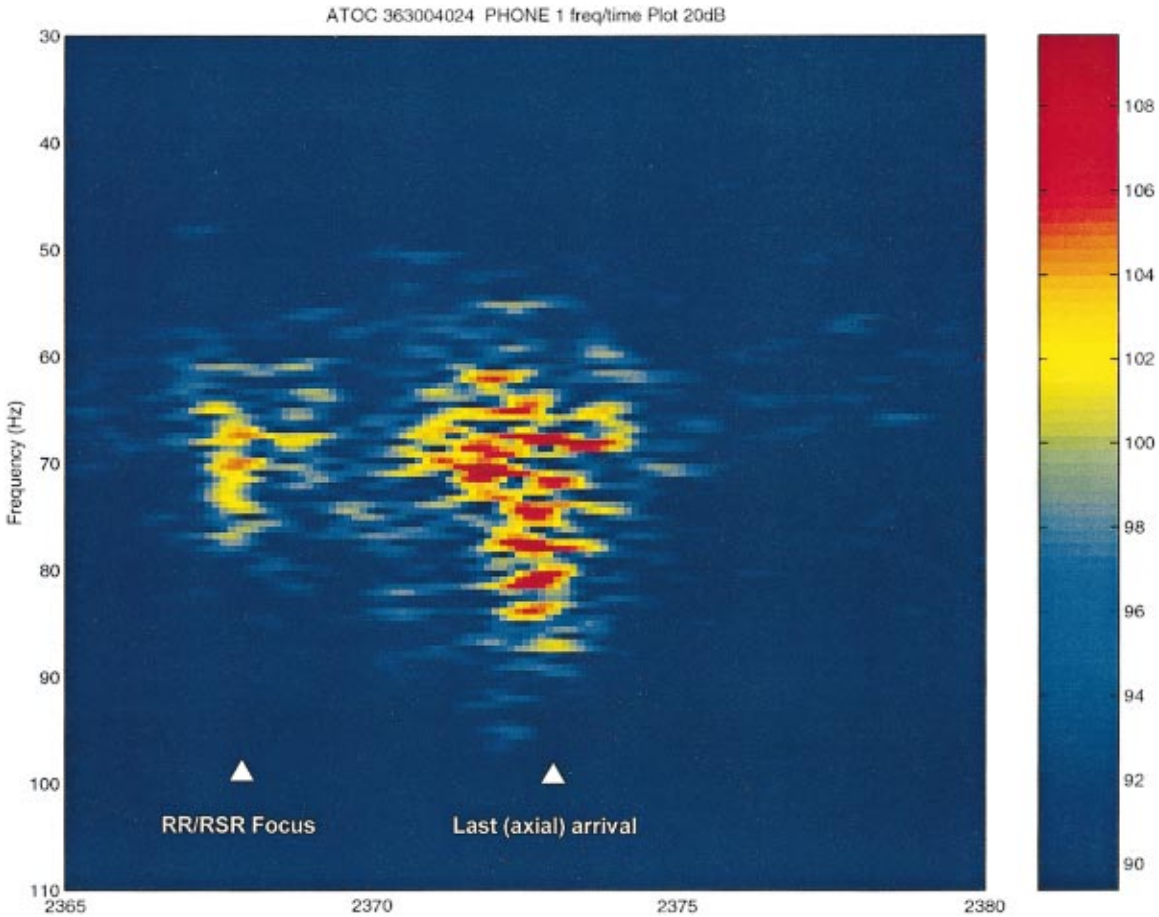


FIG. 7. Spectrograms for path 1 for a receiver at a depth of 326 m.

V. EXTRAPOLATION OF RESULTS TO TWO SINGLE-PHONE RECEIVING STATIONS

The purpose of this section is to investigate the extension of the results just presented to situations where data from two horizontally separated omnidirectional hydrophones are available. Such data may be obtainable from the array geometries envisioned for the hydrophone stations in the International Monitoring System (IMS),[21] provided the source location is such that the propagation paths are not blocked by the surrounding bathymetry.

In a dispersive medium, the arrival from an impulsive source has a duration, τ, that is determined by the range to the source and the difference in group slownesses, Δs_g, between the slowest and fastest components, i.e.:

$$\tau = r \Delta s_g . \tag{8}$$

These components either can be different frequency components of a given mode (multipath), e.g., the ''water wave'' and the Airy wave for a single normal mode in shallow water waveguide propagation,[6] or different multipath components, as presented in the previous section [Eq. (6)]. In either case, prior knowledge of the environment, represented by Δs_g, is required to estimate source range. On the other hand, if data from two horizontally separated receivers are available and the medium is approximately range independent, then the differential spreading of the pulse between receivers, $\Delta\tau$, can be used to estimate source range without requiring *a priori* environmental information. That is, applying Eq. (7) to the data from both receivers, and combining to eliminate Δs_g gives:

$$\frac{\Delta r}{r} = \frac{\Delta \tau}{\tau}. \tag{9}$$

The quantity, Δr, is determined by the separation between the two receivers, d, and the direction towards the source, θ, i.e., $\Delta r = d \cos(\theta)$. The source bearing can be obtained from the difference in total pulse arrival times at the two receivers, e.g., Ref. 22.

One difficulty in applying Eq. (8) to actual ocean acoustic data is that the differential pulse spreading from one receiver to the next may be hard to measure accurately. For the technique to work, the two components of the signal whose arrival times are being compared must be the same at the two receivers. An example where conditions are favorable is when these two components are pulses that have good signal-to-noise ratio (SNR) and that retain their shape as they propagate with range. However, if either or both of the pulses disperse (spread out in time) with increasing range, then the ability to identify them and/or their SNR degrades. The use of the focal regions in spectrograms eliminates this problem to a large extent since the focal regions do not disperse.

A second difficulty is that the horizontal separation between the two receivers must be quite large for pulse spreading over deep ocean propagation paths to be discernible. For example, assuming that a 10% change in pulse spreading can be measured, then the horizontal receiver separation must be order 100 km to permit estimation of the range to nominally 1 mm distant sources. Note, however, these separation distances are comparable to those planned for the pair of hydrophone triads at the IMS hydrophone stations.[23] For sources at shorter range and/or for shallow water waveguide propagation, the requirements on the receiver separation are less stringent.

Finally, the method described in Sec. III assumes the source is located near the sound channel axis, so that it excites those mode groups that create the single spectrogram focus regions having maximum temporal separation. For sources at other depths, a different set of mode groups may be excited. Therefore, the duration of the received pulse at long range is dependent upon source depth, e.g., Ref. 23. The differential pulse spreading technique described in this section provides the identification of the pair of mode groups that determine the received pulse duration, thus yielding information on the depth of the source as well as its range.

VI. DISCUSSION AND CONCLUSION

First we must emphasize that the results of this paper were for a high signal to noise ratio at the receiver. The source level was about 195 dB 1 μPa @ 1 m. From Fig. 6, we see that the SNR was at least 20 dB. Furthermore, this signal to noise ratio was obtained by coherently adding the pulses over a 20 (path 1) or 13.8 (path 2) min time period. Hence, we had a loud source with a stable temporal structure. From an underwater acoustic signal processing point of view, this is an atypical situation. Nevertheless, it is quite amazing that a single receiver can localize a source at such a large range of 3500 km using only archived climatological data. Estimates of source range from a vertical array using these data have also been previously presented.[24] This single receiver method, described in this paper, might be applicable to recent efforts related to the Comprehensive Nuclear Test Ban Treaty (CTBT)[25] in which, among other things, it is desired to be able to localize very loud sources in the ocean. Although energy arriving from earthquakes and other naturally occurring events typically are confined to frequencies below 50 Hz, arrivals from in-water detonations often have frequency content up to 100 Hz or more, even after propagating over very long oceanic paths. Therefore, the present plan for the hydrophone stations in the hydroacoustics network of the International Monitoring System specifies a data sampling rate of 240 samples/s. The 60–90 Hz band of the ATOC data presented in this paper falls completely within the resulting 100 Hz bandwidth of these monitoring stations. Furthermore, as discussed above, using two single-phone receiving stations reduces the requirement of having accurate climatology data.

Finally we mention what we have not addressed in this paper. First, the intensities of the focal regions fluctuate with time. In particular, the RR/RSR paths travel through regions of significant internal wave activity and these may be a diagnostic of the internal wave field. Second, we were only concerned in this paper with either deep water source-receiver configurations or when the source is in a steep slope region so that the paths are essentially deep water paths. For sources (or receivers) that result in even a marginal propagation distance in bottom interacting paths, there is a significant effect on the group/phase speed relation. With respect to

the CTBT problem, archived bottom bathymetry can provide additional data to perform source localization. We will address these issues in future research.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research. The data were kindly provided to us by the ATOC Group (A. B. Baggeroer, T. G. Birdsall, C. Clark, J. A. Colosi, B. D. Cornuelle, D. Costa, B. D. Dushaw, M. A. 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). The ATOC Project was supported by the Strategic Environmental Research and Development Program through Defense Advanced Research Projects Agency (DARPA) Grant No. MDA972-93-1-0003. We would also like to thank Ed McDonald at the SACLANTCEN for pointing out the interference patterns in the ATOC spectrograms.

[1] M. Ewing and J. L. Worzel, ''Long-range sound transmission,'' Geo. Soc. Am., Memoir **27** (1948).
[2] K. D. Heaney, W. A. Kuperman, and B. E. McDonald, Perth-Bermuda sound propagation (1960); ''Adiabatic mode interpretation,'' J. Acoust. Soc. Am. **90**, 2586–2594 (1991).
[3] W. Munk and A. Baggeroer, ''The Heard Island papers: A contribution to global acoustics,'' J. Acoust. Soc. Am. **90**, 2327–2329 (1991); this is the introductory paper of a special issue which contains 17 papers.
[4] A. B. Baggeroer, W. A. Kuperman, and P. N. Mikhalevsky, ''An overview of matched field methods in ocean acoustics,'' IEEE J. Ocean Eng. **18**, 401–424 (1993).
[5] L. N. Frazer and P. I. Pecholcs, ''Single-hydrophone localization,'' J. Acoust. Soc. Am. **88**, 995–1002 (1990).
[6] W. M. Ewing, W. S. Jardetzky, and F. Press, *Elastic Waves in Layered Media* (McGraw Hill, New York, 1957).
[7] G. D. Garland, *Introduction to Geophysics* (W. P. Saunders, Philadelphia, 1979).
[8] K. Aki and P. G. Richards, *Quantitative Seismology: Theory and Methods* (W. H. Freeman, San Francisco, 1980), Vol. 1.
[9] S. D. Chuprov, in *Ocean Acoustics, Current State*, edited by L. M. Brekhovskikh and I. B. Andreevoi (Nauka, Moscow, 1982), pp. 71–91.
[10] L. M. Brekhovskikh and Y. P. Lysanov, *Fundamentals of Ocean Acoustics*, 2nd ed. (Springer, New York, 1991).
[11] V. N. Golubev and V. N. Fokin, ''Space-frequency dependence of the parameters of interference modulation of wideband sound in a surface channel,'' Sov. Phys. Acoust. **37**, 233–237 (1991).
[12] V. N. Lobanov and Yu. V. Petukhov, ''Space-frequency distribution of the intensity of wideband sound in a shallow sea,'' Acoust. Phys. **39**, 574–581 (1993).
[13] G. A. Grachev, ''Theory of acoustic wave invariants in layered waveguides,'' Acoust. Phys. **39**, 33–35 (1993).
[14] 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).
[15] L. M. Brekhovskikh, V. V. Goncharov, and V. M. Kurtepov, in *Full Field Inversion Methods in Ocean and Seismo-Acoustics*, edited by O. Diaschok (Kluwer Academic, Dordrecht, 1995).
[16] P. F. Worcester, B. D. Cornuelle, J. A. Hildebrand, 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**, 3118–3129 (1991).
[17] ATOC Consortium, ''Ocean climate change: Comparison of acoustic tomography, satellite altimetry, and modeling,'' Science **281**, 1327–1332 (1998).
[18] 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-scale vertical array at 3250-km range in the eastern North Pacific,'' J. Acoust. Soc. Am. **105**, 3185–3201 (1999).
[19] J. A. Colosi, E. K. Sheer, S. Flatte, 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,'' J. Acoust. Soc. Am. **105**, 3202–3218 (1999).
[20] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (AIP, Woodbury, NY, 1994).
[21] J. Newton and P. Grenard, Proc. Informal Workshop on Hydroacoustics, Papeete, Tahiti, 1999, p. 30.
[22] J. A. Hanson, ''Seismic and hydroacoustic investigations near Ascension Island,'' Ph.D. thesis, University of California, San Diego, CA, 1998.
[23] P-F. Piserchia, D. Rodriques, J. Virieux, and S. Gaffet, Detection of underwater explosion at very long range, Oceans '98, Nice, France (1998).
[24] K. D. Heaney and W. A. Kuperman, ''Very long-range source localization with a small vertical array,'' J. Acoust. Soc. Am. **104**, 2149–2159 (1998).
[25] M. W. Lawrence, ''Overview of the hydroacoustic monitoring system for the comprehensive nuclear-test-ban treaty,'' J. Acoust. Soc. Am. **105**, 1037 (1999).

Excitation of *T*-phases by seafloor scattering

Catherine D. de Groot-Hedlin and John A. Orcutt
Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093-0225

(Received 8 December 1999; revised 28 August 2000; accepted 30 January 2001)

T-phases excited by suboceanic earthquakes are classified into two types: abyssal phases which are excited near the earthquake epicenter at seafloor depths far below the SOFAR velocity channel, and slope *T*-phases which are excited at continental, or ocean island slopes and ridges at distances up to several hundreds of kilometers from the epicenter. In this article, it is demonstrated that approximate time–frequency characteristics of both classes of *T*-phase can be synthesized under the assumption that *T*-phases are excited by scattering from a rough seafloor. Seafloor scattering at shallow depths preferentially excites low order acoustic modes that propagate efficiently within the ocean sound channel minimum. At greater depths, scattering excites higher order modes which interact weakly with the seafloor along much of the propagation path. Using known variations in near-source bathymetry, *T*-phase envelopes are synthesized at several frequencies for several events south of the Fox Islands that excited both types of *T*-phase. The synthesized *T*-phases reproduce the main time vs frequency features of each type of arrival; a higher frequency, nearly symmetric arrival excited near the epicenter and a longer duration, lower frequency arrival excited near the continental shelf, with a peak amplitude at about 5 Hz. © *2001 Acoustical Society of America.*
[DOI: 10.1121/1.1361057]

PACS numbers: 43.30.Bp, 43.30.Qd, 43.30.Hw, 43.30.Ma [SAC-B]

I. INTRODUCTION

Seismic events below the seafloor generate ocean-borne acoustic phases, called *T*-phases, that propagate great distances within the ocean sound channel minimum with little transmission loss. Because of their efficient propagation, *T*-phases can be detected for low magnitude events that are not observable at land-based seismic networks.[1] Therefore *T*-waves have been used to investigate swarms of tectonic activity associated with seafloor spreading at the Mohns Ridge,[2] at the Gorda Plate,[3] and at the Juan de Fuca Ridge.[4] Further research into *T*-phases has been stimulated by the need to monitor compliance with the Comprehensive Nuclear Test Ban Treaty (CTBT); a hydroacoustic network is being installed as part of the International Monitoring System (IMS) to detect and locate events that generate hydroacoustic phases. Methods to discriminate between explosions and naturally occurring tectonic events are currently being developed.[5] Ultimately, a better understanding of the physics of *T*-phase excitation can aid both in geophysical investigations and in monitoring the CTBT.

Early hydroacoustic recordings indicated that *T*-phases are excited in shallow regions where the seafloor is sloping, i.e., near islands or submarine promontories, and not necessarily at the earthquake epicenters.[6,7] This ''slope'' *T*-phase often features multiple peaks associated with bathymetric highs in the vicinity of the epicenter,[6–8] thus the source region can often be recognized by its signal characteristics.[9] Another class of seismically generated acoustic phases, with source locations coincident with earthquake epicenters far below the ocean sound channel, was later identified.[7,10] These ''abyssal'' *T*-phases are characterized by a higher dominant frequency than that of the slope *T*-phase, and by symmetric coda about the peak frequency arrivals.[11] Abyssal *T*-phases are often weaker than those generated in shallow regions, further from the epicenter.[9]

The distinct characteristics of the slope and abyssal *T*-phases are widely thought to derive from distinct seismic to acoustic coupling mechanisms (e.g., Refs. 7, 10, and 11). Excitation of *T*-phases in shallow regions near islands or near the continental shelf is usually attributed to downslope propagation,[12,13] which occurs where steep slopes intersect the sound channel. In this description, acoustic energy is refracted nearly vertically into the ocean column and is transformed to a horizontally propagating phase by multiple reflections between the sea surface and the sloping seafloor. However, in Ref. 14 we showed that for a gently sloping, sediment covered seafloor, the shape of the *T*-phase envelope is consistent with a coupling mechanism that is dependent mainly on depth rather than slope, and attributed the excitation of the slope *T*-phase to seafloor scattering. Excitation of the abyssal *T*-phase has variously been attributed to coupling between Stoneley waves and the SOFAR sound channel,[15] and reflection scattering from either the seafloor or sea surface,[10] or from the overlying sea ice.[11]

Although downslope propagation may be an efficient *T*-phase excitation mechanism for a steeply sloping, rocky seafloor, it is far less efficient for a gently sloping, sediment covered environment typical of continental shelves. Given that water-saturated seafloor sediments cannot support steep slopes, a large number of seafloor reflections would be required to transform the energy into the horizontal direction. For velocities typical of the sediment covered seafloor,[16,17] reflection coefficients range from 0.2–0.4, thus most of the specularly reflected energy is lost.

In this article we demonstrate that, for a sediment covered seafloor with shallow slope, we can numerically synthesize the general frequency vs time characteristics of both

classes of *T*-phase under the assumption that they are excited primarily by small scale scatterers at a rough seafloor. In the next section we show that, since seafloor scatterers excite acoustic modes in proportion to their intensities at the ocean floor, only leaky modes which interact with the seabottom can be excited. Scattering at shallow depths excites low order acoustic modes that initially propagate in both the ocean column and sedimentary layer, but cease interaction with the seafloor as they propagate seaward. Higher modes, which fill up more of the sound channel and are thus more prone to attenuation at seamounts and ridges, are preferentially excited at greater depth. For several events south of the Fox Islands, we compute approximate source locations for peak *T*-wave arrivals and find that the seismic velocity corresponding to arrivals excited at a distance from the epicenter is consistent with shear wave propagation. Finally, we use known variations in near-source bathymetry to synthesize *T*-wave envelopes at several frequencies, and demonstrate that these show broad agreement with observed *T*-phases.

II. THE SEAFLOOR SCATTERING MODEL

Given the significant velocity contrast between rock at earthquake source depths and the sediment covered seabottom, seismic energy is refracted nearly vertically upwards at the crust–sediment interface. Any method of numerically synthesizing the time–frequency characteristics of earthquake-generated *T*-phases must account not only for the physics of coupling from nearly vertically propagating seismic waves to nearly horizontally propagating acoustic waves, but also the effects of seismic propagation through the attenuative crust and upper mantle, and acoustic transmission blockage by bathymetric features within the ocean column. As in Ref. 14, we assume that *T*-phase excitation along the continental shelf is dominated by seafloor scattering, and that slopes are sufficiently gentle that acoustic propagation through the ocean waveguide is approximately adiabatic (i.e., we neglect mode coupling). If the seafloor is treated as a sheet of point scatterers, the pressure contribution from each point is given by the energy-conserving adiabatic mode solution to the wave equation in a stratified waveguide, i.e.,

$$p(r_r,z_r)=\frac{1}{\sqrt{|r_r-r_s|}}\sum_{m=1}^{\infty}\Psi_m(r_s,z_s)\Psi_m(r_r,z_r)\times\frac{\exp(i\int k_m(s)ds)}{\sqrt{k_m|r_r-r_s|}}, \tag{1}$$

where $\Psi_m(r_s,z_s)$ and $\Psi_m(r_r,z_r)$ are the mode functions at the source and receiver locations, respectively, k_m is the complex wave number, and the summation is over the mode number m.[18] For seafloor scatterers, the source depth is equal to the seafloor depth. Thus a point scatterer on the ocean floor excites modes in proportion to the mode amplitude at the ocean floor. Thus only leaky modes, which propagate in both the ocean column and the ocean bottom, can be excited by scattering of seismic energy into acoustic energy at the seafloor.

Energy is lost in the propagation of leaky modes as long as they interact with the seafloor. Given that k_m is complex valued, the exponential term in Eq. (1) yields along path transmission losses. However, we chose instead to use a result from perturbation theory[19] to compute transmission loss; in this formulation, the attenuation of a given acoustic mode is estimated by the fraction of that mode's energy in the attenuative bottom. The total bottom attenuation along the transmission path from source to receiver is then given in dB by

$$\mathrm{dB}_{\mathrm{loss}}=\beta f\int_{\mathrm{receiver}}^{\mathrm{source}}dr\int_{\infty}^{z_b}\Psi_i^2(z)dz, \tag{2}$$

where β is the attenuation coefficient within the sedimentary layer, expressed in dB/km/Hz, and $z_b(r)$ is the bathymetric depth at any given horizontal location.[20] Bottom losses thus trade off with energy excitation for any given mode and frequency. As seafloor depth increases, the attenuation of a given mode decreases and, at sufficient depths, the attenuation is negligible as the mode is transmitted entirely in the ocean column. An attenuation coefficient β of 0.10 db/km/Hz was used, consistent with mean values for a sedimentary bottom.[15]

The total pressure from the entire sheet of point scatterers is given by the summation over all points in the vicinity of the epicenter, i.e.,

$$\sum_{\mathrm{area}}p(r_r,z_r)\left[\frac{e^{-d_s*\alpha*f}}{d_s}\right]f^n s(\theta), \tag{3}$$

where r_e is the epicentral location, z_e is the hypocentral depth, $d_s=\sqrt{(z_e-z_s)^2+(r_e-r_s)^2}$ is the distance from the earthquake hypocenter to the seafloor scatterer, f is the frequency, α is the coefficient of attenuation within the crust and mantle, and $s(\theta)$ is the scattering coefficient as a function of the angle of incidence on the rough interface. The term in square brackets above describes the pressure incident from below on the rough seafloor, i.e., pressure decreases due to spherical spreading and frequency-dependent attenuation within the crust and upper mantle. We use an α value of 0.02 dB/km, consistent with mean values for oceanic crust.[17]

The forms of the scattering function and the frequency power law depend on the relative scale sizes of the scattering bodies and the wavelength of the sound energy. For scattering bodies that are small compared to the incident field wavelength, the scattered field can be approximated by the method of small perturbations (MSP), first used by Rayleigh[21] to investigate scattering of sound at irregular surfaces. The MSP predicts that the intensity of the scattered field varies with the fourth power of frequency, i.e., pressure varies as f^2, and that the scattered field is nearly omnidirectional.[22] The tangent plane method, also called the Kirchoff approximation, is used for large scale roughness. In this case it can be shown that scattered intensities are independent of the sound frequency, i.e., $n=0$ in Eq. (3). For the small slopes typical of sedimentary seafloors, it can be shown that the scattered field is concentrated in the specular direction for large scale scatterers.[22]

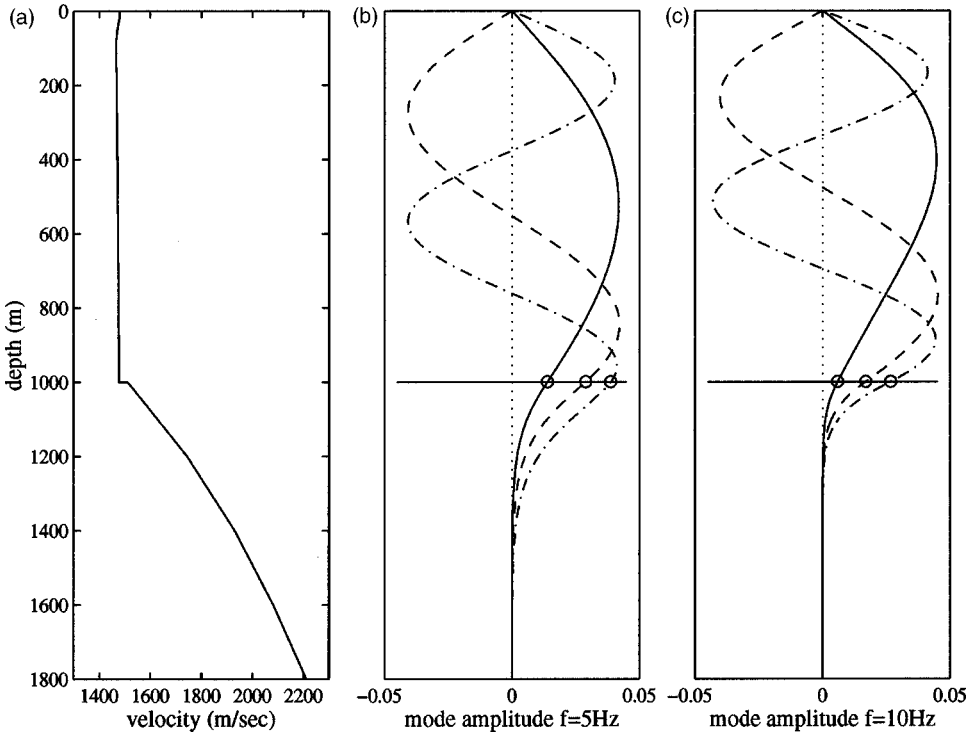


FIG. 1. (a) The velocity profile used to compute acoustic modes for an ocean depth of 1 km. The top km of the profile corresponds to annual average velocities near the epicenters; the bottom part of the velocity profile conforms to a typical seafloor sedimentary layer. (b) The first three modes computed at 5 Hz for the given velocity profile. Mode 1 is indicated by the solid line, mode 2 by the dashed line, and mode 3 by the dash–dotted line. (c) Mode amplitudes computed at 10 Hz.

The seafloor is generally characterized by roughness over a wide range of scale lengths.[23] For a rough interface with both small-scale and large-scale irregularities, it can be shown that scattering due to large-scale roughness dominates at low angles of incidence, and drops off rapidly with increasing angles.[22] Scattering due to small-scale irregularities becomes more important at higher angles of incidence. However, since scattering by large-scale inhomogeneities directs energy in the specular direction, we conclude that it is no more efficient at generating *T*-phases than downslope propagation. In contrast, the scattered sound field excited by small-scale scatterers is nearly omnidirectional; therefore we assume that, although only a small fraction of the total incident field is scattered, *T*-phases are primarily excited by small-scale scattering for gently sloping, sediment covered seafloors typical of continental shelf regions. Thus pressure amplitudes vary as f^2, and since scattering strengths are only weakly dependent on angle of incidence, we set $s(\theta)=1$ in Eq. (3) above.

A more complete description of scattering at a rough seabed, including the effects of nonzero shear velocities in the seafloor,[24] is beyond the scope of this article. However, it should be noted that energy can only be scattered approximately in proportion to the *S* to *P*, or *P* to *P* transmission coefficients. Thus, normally incident shear waves on the ocean waveguide cannot excite low grazing angle acoustic waves within the ocean waveguide. At higher angles of incidence, both *P* and *S* waves can excite *T*-phases, but the relative contribution will depend on the relative partitioning of *P* and *S* energy at the source which is unknown. Furthermore, since the radiated energy depends on the size of the scattering bodies, which is unknown, we assume in this study that seafloor roughness is uniform over a broad region in the vicinity of the epicenter. Therefore, we do not specify an absolute value of the scattered energy, but seek only to demonstrate that small-scale scattering yields the general time–frequency characteristics of both slope and abyssal *T*-phases excited at a sedimentary seafloor.

A. Acoustic excitation as a function of seafloor depth

The combined ocean/seafloor velocity profiles must be used to determine modal excitation at the seafloor. An acoustic velocity structure connecting an ocean velocity profile to a typical sedimentary seafloor *P*-wave velocity profile[16] is shown in Fig. 1 for a water depth of 1 km. The water column velocity profile was computed using an equation relating ocean temperatures and salinities to acoustic velocities[26] for temperature and salinity data from the Levitus database.[25] The seafloor can be treated as an acoustic medium, as *S*-wave velocities in the upper 100 m of the seafloor are very low, on the order of 125 m/s[16,17] or even less. The first three modes corresponding to the velocity profile of Fig. 1(a), computed using a WKBJ method,[27] are shown in Figs. 1(b) and 1(c) at 5 Hz and 20 Hz. As indicated, low-order modes penetrate the ocean bottom at this depth and thus can be excited by scattering. At much higher frequencies, or greater depths, low order modes have zero amplitude at the seafloor, so they cannot be excited by seafloor scattering. Thus the modal excitation due to seafloor scattering depends on both frequency and seafloor depth.

To estimate the modal excitation as a function of frequency and seafloor depth the computation indicated in Fig. 1 was repeated for a series of frequencies and seafloor depths. At each depth, standard sedimentary seafloor velocity profiles were appended to ocean velocity profiles and mode amplitudes were computed at the seafloor. The mode intensity as a function of the seafloor depth, which we term the excitation function, is shown in Fig. 2 for each of three modes and frequencies. As indicated, the excitation functions are strongly dependent on both frequency and depth. The

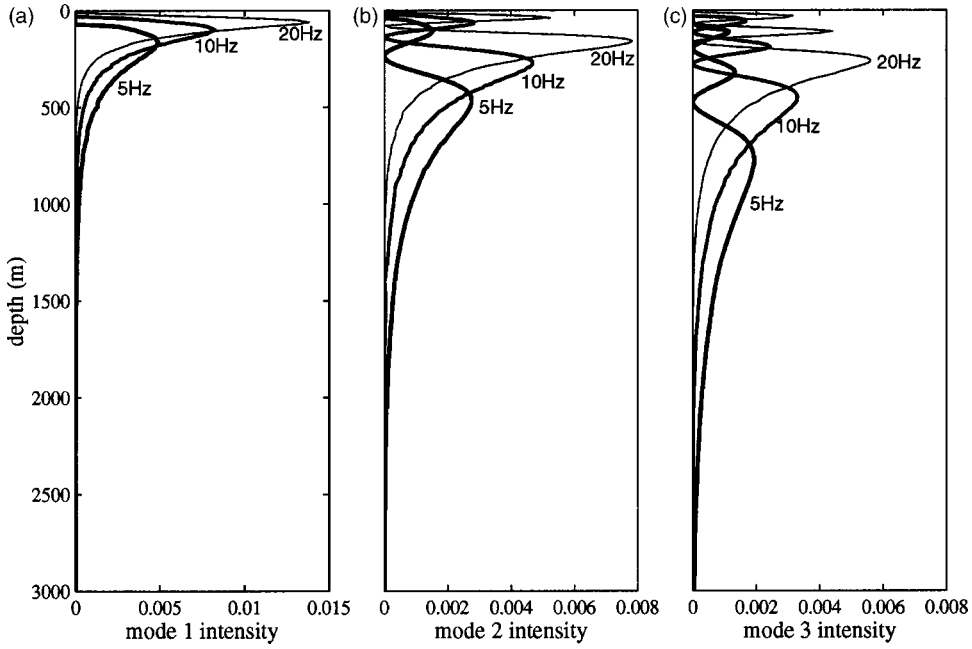


FIG. 2. Acoustic energy excitation at the seafloor, as a function of seafloor depth, computed at frequencies of 5 Hz, 10 Hz, and 20 Hz, for (a) mode 1, (b) mode 2, and (c) mode 3, for annual average velocity profiles near the earthquake epicenters.

optimal depth of excitation for any given mode increases with increasing mode number, and is approximately equal to the corresponding cutoff depth. Thus lower modes, which dominate acoustic signals at long ranges,[20] are preferentially excited in shallower regions. Furthermore, the peaks of the excitation functions decrease and become broader with increasing mode number, thus excitation of high modes is less efficient and less depth dependent than for low modes. Overall, this implies that the scattering of seismic to acoustic energy in shallow areas results in the excitation of low acoustic modes and that the strength of this coupling is strongly dependent on bathymetric depth. At a sloping interface, the acoustic energy in a given low mode ceases interaction with the bottom as it propagates into deeper water.

Modes excited by a point source on a flat seafloor at 4.5 km depth are shown in Fig. 3(a). Since a seafloor source can excite only bottom-interacting modes, lower-order modes can be generated only at the lowest frequencies. As the frequency increases, only higher-order modes are excited. The corresponding transmission coefficients, i.e., the fraction of energy that is propagated to the receiver, are shown in Fig. 3(b) for a path length of 4000 km. As shown here, the transmission coefficients increase with frequency and decrease with mode number, reflecting the fact that higher modes and

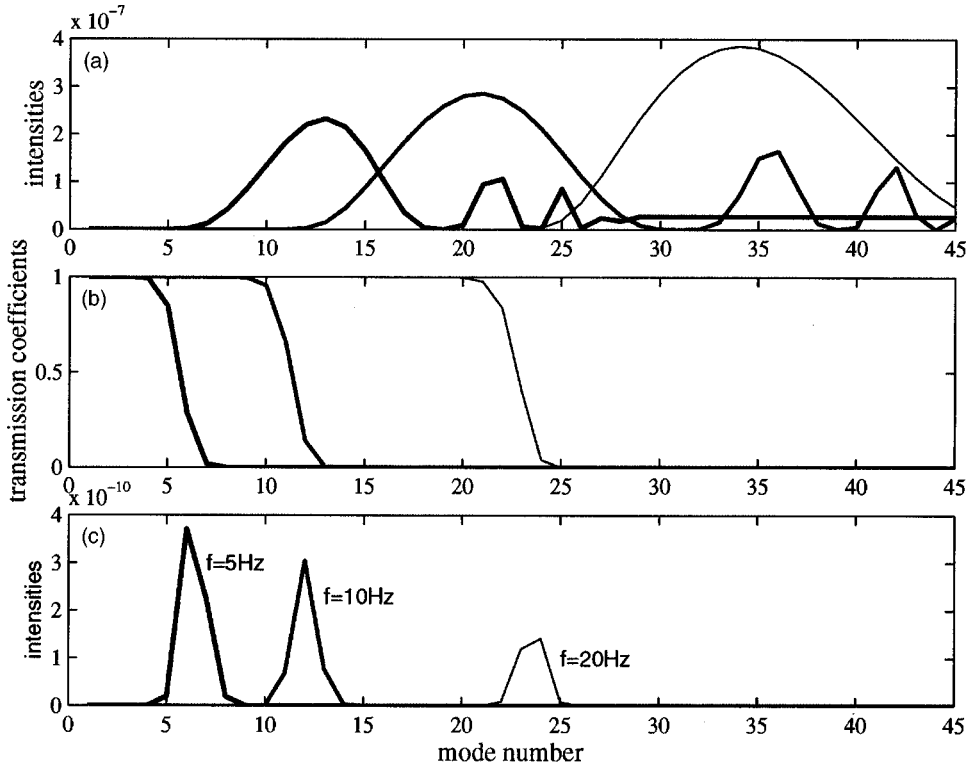


FIG. 3. (a) Modal intensities excited at a seafloor depth of 4.5 km for frequencies of 5 Hz (thickest line), 10 Hz (gray line), and 20 Hz (thin line). (b) Corresponding transmission coefficients for a path length of 4000 km, for a flat seafloor at 4.5 km depth. (c) Modal intensities observed at a receiver at a distance of 4000 km, given by the products of values in (a) and (b).

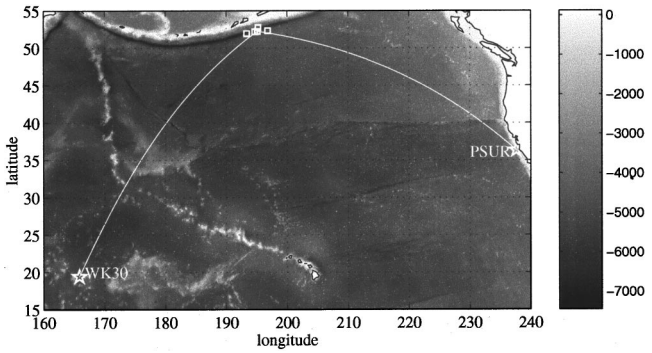


FIG. 4. Bathymetry of the north Pacific Ocean. Locations of the Pt. Sur (PSUR) and Wake Island (WK30) hydrophones are marked by stars; epicentral locations are labeled by squares. Geodesic paths from one of the epicenters to each hydrophone are indicated by solid lines.

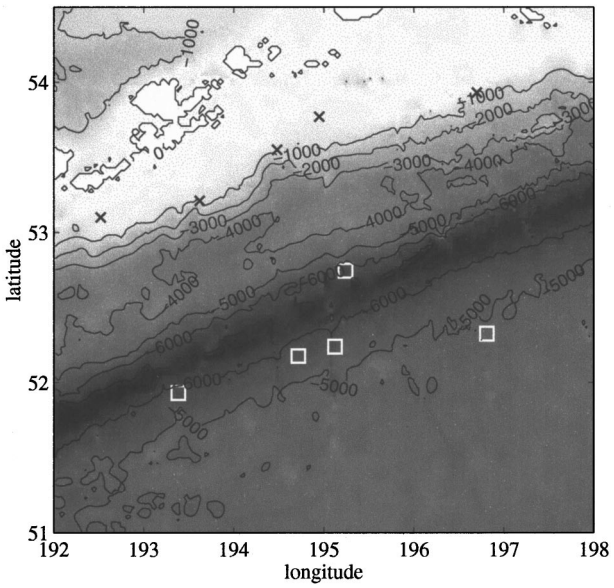


FIG. 5. Bathymetry near the Fox Islands, in the Aleutian Chain, with land indicated by white. Epicenters for earthquakes that excited both slope and abyssal *T*-phases are indicated by squares; ×'s show the source locations for the associated high amplitude *T*-phase arrivals. Bathymetries are derived from satellite altimetry data.[11]

lower frequencies penetrate more deeply into the bottom. In general, the attenuation depends strongly on the bathymetry along the transmission path; higher modes are stripped away by bathymetric promontories. However, even for an ocean floor that is perfectly flat from source to receiver, higher order modes are stripped away with increasing source–receiver distance, as they interact with the seafloor along the entire transmission path. The modal energy observed at the receiver is a product of the modal excitation and transmission coefficients and is shown in Fig. 3(c). As indicated, the lowest modes excited at the epicenter are transmitted most efficiently. These are modes that interact very weakly with the bottom, thus are only weakly excited. At these depths, excitation is nearly independent of seafloor depth.

B. Computation of travel times

Finally, acoustic travel times for each point are derived by integrating the modal group velocities along geodesic paths from each gridpoint to the receiver. The average seismic phase velocity for a cluster of events may be estimated by comparing the length of the seismic travel path to time delays between observed *T*-phase arrivals and computed acoustic travel times, as in the next section.

Several events south of the Fox Islands in the Aleutian chain excited both classes of *T*-phases which were recorded at hydrophones near Wake Island and off the coast of California. A map relating source and hydrophone locations is shown in Fig. 4. A detailed map showing bathymetry and estimated epicentral locations is shown in Fig. 5. As indicated, ocean depths at the epicenters range from 5–6 km, far below the sound channel. Spectral ratio sonograms computed for the *T*-phase recordings at WK30 and Pt. Sur are shown in Fig. 6. For the WK30 recordings, each sonogram exhibits at least two arrivals, with the high frequency abyssal phase preceeding the lower frequency, longer duration, slope phase. A notable feature of the later arrivals is that their signal-to-noise ratios are greater than those of the abyssal *T*-phase although they are generated further from the earthquake hypocenter. Also, the strength of the slope and abyssal *T*-phase arrivals near Wake Island are comparable, but at Pt. Sur the abyssal *T*-phases generally have much lower amplitudes than the slope phases, and are barely discernible for two of the events. This suggests that the abyssal *T*-phase is more susceptible to transmission blockage by bathymetric obstacles than the slope phase, consistent with our hypothesis that the abyssal phases are made up of higher order modes, which are susceptible to mode-stripping by bathymetric obstacles along the travel path.

Integration of modal group velocities indicates that modal dispersion along the source–receiver path is on the order of several seconds. This is negligible compared to the length of the *T*-phase wavetrains, and suggests that *T*-phases are generated over a distributed region of the seafloor. However, we define the *T*-phase source location as that corresponding to the peak arrival time, and estimate this for each phase. The total travel time for each phase is equal to the sum of the seismic travel time through the crust and upper mantle, and the acoustic travel time for oceanic propagation. The latter quantity is computed for a grid of points in the vicinity of the epicenter by integrating SOFAR channel velocities, derived from the Levitus database,[25] along geodesic travel paths. Since the seismic velocity, and hence the seismic travel times, are unknown, *T*-phase source regions are not uniquely identified. Therefore, given earthquake onset times and arrival times for each phase at only two hydrophones, the *T*-phase source locations are defined along arcs, which trend north–south for the given source–receiver geometry. Given additional reasonable assumptions about the seismic velocity—we assume that it must be greater than the average SOFAR velocity and less than 8 km/s—and given that the observed *T*-phases cannot be generated where acoustic transmission is blocked along the source–receiver path, *T*-phase source locations can be pinpointed to within 0.1 deg in both longitude and latitude.

Source location estimates for the slope *T*-phases are also shown for each event in Fig. 5. Abyssal *T*-phase source locations are coincident with epicentral locations to within the

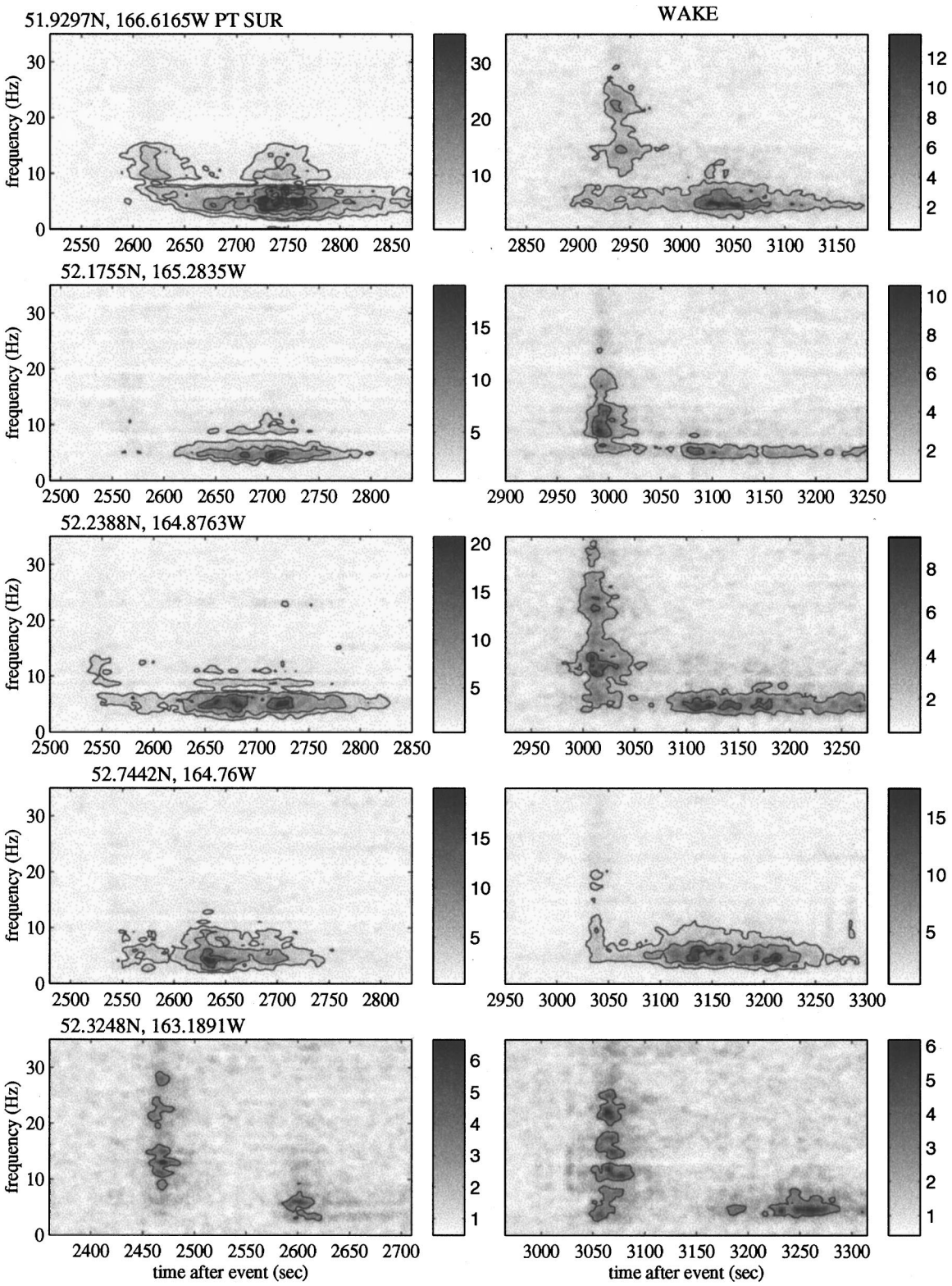


FIG. 6. Spectral ratio sonograms for each recording at the Pt. Sur (left) and Wake Island (right) hydrophones, formed by dividing the power spectrum for each time slice by average noise powers computed for quiet intervals before the *T*-phase arrivals. Events are shown from west (top) to east (bottom); epicentral locations are given for each sonogram pair. Contours indicate signal to noise ratios at intervals of 3, 6, 12, 18, 24, and 30. Depth estimates are between 30 and 40 km for all events.

accuracy of the location estimates and time picks. The lower-frequency slope *T*-phases are generated at greater distances from the epicenter, in regions of shallow, sloping bathymetry, in agreement with previous studies.[6,7,10] The approximate linearity between the seismic delay times (defined as the difference between the *T*-phase arrival time and the acoustic travel time) and the distance between the epicenters and corresponding *T*-phase source locations gives us confidence that the *T*-phase source locations are accurately estimated. These values are plotted in Fig. 7; from the slope we

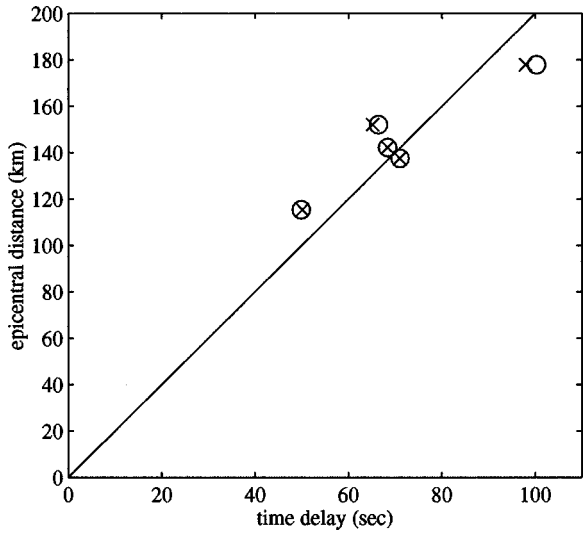


FIG. 7. Hypocentral distance of the peak slope *T*-phase source location versus the seismic travel time through the crust and upper mantle, defined as time delay between the observed *T*-phase travel time and the computed acoustic travel times for the *T*-phase source. Circles indicate the seismic travel time derived for the Wake arrival, x's indicate the corresponding value for Pt. Sur. The seismic velocity is given by the slope of the line, which is 2 km/s.

compute a velocity of 2.0 ± 0.3 km/s for the seismic phase coupling to the slope arrivals. The slow seismic velocity suggests that *T*-phases excited at large distances from the epicenter result from the conversion of shear waves to the ocean-borne acoustic phase.

The corresponding velocity of the seismic phase that couples to the abyssal *T*-phase is poorly determined, given the accuracy of the picks and of the estimated source locations. Given that only *P*-waves can excite acoustic energy within the waveguide at normal incidence, the seismic velocity associated with the abyssal phase is likely greater than 2 km/s. However, given the short distance travelled by the seismic wave prior to conversion to an abyssal phase, this value is sufficiently accurate for our computations.

III. NUMERICAL EXAMPLES

In this section, we model *T*-phase arrivals at the WK30 hydrophone for several events south of the Fox Islands. Only arrivals at WK30 are synthesized, since instrument corrections are not available for the Pt. Sur hydrophones. We assume that the source rupture duration and area may be treated as a point source in both time and location. Further-

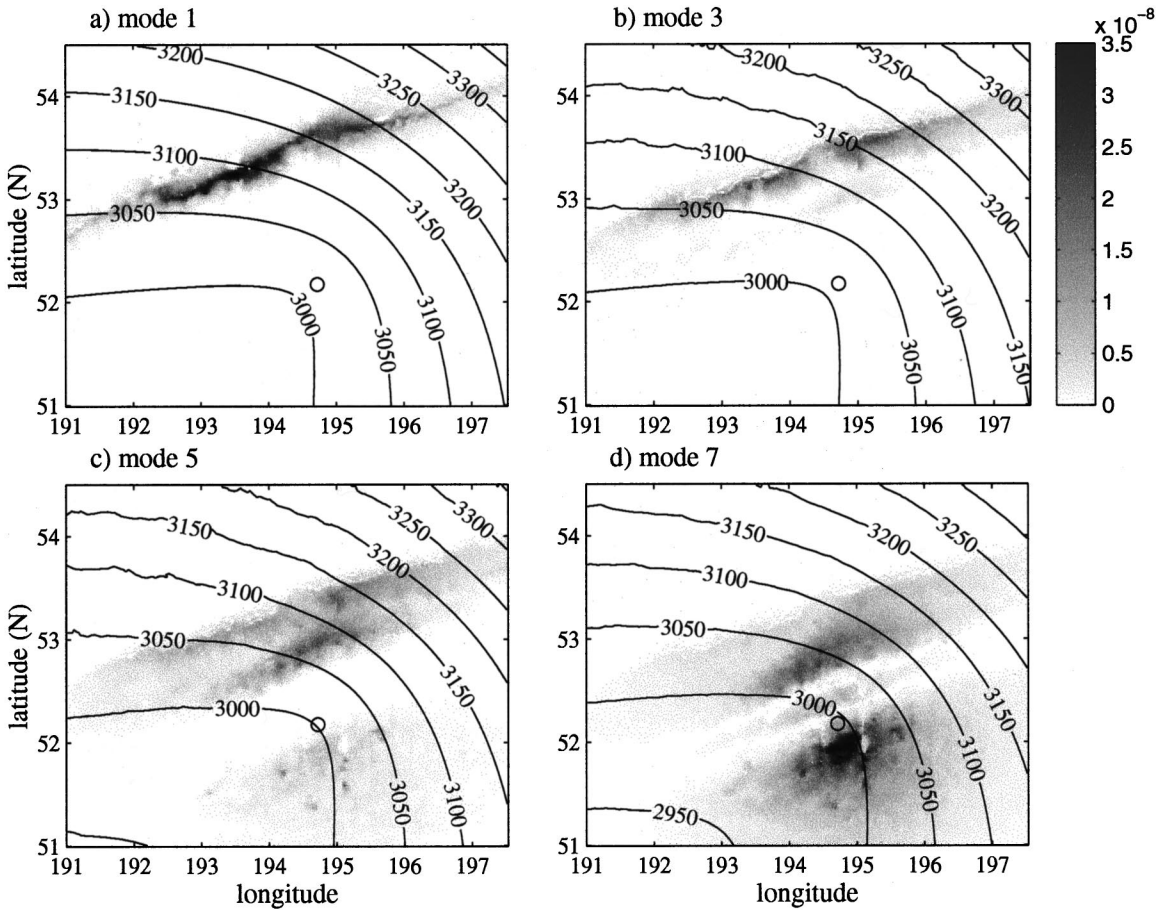


FIG. 8. Acoustic source regions for *T*-phases observed at WK30, for several modes at a frequency of 5 Hz: (a) mode 1, (b) mode 3, (c) mode 5, (d) mode 7. Dark areas indicate areas of greatest *T*-phase excitation. All plots are shown to the same colorscale. The epicenter is marked by a circle. Superimposed contour lines show the *T*-phase travel time in seconds after the earthquake.

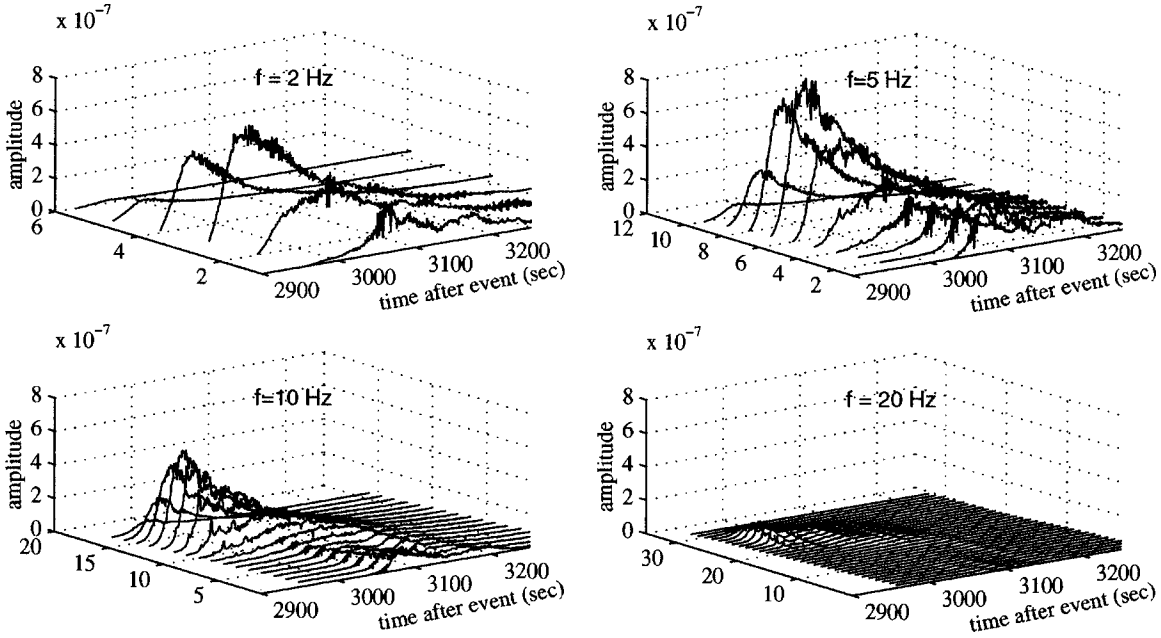


FIG. 9. Computed acoustic *T*-phase envelopes for a range of modes at several frequencies for an event at 52.1755N, 165.2835W, at a depth of 37.5 km.

more, we assume that the source radiates seismic energy uniformly across the frequency band of interest.

Synthesized source regions for *T*-phases excited by the event at 52.1755N, 165.2835W, with a depth estimate of 37.5 km are mapped in Fig. 8 for several modes, at a frequency of 5 Hz. The pressure at each point is given by

$$\frac{1}{\sqrt{|r_r - r_s|}} \Psi_m(r_s, z_s) \frac{\exp(i \int k_m(s) ds)}{\sqrt{k_m |r_r - r_s|}} \left[\frac{e^{-d_s * \alpha * f}}{d_s} \right] f^2. \tag{4}$$

As shown, *T*-phase energy in the lowest modes is excited along the shallow regions at the edge of the continental shelf. Areas to the western and northern ends of the source region are in an acoustic shadow, blocked by the Emperor seamount chain and by the Aleutian Islands, respectively. With increasing mode number, the *T*-phase excitation increases in the vicinity of the epicenter and decreases in the shallow areas. For yet higher acoustic modes, the acoustic energy observed at Wake Islands becomes negligible due to mode stripping by bathymetric obstacles along the source–receiver path. Thus, the low modes, that are excited further from the epicenter, are associated with the slope *T*-phase. Higher modes, excited nearest to the epicenter, are associated with the abyssal *T*-phase. At higher frequencies, the incident energy at the continental shelf decreases due to the frequency dependent attenuation of the seismic phase, so the slope phases are only weakly excited.

The synthesized mode envelopes for this event, formed by summing arrivals within 2-s bins, are plotted in Fig. 9 as a function of time after event rupture for several frequencies. *T*-phase energy excited in the shallow regions is concentrated in lower modes and has greater duration than the *T*-phase energy excited in the epicentral regions. The abyssal phases appear in the higher modes, and are nearly symmetric about the peak amplitudes. The separation of ''slope-generated'' *T*-phases in the low modes, and the abyssal energy in the high modes, becomes more apparent with increasing frequency. Note the decrease in amplitude of the slope *T*-phase compared to that of the abyssal *T*-phase at the higher frequencies, due to the greater attenuation of seismic energy in the crust and upper mantle.

Envelopes were formed at frequencies of 2, 5, 10, and 20 Hz by summing intensities at each mode, weighted by the mode functions at the receiver locations. Finally, the synthesized sonograms were formed by linear interpolation between the envelopes; these are shown in the left column of Fig. 10. The real, instrument-corrected sonograms are shown to the right. Sonograms are shown only for four of the events shown in Fig. 6; the center one, at 52.2388N, 164.8763W was left out due to its proximity, hence its similarity to the event at 52.1755N, 165.2835W.

Although not a perfect match, several characteristics of the synthetic sonograms agree with corresponding features of the WK30 sonograms. Both real and synthetic slope phases, which trail the abyssal phases, show a peak in energy at about 5 Hz, and have longer duration than the abyssal events. The synthetic sonograms also correctly predict that the abyssal *T*-phases have higher frequency content than the slope phase, and are nearly symmetric about the peak amplitudes. Arrival times for the abyssal phases agree, with the exception of the third event, which had a poorly constrained location. However, the computed abyssal phase has too low a dominant frequency, and falls off too rapidly with increasing frequency. Discrepancies between real and computed sonograms may be due to the simplifying assumptions made in the modeling, as discussed in the next section.

IV. DISCUSSION

A number of simplifying assumptions were made in the synthesis of the *T*-phase envelopes in order to isolate the effects of seafloor scattering from other geophysical effects.

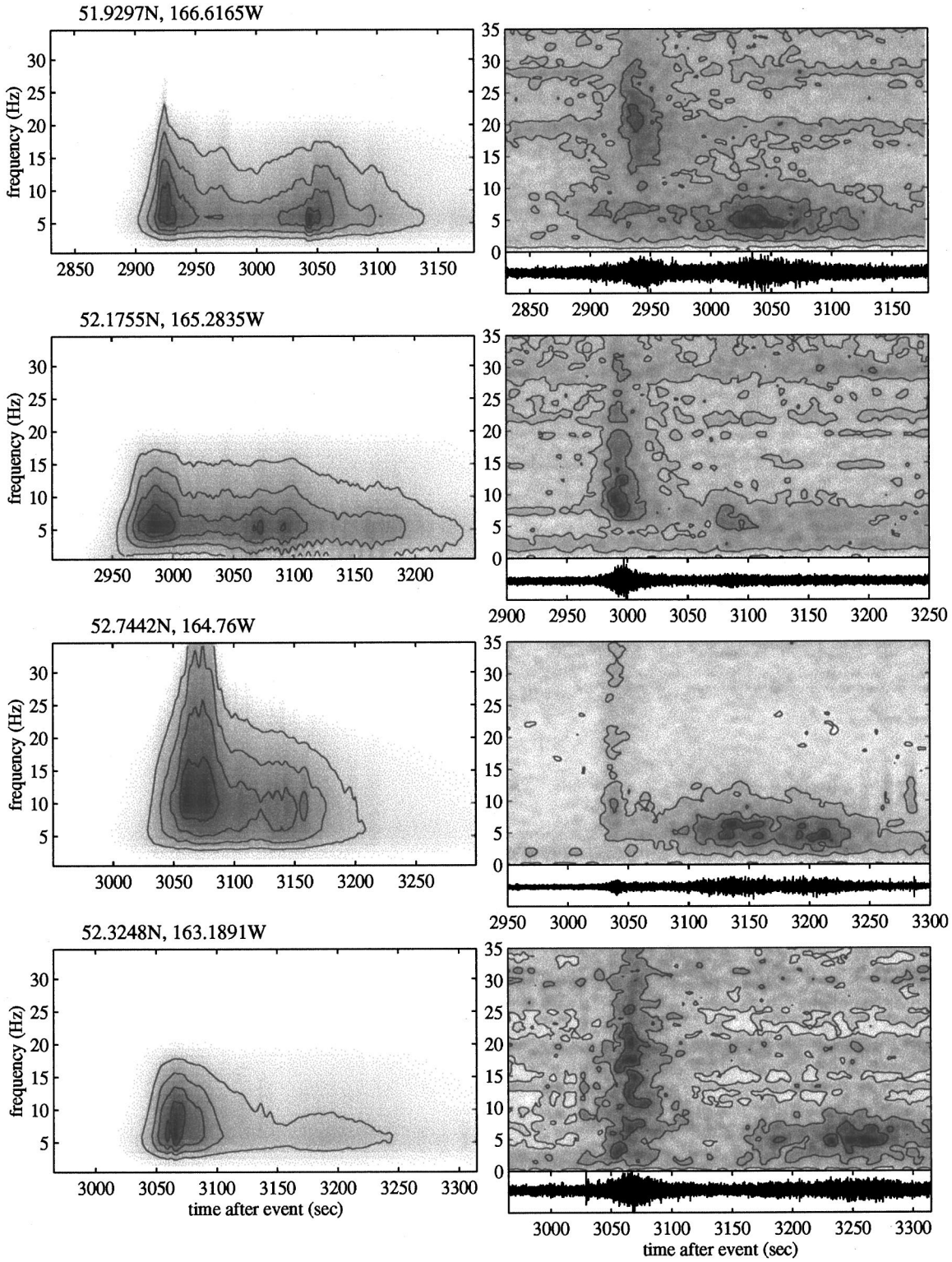


FIG. 10. Comparison of synthetic (left column) and real sonograms (right column) recorded at WK30. Recorded waveforms, bandpassed from 2–20 Hz are also shown. The epicentral locations used in computing synthetics for each event are given.

As a result, variations in amplitude with frequency and the relative scale sizes of the abyssal and slope phases are poorly constrained in this analysis. Factors which affect the frequency dependence of the observed *T*-phases are the earthquake source spectrum and the scale size of the seafloor scatterers. We have made the assumption that the characteristics of the seafloor near the Aleutian trench are similar to those of the continental shelf region. Furthermore, our assumption that the earthquake source spectrum is flat over the entire frequency range is true only if both the fault rupture surface and rupture duration are infinitesimally small.[28] For more realistic sources, the frequency spectrum drops off approximately as $1/f$ for frequencies greater than about 1 Hz, this makes the excitation of higher frequencies even more problematic.

Factors which affect both the variation in *T*-phase am-

plitude with frequency and the relative amplitudes of the slope and abyssal *T*-phases are the seismic attenuation coefficients (α and β), the source depth, and the bottom velocity profile. An increase in β, the seismic attenuation of the sedimentary seafloor, results in a decrease in the abyssal *T*-phase amplitude relative to the slope phase since the abyssal *T*-phase interacts with the seafloor along a much greater portion of its propagation path. However, increasing α, the attenuation coefficient for the upper mantle and lower crust, decreases the computed amplitudes of both the slope *T*-phase and the high-frequency abyssal *T*-phase since the loss is proportional to the number of wavelengths traveled by the seismic phase through the crust and upper mantle. Unfortunately, although the frequency dependence and relative amplitudes of the slope and abyssal phases are strongly dependent upon the value of the seismic attenuation coefficients, these values are only approximately known, so that only order of magnitude computations can be done. Furthermore, shallow sources are predicted to be associated with high amplitude abyssal phases with higher dominant frequencies than deep sources, given the shorter seismic travel path through the attenuative crust and mantle. Finally, the velocity gradient within the seabottom affects the mode functions, and hence the computed amplitudes. For a high velocity gradient, the mode functions decrease rapidly with depth so the propagation loss due to seafloor interaction decreases. The effects of decreasing the velocity gradient is thus similar to increasing the attenuation coefficient β, i.e., it decreases the computed amplitude of the abyssal *T*-phase.

An unexpected result of our analysis is that, although shear wave velocities in the sediments are so low that the seafloor may be treated as a fluid in computing the mode functions, the velocity of the seismic phase coupling to slope *T*-phases is consistent with shear waves. A mechanism to resolve this apparent incongruity is that *S*-waves convert to compressional waves at the interface between the elastic, oceanic crust and the seafloor sediments; this is supported by observations[29] that shear waves are converted at the sediment–rock interface and not at the sediment–seawater interface. Computations of *S*- and *P*-energy flux densities at the crust–sediment contact indicate that, at high angles of incidence, crustal shear waves impinging on the sediments are mainly refracted into the sediments as acoustic energy. At low angles of incidence, the shear waves are mainly reflected back into the crust. Compressional waves are mainly reflected back into the crust at high incidence angles, with a smaller portion refracted into the sediments. In this case, the abyssal phase is more likely to result from the conversion of *P*-waves and the slope phase from the transformation of *S*-waves.

V. SUMMARY

The major contribution of this article is to demonstrate that scattering of seismic to acoustic energy at a rough seafloor yields the approximate time–frequency characteristics of *T*-phases excited both at shallow regions within the SOFAR channel, as well as at depths far below the sound channel. The modeling predicts that, although *T*-phases are generated most efficiently at shallow depths, significant *T*-phase energy can be excited at depths of several kilometers. The abyssal phases consist of high-order acoustic modes that interact weakly with the sediment layer along all, or almost all, of the transmission path but can have significant amplitude for paths with few bathymetric obstacles along the transmission path. For *T*-phase excitation distant from the epicenter, shear wave energy transforms to low-order acoustic modes at bathymetric promontories. The abyssal phase results from direct conversion of *P*-waves to higher-order acoustic modes. For both the slope and abyssal phases, the scattered energy correspond to rays that initially have grazing incidence on the seafloor; as depth increases, the slope ''rays'' become entrained within the sound channel minimum. The model predictions agree with the observations[30] that sources within the SOFAR sound channel excite low mode energy, whereas sources far below the channel excite higher acoustic modes.

ACKNOWLEDGMENTS

We thank M. Collins and M. Dzieciuch for discussions on scattering and acoustic propagation. This work was supported by the Defense Threat Reduction Agency, contract DSWA01-98-0004.

[1] C. G. Fox, R. P. Dziak, H. Matsumoto, and A. E. Schreiner, ''Potential for monitoring low-level seismicity on the Juan de Fuca Ridge using military arrays,'' Mar. Technol. Soc. J. **27**, 22–30 (1994).
[2] D. K. Blackman, C. E. Nishimura, and J. A. Orcutt, ''Seismoacoustic recordings of a spreading episode on the Mohns Ridge,'' J. Geophys. Res. **105**, 10961–10973 (2000).
[3] C. G. Fox and R. P. Dziak, ''Internal deformation of the Gorda Plate observed by hydroacoustic monitoring,'' J. Geophys. Res. **104**, 17603–17615 (1999).
[4] A. E. Schreiner, C. G. Fox, and R. P. Dziak, ''Spectra and magnitudes of *T*-waves from the 1993 earthquake swarm on the Juan de Fuca Ridge,'' Geophys. Res. Lett. **22**, 139–142 (1995).
[5] J. Talandier and E. A. Okal, ''Identification criteria for sources of *T*-waves recorded in French Polynesia,'' Pure Appl. Geophys. (in press).
[6] D. H. Shurbet and M. Ewing, ''*T*-phases at Bermuda and transformation of elastic waves,'' Bull. Seismol. Soc. Am. **47**, 251–262 (1957).
[7] R. H. Johnson and R. A. Norris, ''*T*-phase radiators in the Western Aleutians,'' Bull. Seismol. Soc. Am. **58**, 1–10 (1968).
[8] D. A. Walker, C. S. McCreery, and Y. Hiyoshi, ''*T*-phase spectra, seismic moments, and tsunamigenesis,'' Bull. Seismol. Soc. Am. **82**, 1275–1305 (1992).
[9] R. H. Johnson, J. Northrup, and R. Eppley, ''Sources of Pacific *T*-Phases,'' J. Geophys. Res. **68**, 4251–4260 (1963).
[10] R. H. Johnson, R. A. Norris, and F. K. Duennebier, ''Abyssally generated *T*-phases,'' in *The Crust and Upper Mantle of the Pacific Area*, edited by L. Knopoff, C. L. Drake, and P. J. Hart, Geophysics Monograph, Vol. 12 (American Geophysics Union, Washington, D.C., 1968), pp. 70–78.
[11] R. E. Keenan and L. R. L. Merriam, ''Arctic abyssal *T* phases: Coupling seismic energy to the ocean sound channel via under-ice scattering,'' J. Acoust. Soc. Am. **89**, 1128–1133 (1991).
[12] A. R. Milne, ''Comparison of spectra of an earthquake *T*-phase with similar signals from nuclear explosions,'' Bull. Seismol. Soc. Am. **49**, 317–330 (1959).
[13] J. Talandier and E. A. Okal, ''On the mechanism of conversion of seismic waves to and from *T* waves in the vicinity of Island Shores,'' Bull. Seismol. Soc. Am. **88**, 621–632 (1998).
[14] C. D. de Groot-Hedlin and J. A. Orcutt, ''Synthesis of earthquake-generated *T*-waves,'' Geophys. Res. Lett. **26**, 1227–1230 (1999).
[15] M. A. Biot, ''The interaction of Rayleigh and Stoneley waves in the ocean bottom,'' Bull. Seismol. Soc. Am. **42**, 81–93 (1952).
[16] E. L. Hamilton, ''V_p/V_s and Poisson's ratio in marine sediments and rocks,'' J. Acoust. Soc. Am. **66**, 1093–1101 (1979).

[17] E. L. Hamilton, ''Geoacoustic modeling of the seafloor,'' J. Acoust. Soc. Am. **68**, 1313–1340 (1980).
[18] M. D. Collins, ''The adiabatic mode parabolic equation,'' J. Acoust. Soc. Am. **94**, 2269–2278 (1993).
[19] F. Ingenito, ''Measurement of mode attenuation coefficients in shallow water,'' J. Acoust. Soc. Am. **53**, 858–863 (1973).
[20] K. D. Heaney, W. A. Kuperman, and B. E. MacDonald, ''Perth–Bermuda sound propagation (1960): Adiabatic mode interpretation,'' J. Acoust. Soc. Am. **90**, 2586–2594 (1991).
[21] Lord Rayleigh [J. W. Strutt], *The Theory of Sound*, published in one volume (Dover, New York, 1945).
[22] L. M. Brekhovskikh and Yu. P. Lysanov, *Fundamentals of Ocean Acoustics*, 2nd ed. (Springer-Verlag, Berlin, 1982).
[23] U. C. Herzfeld, I. I. Kim, and J. A. Orcutt, ''Is the ocean floor a fractal?,'' Math. Geol. **27**, 421–462 (1995).
[24] W. A. Kuperman and H. Schmidt, ''Self-consistent perturbation approach to rough surface scattering in stratified elastic media,'' J. Acoust. Soc. Am. **86**, 1511–1522 (1989).
[25] S. Levitus, *World Ocean Atlas 1994* (NOAA Atlas NESDIS, Washington D.C., 1994).
[26] J. V. Mackenzie, ''Nine-term equation for sound speed in the ocean,'' J. Acoust. Soc. Am. **70**, 807–812 (1981).
[27] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (AIP, Woodbury, New York, 1994).
[28] K. Aki and P. G. Richards, *Quantitative Seismology* (W. H. Freeman and Co., San Francisco, 1980).
[29] P. Spudich and J. Orcutt, ''Petrology and porosity of an oceanic crustal site: results from wave form modeling of seismic refraction data,'' J. Geophys. Res. **85**, 1409–1433 (1980).
[30] G. L. D'Spain, L. P. Berger, and W. A. Kuperman, ''Normal mode composition of earthquake *T*-phases,'' Pure Appl. Geophys. (submitted).

Acoustic attenuation in three-component gas mixtures—Theory

Y. Dain and Richard M. Lueptow[a)]
Department of Mechanical Engineering, Northwestern University, 2145 Sheridan Road, Evanston, Illinois 60208-3111

(Received 23 March 2000; revised 29 December 2000; accepted 4 January 2001)

Vibrational relaxation accounts for absorption and dispersion of acoustic waves in gases that can be significantly greater than the classical absorption mechanisms related to shear viscosity and heat conduction. This vibrational relaxation results from retarded energy exchange between translational and intramolecular vibrational degrees of freedom. Theoretical calculation of the vibrational relaxation time of gases based on the theory of Landau and Teller [Phys. Z. Sovjetunion **10**, 34 (1936); **1**, 88 (1932); **2**, 46 (1932)] and Schwartz *et al.* [J. Chem. Phys. **20**, 1591 (1952)] has been applied at room temperature to ternary mixtures of polyatomic gases containing nitrogen, water vapor, and methane. Due to vibrational-translational and vibrational–vibrational coupling between all three components in ternary mixtures, multiple relaxation processes produce effective relaxation frequencies affecting the attenuation of sound. The dependence of effective relaxation frequencies and the attenuation on mole fractions of the constituents was investigated. The acoustic attenuation in a mixture that is primarily nitrogen is strongly dependent on the concentrations of methane and water vapor that are present. However, the attenuation in a mixture that is primarily methane is only weakly dependent on the concentrations of nitrogen and water vapor. The theory developed in this paper is applicable to other multicomponent mixtures.
[DOI: 10.1121/1.1352087]

PACS numbers: 43.35.Ae, 43.35.Fj [SGK]

I. INTRODUCTION

The anomalous absorption and dispersion of sound waves in polyatomic gases has been studied for nearly seven decades. The pioneering studies of Zener,[1] and Landau and Teller[2–4] laid the foundation of the theory of vibrational energy transfer in diatomic gases. Later, the theory of vibrational relaxation was successfully applied by Kneser,[5] Schwartz *et al.* (SSH),[6] and Tanczos[7] to explain the absorption and dispersion of sound waves in polyatomic gases. The further application of the SSH theory to the investigation of sound in pure gases and binary gas mixtures was reviewed by Herzfeld and Litovitz,[8] Burnett and North,[9] and Lambert.[10] The complexity of molecular relaxation processes can lead to a complicated mechanism of vibrational-translational and vibrational–vibrational energy transfer between different molecular energy levels producing several effective relaxation frequencies depending on the vibrational modes involved in the relaxation process. Consequently, the effective relaxation frequencies of multi-component gas mixtures depend on the concentrations of the gases in the mixture.

Several phenomenological alternatives to the SSH theory have been used for binary and multi-component gas mixtures. Bauer developed a method to obtain relaxation equations for the phenomenological theory of simultaneous multiple relaxation processes.[11] His general theory calculates the eigenvalues of coupled relaxations in terms of the individual relaxation times including the cross terms. Each energy process between two vibrational levels is considered as a reaction. Zuckerwar and Miller applied Bauer's phenomenological treatment to the case of a mixture of nitrogen, oxygen, water vapor, and carbon dioxide.[12] They provided analytical expressions for coupled effective relaxation frequencies of oxygen and nitrogen in air depending on small concentrations of water vapor and carbon dioxide. Henderson and Herzfeld applied a similar analysis to air/water vapor mixtures using the semi-empirical dependence of the effective relaxation frequency of oxygen on water vapor.[13] The general formalism of the theory of sound absorption of Bauer has been used to deduce relaxation equations with the functional form of relaxation frequencies of moist oxygen and nitrogen determined empirically from experimental measurements of sound absorption in air. A thorough review of the relaxation frequencies of nitrogen and oxygen as functions of humidity is provided by Bass *et al.*, with a focus on acoustic attenuation in the atmosphere.[14]

Townsend and Meador used the usual gas species continuity and momentum equations together with reaction kinetics to derive a wave equation for a single parameter (pressure or density) in a binary mixture.[15] The absorption coefficient was decomposed into the sum of two contributions, which cannot be uniquely identified with either of the two species. This formula was used to calculate the absorption coefficient for a nitrogen/water vapor mixture, comparing favorably with experimental data of Zuckerwar and Griffin.[16]

Experimental results for multi-component mixtures of the gases that we consider in this paper are sparse. The dependence of the relaxation frequency of nitrogen on mole fractions of carbon dioxide and water vapor was experimentally investigated at 448 K by Henderson *et al.*[17] They observed a synergistic effect of both additives resulting in a

a)Electronic mail: r-lueptow@northwestern.edu

nonlinear shift of the relaxation frequency of nitrogen with the mole fraction of the additives. Zuckerwar and Griffin experimentally studied the vibrational relaxation peak in nitrogen/water vapor binary mixtures as a function of humidity and derived the reaction rate constants assuming that vibrational–vibrational energy transfer provides the dominant relaxation path.[16]

The present study is an investigation of the relaxational component of the attenuation coefficient in a ternary mixture of polyatomic gases at room temperatures based on the SSH theory. Specifically, we consider two cases: nitrogen with small amounts of water vapor and methane, and methane with small amounts of water vapor and nitrogen. This work is unique in that it is the first application of SSH theory to multi-component mixtures to our knowledge. Previously, SSH theory has only been applied to pure gases and binary mixtures of gases.

In this paper the plane acoustic wave propagation in a gas mixture is described by the Euler equations linearized around the equilibrium density, pressure, and temperature. The total gas energy is the sum of translational energy and internal molecular vibrational energies,[6] depending on kinetic processes between vibrational modes of the constituents. The relaxation equations, which are written following Landau and Teller,[2] Schwartz *et al.*, [6] and Tanczos[7] establish a link between the internal molecular vibrational ''temperatures'' and the translational temperature of the gas. In the case of acoustic attenuation, the deviations from equilibrium temperature are small, and the relaxation system is a linear system of ordinary differential equations. Coefficients of the matrix of the relaxation system can be expressed in terms of transition probabilities. For the calculation of the transition probabilities, the SSH theory, which agrees relatively well with experimental data for nitrogen and methane at room temperature,[10] is used.

II. THEORY FOR RELAXATION IN MULTI-COMPONENT MIXTURES

The theory is based on the Euler gas equations as the model of a continuous medium for a polyatomic gas mixture accompanied by nonlinear semi-macroscopic population equations for the number of molecules in a given energy state.[18] In the acoustic approximation, these equations yield acoustical equations and a linear system of ordinary differential equations describing multiple relaxation processes. Consequently, we do not include the classical effects of viscosity and heat transfer on the acoustic attenuation. These effects can be calculated separately. Assuming that there is no diffusion of gas components, the acoustic equations for a gas mixture including the internal temperatures of molecular vibrational modes can be written in the form

$$\frac{p}{p_0}=\frac{T}{T_0}+\frac{\rho}{\rho_0},\quad \frac{\partial\rho}{\partial t}+\rho_0\frac{\partial u}{\partial x}=0,\quad \frac{\partial u}{\partial t}+\rho_0^{-1}\frac{\partial p}{\partial x}=0,$$

$$\frac{\partial\epsilon}{\partial t}-p_0\rho_0^{-2}\frac{\partial\rho}{\partial t}=0, \tag{1}$$

$$\epsilon=c_v T+\alpha_1 c_1^{\mathrm{vib}}T_1^{\mathrm{vib}}+\alpha_2 c_2^{\mathrm{vib}}T_2^{\mathrm{vib}}+\alpha_3(c_3^{\mathrm{vib}}T_3^{\mathrm{vib}}+c_4^{\mathrm{vib}}T_4^{\mathrm{vib}}),$$

$$\sum_{i=1}^{3}\alpha_i=1,\quad c_v=\sum_{i=1}^{3}\alpha_i c_v^i,$$

where p, ρ, u, ϵ, T, and T_j^{vib} are small fluctuations of pressure, density, velocity, energy, and temperatures around the equilibrium value. Here α_i are mole fractions of the three gas constituents, and c_v^i are the translational specific heat capacities of mixture constituents, i. For molecular vibrational mode j, c_j^{vib} are the vibrational specific heat capacities, and T_j^{vib} are the internal temperatures.

The fifth equation of Eq. (1) represents the fluctuation of the total energy ϵ of the gas mixture. The total energy depends not only on the gas temperature T, but also on the internal temperatures of excited molecule mode T_j^{vib}. We have assumed that there are only three gas components. The first two gases having mole fractions α_1 and α_2 are assumed to each have one vibrational mode with energies $c_1^{\mathrm{vib}}T_1^{\mathrm{vib}}$ and $c_2^{\mathrm{vib}}T_2^{\mathrm{vib}}$, respectively. We consider two vibrational modes with energies $c_3^{\mathrm{vib}}T_3^{\mathrm{vib}}$ and $c_4^{\mathrm{vib}}T_4^{\mathrm{vib}}$ for the third gas having mole fraction α_3. As will be discussed later, nitrogen and water vapor are modeled with a single vibrational mode, while methane is modeled with two vibrational modes. Clearly the fifth equation of Eq. (1) could be extended to accommodate any number of constituent gases, each with any number of vibrational modes.

We assume that the translational degrees of freedom of each of the gases in the mixture are in equilibrium at equilibrium temperature T_0. The mean equilibrium pressure and density are p_0 and ρ_0, where

$$p_0=\rho_0 R T_0,\quad R=\frac{\mathsf{R}}{M}. \tag{2}$$

Here R is the gas constant per unit mass, R is the universal gas constant, and M is the mean molecular weight of the mixture. By definition, the total fluctuation of gas density of the mixture is

$$\rho=\sum_{i=1}^{3}\alpha_i\rho_i. \tag{3}$$

Molecules gain and lose vibrational and translational energy in collisions. The kinetic nature of the collision process is thus of fundamental importance for investigation of relaxation processes and acoustic absorption. The relaxation equations for internal molecular temperatures depend to a large measure on the model of energy transitions and kinetic processes including: (1) vibrational-translational (V-T) energy exchange between vibrational and translational degrees of freedom for one type of molecule; (2) vibrational-translational (V-T) energy exchange for vibrational and translational degrees of freedom between different types of molecules; (3) vibrational–vibrational (V–V) energy exchange within a given type of molecule; (4) vibrational–vibrational (V–V) energy exchange between vibrational modes of different types of molecules. Thus the collisional reactions of excitation and deexcitation are of the form

$$M+M\Leftrightarrow M+M^*,$$

$$M^*(p+1,q)+M(p,q)\Leftrightarrow M(p,q)+M^*(p,q+1),$$

TABLE I. Vibrational frequencies and collisional parameters for Lennard-Jones potential (Refs. 10, 19).

	r_0 (Å)	ϵ_{LJ} (cal mol^{-1})	Normal modes of vibration (cm^{-1})	g	Vibrational amplitude coefficients (a.m.u.$^{-1}$)
N_2	3.749	159	$\nu=2331$	1	0.0354
H_2O	2.65	760	$\nu_1=3657$	1	0.9539
			$\nu_2=1596$	1	0.9527
			$\nu_3=3756$	1	0.9241
CH_4	3.796	286	$\nu_1=2915$	1	0.9921
			$\nu_2=1534$	2	0.9921
			$\nu_3=3019$	1	0.9923
			$\nu_4=1306$	3	0.8368

$$M+N \Leftrightarrow M+N^*,$$

$$M^*(p+1)+N(q) \Leftrightarrow M(p)+N^*(q+1).$$

Here M and N denote species of the mixture, and p and q are molecular vibrational modes. An asterisk indicates a molecule excited to the lowest vibrational level above the ground level.

Table I provides the vibrational modes ν_i of nitrogen, water, and methane. The subscripts are numbered according to the spectroscopic convention.[19] The frequencies are expressed using the spectroscopic convention in terms of inverse wavelength. Multiplying the numerical value by the speed of light (3.00×10^{10} cm/s) provides the value for the mode in units of frequency (Hz).

In this analysis we consider only relatively low temperatures. We assume that only the lowest modes are significant, since the contribution of higher modes to molecular energy transfer at room temperature is small. Consequently, energy exchange occurs only between the following vibrational modes: $\nu=2331\,\mathrm{cm}^{-1}$ of N_2, $\nu_2=1596\,\mathrm{cm}^{-1}$ of H_2O, and $\nu_2=1534\,\mathrm{cm}^{-1}$ and $\nu_4=1306\,\mathrm{cm}^{-1}$ of CH_4. The second vibrational mode ν_2 of CH_4 is included along with the lowest modes of other constituents due to its near resonance with the mode $\nu_2=1596\,\mathrm{cm}^{-1}$ of H_2O. For notational simplicity, n_i denotes the vibrational modes when multiplied by the speed of light. Thus n_1 corresponds to the vibrational mode of N_2, n_2 corresponds to the lowest vibrational mode of H_2O, and n_3 and n_4 are the two lowest vibrational modes of CH_4.

The vibrational specific heats for the vibrational modes in Eq. (1) are given by the Planck–Einstein function[20] for a harmonic oscillator

$$c_i^{\mathrm{vib}}=g_i\mathsf{R}\left(\frac{\theta_i^{\mathrm{vib}}}{T_0}\right)^2\frac{\exp(\theta_i^{\mathrm{vib}}/T_0)}{(\exp(\theta_i^{\mathrm{vib}}/T_0)-1)^2},\quad \theta_i^{\mathrm{vib}}=\frac{hn_i}{k}, \tag{4}$$

where $h=6.626\times10^{-34}$ J s is Planck's constant, $k=1.380\times10^{-23}$ J/K is Boltzmann's constant, θ_i^{vib} is the characteristic temperature for vibration, and g_i is the degeneracy of the mode (listed in Table I), which is related to the geometric structure of the molecule.

We assume that at room temperature only one-quantum collisional reactions are possible. By analogy to Landau and Teller[2] and Schwartz *et al.*[6] we can deduce relaxation equations for the internal energies of the four vibrational modes n_j participating in the relaxation process of the form

$$\frac{dT_j^{\mathrm{vib}}}{dt}=\frac{T-T_j^{\mathrm{vib}}}{\tau_j^{\mathrm{tran}}}+\sum_{\substack{k=1\\k\neq j}}^{4}\frac{1}{\tau_{j,k}^{\mathrm{vib}}}\frac{1-\exp(-hn_j/kT_0)}{1-\exp(-hn_k/kT_0)}$$
$$\times\left[(T-T_j^{\mathrm{vib}})-\frac{n_k}{n_j}(T-T_k^{\mathrm{vib}})\right],\quad j=1,...,4. \tag{5}$$

The translational (V-T) relaxation times that appear in Eq. (5) are

$$\frac{1}{\tau_j^{\mathrm{tran}}}=\sum_{i=1}^{3}\frac{\alpha_i}{\tau_{j,i}^{\mathrm{tran}}},\quad \frac{1}{\tau_4^{\mathrm{tran}}}=\sum_{i=1}^{2}\frac{\alpha_i}{\tau_{4,i}^{\mathrm{tran}}}+\frac{\alpha_3}{\tau_{4,4}^{\mathrm{tran}}},\quad j=1,...,3; \tag{6}$$

$$\tau_{j,i}^{\mathrm{trans}^{-1}}=Z(j,i)P_{0\to0}^{1\to0}(j,i)(1-\exp(-hn_j/kT)). \tag{7}$$

The paired vibrational (V–V) relaxation times for excitation processes with two vibrational modes involved are

$$\tau_{j,k}^{\mathrm{vib}^{-1}}=\alpha_k g_k Z(j,k)P_{0\to1}^{1\to0}(j,k),$$
$$j,k=1,...,4,\ j\neq k,\quad \alpha_3=\alpha_4. \tag{8}$$

Here $Z(j,k)$ are collision rates of molecules of specie j with molecules of specie k. $P_{0\to0}^{1\to0}(j,k)$ and $P_{0\to1}^{1\to0}(j,k)$ are transitional probabilities of V-T and V–V exchange of vibrational energy per collision between different vibrational modes.

The calculation of the number of collisions per molecule is conveniently based on the kinetic theory expression for a gas of rigid spheres[10]

$$Z(j,k)=2N_k\left(\frac{\sigma_j+\sigma_k}{2}\right)^2(2\pi kT(m_j+m_k)/m_jm_k)^{1/2}, \tag{9}$$

where N_k is the number of molecules of species k per unit volume, σ_j,σ_k are collision diameters, and m_j,m_k are the molecular masses of species.

The calculation of transition probabilities $P_{0\to0}^{1\to0}(j,k)$ and $P_{0\to1}^{1\to0}(j,k)$ is a more complicated problem. It is based on a solution of the Schrödinger equation for binary collisions of two molecules in a mixture. The solution depends on assuming special collisional conditions and is based on the molecular properties of the species. There are various approximations and approaches to estimate such a solution. Here we will use the approximate formulas of transition probabilities for polyatomic gases derived by Tanzcos[7] [Eq. (1)] for the evaluation of V-T and V–V transition probabilities $P_{0\to0}^{1\to0}(j,k)$ and $P_{0\to1}^{1\to0}(j,k)$. The general Tanczos equation for the transition probabilities is quite complicated and not included here. Suffice it to say that the equation depends on geometrical factors, collision cross-section factors, vibrational factors, the total change in translational energy, the Lennard-Jones potential, and an intermolecular force constant. Methods for determining these items are outlined by Tanzcos.[7] Values for constants for the three gases of interest are listed in Table I, based on Tables 1.1, 3.1 and 3.2 in the book by Lambert.[10] The values for $P_{0\to1}^{1\to0}(j,k)$ are given in Table II.

The relaxation system Eq. (6) can be written in the vector form as

$$\frac{dT^{\mathrm{vib}}}{dt}=-AT^{\mathrm{vib}}+qT, \tag{10}$$

TABLE II. $P_{0\to1}^{1\to0}(j,k)$, where j is the row number, k is the column number.

not used	$1.43e-005$	$1.54e-006$	$1.12e-005$
$3.67e-007$	not used	0.12	0.41
$6.27e-009$	0.02	not used	0.05
$1.48e-007$	0.21	0.16	not used

where matrix $A=A(\alpha_1,\alpha_2,\alpha_3)$ of the system and the vector $q=q(\alpha_1,\alpha_2,\alpha_3)$ depend on the mole fractions α_i of gas constituents according to:

$$A_{jj}=\frac{1}{\tau_j^{\text{tran}}}+\sum_{\substack{k=1\\k\neq j}}^{4}\frac{1}{\tau_{j,k}^{\text{vib}}}\frac{1-\exp(-hn_j/kT_0)}{1-\exp(-hn_k/kT_0)},$$

$$A_{jk}=-\frac{1}{\tau_{j,k}^{\text{vib}}}\frac{1-\exp(-hn_j/kT_0)}{1-\exp(-hn_k/kT_0)}\frac{n_k}{n_j}, \tag{11}$$

$$q_j=\frac{1}{\tau_j^{\text{tran}}}+\sum_{\substack{k=1\\k\neq j}}^{4}\frac{1}{\tau_{j,k}^{\text{vib}}}\frac{1-\exp(-hn_j/kT_0)}{1-\exp(-hn_k/kT_0)}\left[1-\frac{n_k}{n_j}\right],$$

$$j,k=1,...,4, j\neq k.$$

To summarize the calculation procedure for molecular relaxation: The collision rates $Z(j,k)$ and transition probabilities $P_{0\to0}^{1\to0}(j,k)$ and $P_{0\to1}^{1\to0}(j,k)$ at a given temperature and pressure are calculated based on Eq. (9) and the SSH–Tanzcos theory. Then the relaxation times $\tau_{j,k}^{\text{vib}}$, and τ_j^{tran} are calculated based on the constituent gas concentrations as well as the collision rates and transition probabilities using Eqs. (7) and (8). Substitution of these values into Eqs. (10) and (11) provides a system of first order differential equations that can be solved numerically for T_j^{vib}. Then these values can be substituted into the fifth equation of Eq. (1) for the total energy of the gas mixture. Finally Eqs. (1) are solved as described below.

III. SOLUTION PROCEDURE

The calculation of the attenuation of the acoustic wave in a gas mixture requires the solution of the united system of differential equations (1) and (10), is sought in the form of a harmonic plane wave with all quantities p, ρ, u, ϵ, T, and T_j^{vib} expressed in the plane wave form

$$f=\overline{f}e^{-i(\omega t-kx)}, \tag{12}$$

where $\overline{f}$ is the amplitude of the quantity. The subsitution of Eq. (12) into the system of differential equations leads to a homogeneous system of algebraic equations in terms of the amplitudes

$$\frac{\overline{p}}{p_0}=\frac{\overline{T}}{T_0}+\frac{\overline{\rho}}{\rho_0},\quad -\omega\overline{\rho}+\rho_0 k\overline{u}=0,\quad \omega\overline{u}-\rho_0^{-1}k\overline{p}=0,$$
$$\overline{\epsilon}-p_0\rho_0^{-2}\overline{\rho}=0, \tag{13}$$

where

$$\overline{\epsilon}=c_v\overline{T}+\sum\ \alpha_i c_i^{\text{vib}}\overline{T}_i^{\text{vib}}. \tag{14}$$

Likewise, the substitution of Eq. (12) into the system of differential equations (10) leads to the algebraic equation

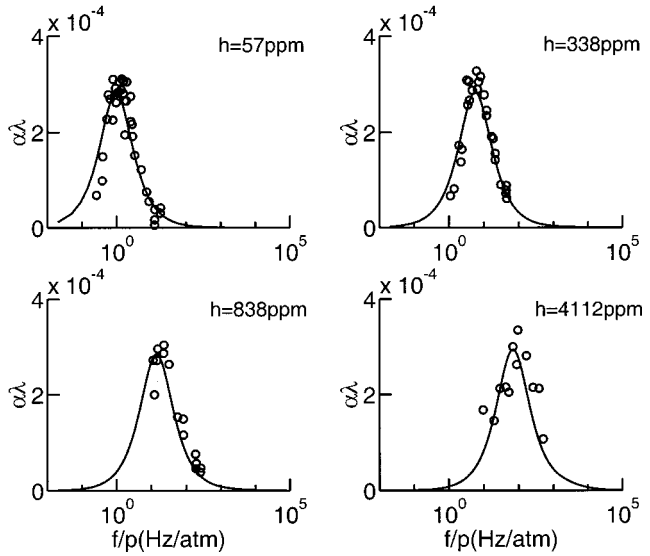


FIG. 1. Comparison of theoretical sound absorption curves with experimental data in N_2–H_2O mixtures for different H_2O concentrations (h). Symbols: Experimental data from Zuckerwar and Griffin (Ref. 16). Curves: Theoretically predicted attenuation curves ($T=297$ K, $P=1$ atm).

$$(i\omega I+A)\overline{T}^{\text{vib}}=q\overline{T}. \tag{15}$$

This equation relates the amplitudes of the internal molecular temperatures $\overline{T}^{\text{vib}}$ to the amplitude of the gas temperature $\overline{T}$, where the rows of the matrix A correspond to the vibrational modes under consideration, the columns correspond to the vibrational reactions, and I is the identity matrix.

Equations (13), (14), and (15) can be combined in matrix form as

$$\mathbf{B}y=0,\ y=(\overline{\rho},\overline{u},\overline{p},\overline{T},\overline{T}_1^{\text{vib}},\overline{T}_2^{\text{vib}},\overline{T}_3^{\text{vib}},\overline{T}_4^{\text{vib}}), \tag{16}$$

where $\mathbf{B}$ is an 8×8 matrix of coefficients. These coefficients depend on the gas concentration α_i, the frequency ω, and the coefficients of the matrix A calculated for the vibrational modes n_i under consideration. Setting the determinant of ma-

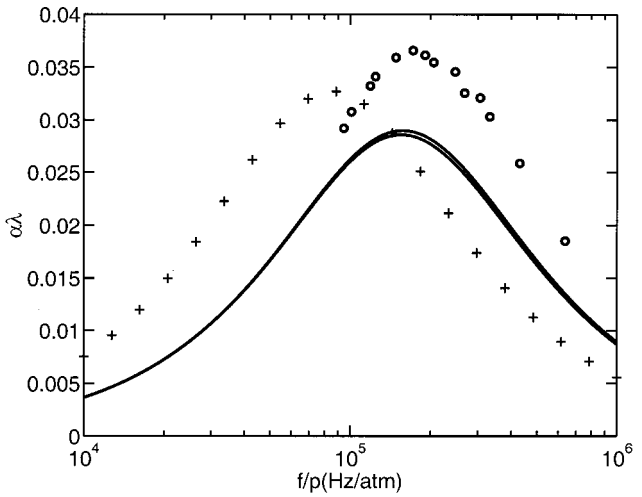


FIG. 2. Comparison of theoretical relaxation attenuation with experimental data for CH_4. Symbols: Experimental data from Edmonds and Lamb (Ref. 22) ($T=298.15$ K, $P=1$ atm, curve fit through data, +), and Gravitt *et al.* (Ref. 24) ($T=299.15$ K, $P=1$ atm, circles). Curves: Theoretically predicted attenuation (lower curve 298.15 K, upper curve 299.15 K).

trix **B** equal to zero provides a dispersion relation, which can be solved numerically to provide k as a function of ω. The wave number k is in general a complex value: $k=k_R+i\alpha$. The real part is proportional to the reciprocal of the wavelength, $k_R=2\pi/\lambda$, and determines the actual speed of wave propagation, $a=\omega/k_R$. The dimensionless attenuation coefficient per wavelength is $\alpha\lambda$.

To test our procedure, we compared our calculation of sound absorption in a binary N_2-H_2O mixture with experimental results. In this case, the coefficient matrix is much simpler, since $\alpha_3=0$. Using the procedure described above, the attenuation can be calculated as a function of frequency. The results of the calculation are compared to the experimental data of Zuckerwar and Griffin[16] in Fig. 1 for four different concentrations of water vapor. Although the theory slightly under-predicts the relaxation frequency at which the maximum attenuation occurs, the overall fit of the model to the experimental results is quite good.

The vibrational relaxation of methane has been measured based upon reverberation experiments, acoustic resonance tube measurements, and the ultrasonic velocity method.[21–25] The experimental results at room temperatures (shown in Table III) indicate a relaxation frequency either near 1.5×10^5 Hz, which agrees quite well with our calculations, or slightly less than 1.0×10^5 Hz. In two cases,[22,24] it is possible to compare results over a range of frequencies, as shown in Fig. 2. The amplitude of the calculated attenuation is somewhat less than the experimental value, but clearly in the proper range. The relaxation frequency matches one experiment quite well and is only slightly different from the other. Given the variability in the experimental results evident in Fig. 2 and Table III, it is clear that the calculations agree quite well with the experiments.

TABLE III. Comparison of theoretical calculations with experimental data for CH_4 relaxation.

Temperature (K)	Calculated relaxation frequency (10^5 Hz)	Experimental relaxation frequency (10^5 Hz)	Reference
382	3.10	1.89	Eucken and Aybar (Ref. 21)
298.15	1.46	0.85	Edmonds and Lamb (Ref. 22)
296	1.43	1.48	Parker and Swope (Ref. 23)
299.15	1.48	1.7	Gravitt *et al.* (Ref. 24)
303	1.53	0.94	Cottrel and Day (Ref. 25), 0.3%N_2

IV. RESULTS

We first consider a $N_2-H_2O-CH_4$ mixture, when water vapor and methane are small additives with volume concentrations of 0.03. Evaluation of matrix A at a temperature of 297 K and a pressure of 1 atm produces the relaxation matrix

$$A=\begin{pmatrix} 1.17e+004 & -2.95e+003 & -715 & -4.05e+003 \\ -4.90e+003 & 4.28e+008 & -1.05e+008 & -2.88e+008 \\ -96.05 & -9.32e+006 & 4.31e+007 & -4.15e+007 \\ -1.94e+003 & -8.36e+007 & -1.36e+008 & 2.40e+008 \end{pmatrix}. \tag{17}$$

Eigenvalues λ_i of matrix (17) define effective relaxation frequencies $f_i=\lambda_i/2\pi$ of the ternary mixture which are equal to

$$f_1=1860\ \text{Hz}, \quad f_2=2.19\times10^4\ \text{Hz}, \quad f_3=3.12\times10^7\ \text{Hz}, \quad f_4=8.21\times10^7\ \text{Hz}. \tag{18}$$

The corresponding normalized eigenvectors V_i of matrix A for eigenvalues λ_i are presented in Table IV. Elements of the eigenvectors describe the degree of participation of a vibrational mode in the relaxation of the complex. The first eigenvector V_1 in Table IV indicates that the nitrogen contribution to the relaxing complex is unity. Therefore, the effective relaxation frequency f_1 can be considered as the effective relaxation frequency of nitrogen as modified by the two small additives of water and methane. Eigenvector V_2 corresponding to the effective relaxation frequency f_2 has detectable but small nitrogen contribution. The effective relaxing complex comprises water and both modes of methane. The nitrogen contribution in relaxing complexes for eigenvectors V_3 and V_4 corresponding to effective relaxation frequencies f_3 and f_4 is negligible. The effective complex for f_3 is water and both modes of methane, whereas the relaxing complex for f_4 is water and mode n_4 of methane.

The results in Eqs. (17), (18), and Table IV are for a mixture of 3% water vapor, 3% methane, and 94% nitrogen. Of course similar results can be obtained for other concentrations of constituents. The constituent concentration alters the translational and vibrational relaxation times [Eqs. (6), (7), and (8)] thereby altering the relaxation matrix A. Thus the relaxation frequencies and eigenvectors can be found as functions of concentrations of water vapor and methane. Figure 3(a) shows the dependence of the lowest relaxation frequency f_1 on the concentration of methane and water vapor. The relaxation frequency is linearly dependent on the concentration of both methane and water vapor, except at very low methane concentrations. Likewise, the second lowest re-

TABLE IV. Normalized eigenvectors V_i of matrix A for 94%N_2–3%H_2O–3%CH_4.

	V_1	V_2	V_3	V_4
$N_2(\nu_1)$	1.000	−0.034	−0.000	0.000
$H_2O(\nu_2)$	0.005	0.529	0.717	−0.956
$CH_4(\nu_3)$	0.006	0.645	−0.223	−0.007
$CH_4(\nu_4)$	0.005	0.550	0.661	0.293

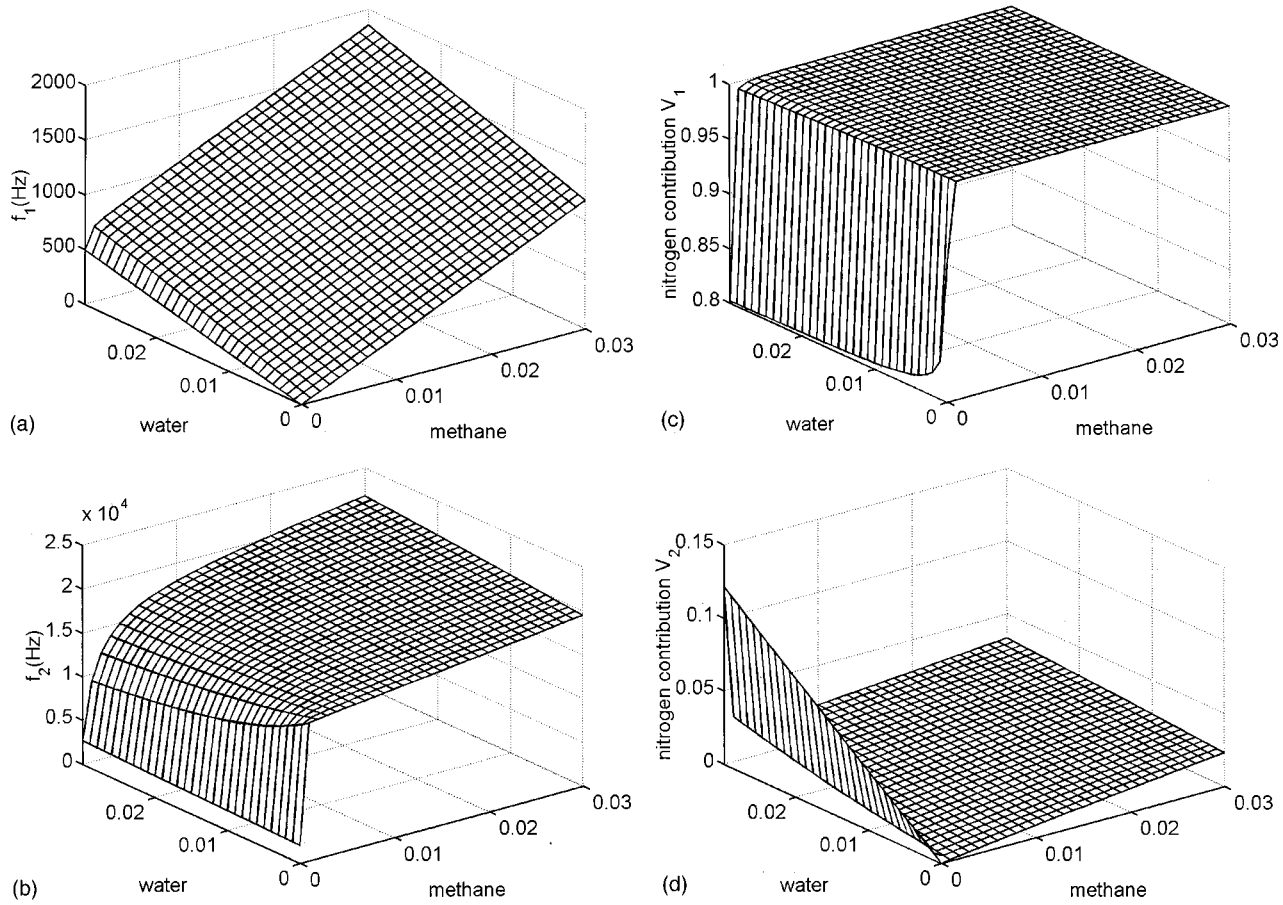


FIG. 3. Dependence of effective relaxation on the concentration in a ternary N_2–H_2O–CH_4 mixture ($T=297$ K, $P=1$ atm). (a) Relaxation frequency f_1; (b) Relaxation frequency f_2; (c) Nitrogen contribution term of the eigenvector for f_1; (d) Nitrogen contribution term of the eigenvector for f_2.

laxation frequency f_2 is also linearly dependent on concentration except at low methane concentrations, as shown in Fig. 3(b). The reason for the nonlinearity at low concentrations of methane is evident from the nitrogen contribution to the relaxation as indicated in the first term of the eigenvector. The value of this term is plotted as a function of concentration in Fig. 3(c). This term has a value near unity except at very low methane concentrations where its value drops precipitously. The first term of the eigenvector related to the contribution of nitrogen is much smaller for the second lowest frequency f_2. The plot of its value as a function of concentration in Fig. 3(d) shows a sharp change in its value at low methane concentration. The concentration for which the value changes sharply corresponds to the concentration for which the second relaxation frequency varies nonlinearly with methane concentration.

The third and fourth relaxation frequencies are linearly dependent on the water vapor and methane concentration as shown in Fig. 4. In both cases, the contribution of the nitrogen to the relaxation is so small that it does not result in any nonlinearity.

While the variation of the relaxation frequencies with concentration is of interest with respect to the physical phenomenon of vibrational relaxation, the key physical result is the attenuation of sound. The attenuation of sound is found by solving for the dispersion relation of Eq. (16). For a specified concentration of constituent gases in the mixture, the dispersion relation provides the imaginary part of the wave number α, which is the sound absorption, as a function of frequency ω. Figure 5 shows the relaxational and classical attenuation curves of a $94\%N_2-3\%H_2O-3\%CH_4$ mixture of gases as a function of frequency. For comparison, the relaxation attenuation of a $97\%N_2-3\%H_2O$ and a $97\%N_2-3\%CH_4$ binary gas mixtures are also shown in the figure. Only the peak corresponding to effective relaxation frequency f_2 of the ternary mixture (bold curve) is evident. Effective relaxation frequencies f_3 and f_4 are at much higher frequencies and the absorption related to these frequencies is quite small. There is a slight inflection in the $N_2-H_2O-CH_4$ curve near $f/p\cong 10^3$ Hz/atm that corresponds approximately to the peak in the N_2-H_2O curve. The major peak at relaxation frequency f_2 in the $N_2-H_2O-CH_4$ curve corresponds to the major peak in the N_2-CH_4 curve, although a slight inflection is evident in the N_2-CH_4 curve at a lower frequency, probably corresponding to the interaction with nitrogen.

For comparison, the classical attenuation due to viscosity and heat conduction was calculated for the gas mixture using the classical formula.[8] The viscosity and thermal conductivity used in calculations were based on physical properties software.[26] The classical attenuation ranges from 4.5×10^{-7} to 4.5×10^{-3} for $100\,\text{Hz/atm}<f/p<10^6\,\text{Hz/atm}$. Thus it is clear that the relaxational attenuation is dominant in the range of effective relaxation frequencies f_1 and f_2.

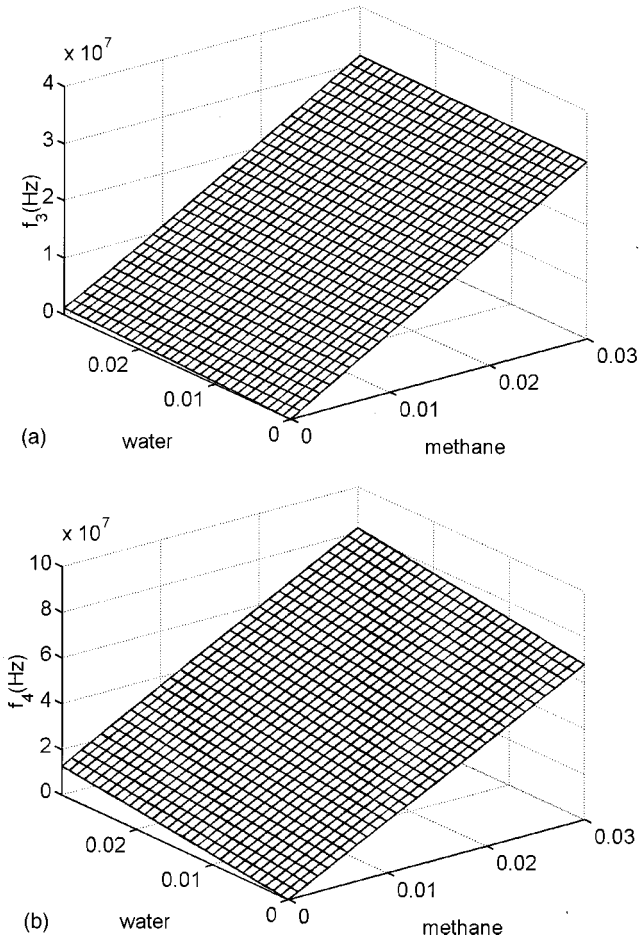


FIG. 4. Dependence of effective relaxation on the concentration in a ternary N_2–H_2O–CH_4 mixture ($T=297$ K, $P=1$ atm). (a) Relaxation frequency f_3; (b) Relaxation frequency f_4.

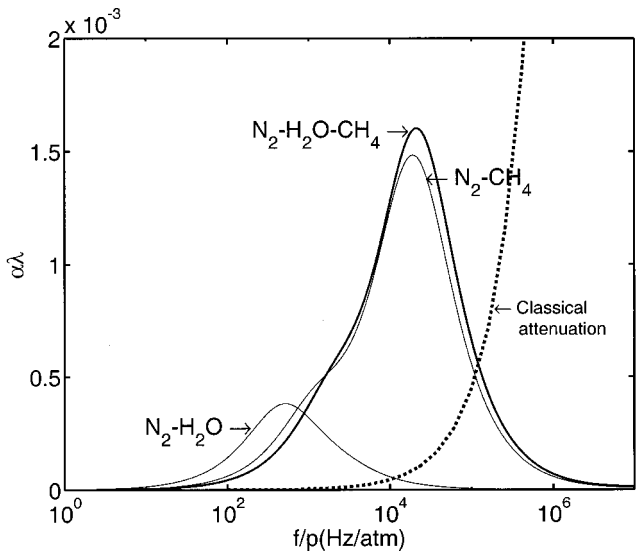


FIG. 5. Comparison of relaxation attenuation curves of a ternary N_2–H_2O–CH_4 mixture with that for binary N_2–H_2O and N_2–CH_4 mixtures ($T=297$ K, $P=1$ atm), and classical attenuation. Bold Solid: Ternary N_2–H_2O–CH_4 mixture with small H_2O and CH_4 additives. Fine Solid: Binary N_2–CH_4 mixture with small CH_4 additive. Fine Solid: Binary N_2–H_2O mixture with small H_2O additive. Dotted: Classical attenuation for the ternary mixture. In all cases the concentration of the small additives is 0.03.

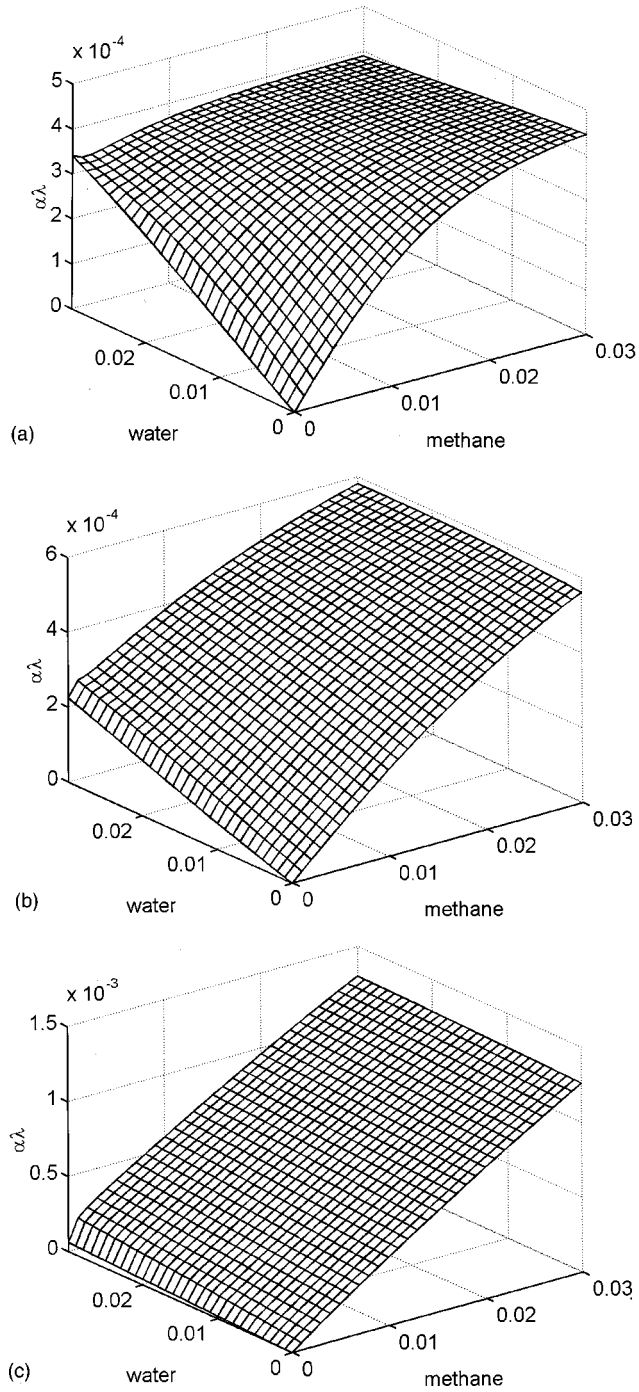


FIG. 6. Dependence of attenuation due to relaxation processes on small concentrations of water and methane for ternary N_2–H_2O–CH_4 mixture ($T=297$ K, $P=1$ atm). (a) 1 kHz; (b) 2 kHz; (c) 10 kHz.

Classical attenuation dominates at higher frequencies.

Figure 6 shows the attenuation coefficient per wavelength versus concentration of the small additives water vapor and methane at different frequencies of the sound wave. At a frequency comparable with the effective relaxation frequency of nitrogen f_1(1 kHz) the surface of the attenuation coefficient demonstrate nonlinear dependence on the concentrations of both small additives. When the concentration of one of the additives is equal to zero the attenuation coefficient is a monotonically increasing function of remaining

TABLE V. Normalized eigenvectors V_i of matrix A for $94\%CH_4-3\%H_2O-3\%N_2$.

	V_1	V_2	V_3	V_4
$N_2(\nu_1)$	1.000	−0.132	−0.000	0.000
$H_2O(\nu_2)$	0.000	−0.524	0.734	−0.999
$CH_4(\nu_3)$	0.000	−0.650	−0.172	−0.000
$CH_4(\nu_4)$	0.000	−0.546	0.657	0.010

additive's concentration. The attenuation coefficient is not a strong function of the concentrations of either constituent when the concentrations are sufficiently large. At 2 kHz, slightly above the lowest relaxation frequency f_1, the attenuation almost linearly depends on concentrations of water and methane. At 10 kHz, a frequency between relaxation frequencies f_1 and f_2, the attenuation is more strongly dependent on methane concentration than on water concentration. Again it is helpful to consider the classical attenuation for comparison. The maximum classical attenuation for any of the three frequencies in Fig. 6 is 4.6×10^{-5} for $100\%N_2$ at 10 kHz/atm. Clearly, the relaxational attenuation is dominant at the frequencies shown in Fig. 6 except at very low methane and water concentrations.

Let us now examine attenuation of sound in a $CH_4-H_2O-N_2$ mixture, when water vapor and nitrogen are small additives. Evaluation of matrix A at 297 K and 1 atm for a $94\%CH_4-3\%H_2O-3\%N_2$ mixture produces the relaxation matrix

$$A=\begin{pmatrix} 2.37e+005 & -2.95e+003 & -2.24e+004 & -1.27e+005 \\ -157 & 1.34e+010 & -3.31e+009 & -9.01e+009 \\ -3.07 & -9.32e+006 & 1.12e+009 & -1.30e+009 \\ -61.8 & -8.36e+007 & -4.28e+009 & 5.10e+009 \end{pmatrix}. \tag{19}$$

The effective relaxation frequencies are

$$f_1=3.77\times10^4\ \text{Hz}, \quad f_2=1.40\times10^5\ \text{Hz},$$
$$f_3=9.75\times10^8\ \text{Hz}, \quad f_4=2.15\times10^9\ \text{Hz}. \tag{20}$$

The corresponding normalized eigenvectors V_i of matrix A for the eigenvalues are shown in Table V. Nitrogen plays a primary role only in the lowest relaxation frequency. The higher relaxation frequencies result from the relaxing complex of methane and water. In Fig. 7 only one peak corresponding to the effective relaxation frequency $f_2=1.4\times10^5$ is evident in the attenuation curve for the ternary mixture. The peak can be attributed to the relaxation of a water-methane complex modified slightly by the presence of nitrogen. The effective relaxation frequencies f_1, f_3, and f_4 do not play a significant role. The relaxation attenuation curves of a $97\%CH_4-3\%H_2O$ and a $97\%CH_4-3\%N_2$ binary gas mixtures merge and are similar to the attenuation of a ternary mixture, indicating the dominant role of methane. The dependence of the effective relaxation frequency f_2 on the nitrogen and water vapor concentrations is quite small, varying by only about 2% over 0% to $3\%CH_4$ and 0% to $3\%N_2$. The classical attenuation is negligible in the range of effective relaxation frequency f_2, but dominates at higher frequencies.

Figure 8 shows the attenuation versus concentrations of the small additives water vapor and nitrogen at different frequencies of the sound wave. The attenuation linearly depends on concentrations of water and nitrogen for all frequencies, although the variation in attenuation is quite small. At 10 kHz, which is much smaller than the effective relaxation frequency f_2, and at 1 MHz, which is much larger than the effective relaxation frequency f_2, the attenuation due to relaxation is quite small compared to the attenuation at 150 kHz, which is quite close to the effective relaxation frequency f_2. The maximum classical attenuation over the range of concentrations is 2.9×10^{-5} at 10 kHz, 4.4×10^{-4} at 150 kHz, and 2.9×10^{-3} at 1 MHz. Thus the classical attenuation is quite small compared to the attenuation due to relaxation at the two lower frequencies. However, at the higher frequency, the classical attenuation is the same order as the relaxational attenuation.

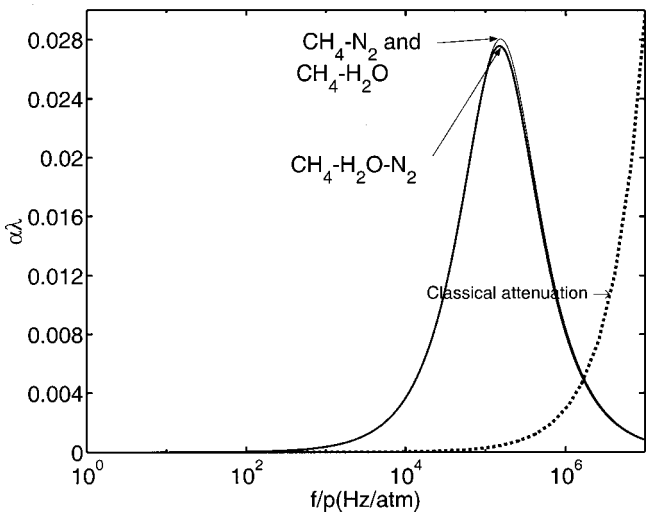


FIG. 7. Comparison of relaxation attenuation curves of a ternary $CH_4-H_2O-N_2$ mixture with that for binary CH_4-H_2O and CH_4-N_2 mixtures ($T=297$ K, $P=1$ atm), and classical attenuation. Bold Solid: Ternary $CH_4-H_2O-N_2$ mixture with small H_2O and N_2 additives. Fine Solid: Binary CH_4-N_2 mixture with small N_2 additive and CH_4-H_2O mixture with small H_2O additive. Dotted: Classical attenuation for the ternary mixture. In all cases the concentration of the small additives is 0.03.

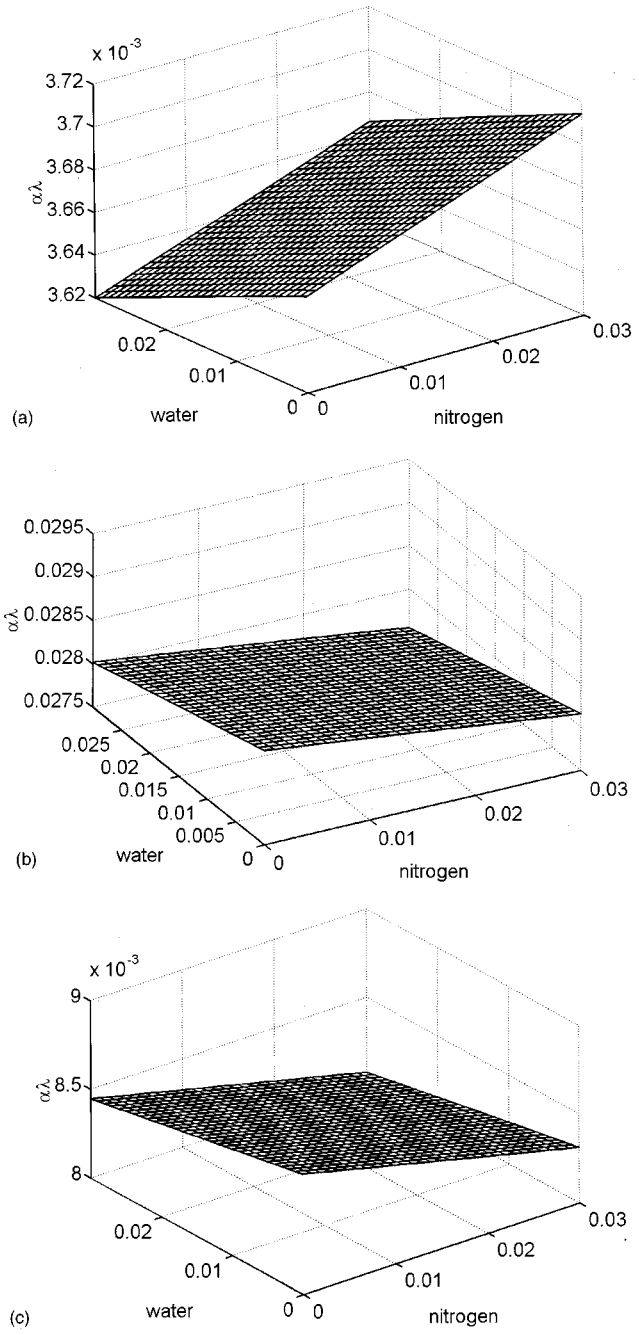


FIG. 8. Dependence of attenuation due to relaxation processes on the concentration of water and nitrogen for ternary $CH_4-H_2O-N_2$ mixtures ($T=297$ K, $P=1$ atm). (a) 10 kHz; (b) 150 kHz; (c) 1 MHz.

V. CONCLUSIONS

The attenuation due to relaxation processes is the major contributor to acoustic attenuation at low frequencies. Unfortunately, until now the attenuation due to relaxation has been quite difficult to calculate. In this paper we have applied the theory for vibrational energy transfer to multicomponent mixtures. Previously this theory had been applied to single-component and binary mixtures. We have extended the theory to three-component mixtures and indicated how the theory could be extended to mixtures with more components. The theory matches experimental data quite well for a N_2-H_2O binary mixture and for pure methane.

We have applied the theory to mixtures of nitrogen, water vapor, and methane where either nitrogen or methane has been the dominant component of the mixture. The relaxation frequencies that were calculated are strongly dependent on the concentration of small additives, even though the concentration of small additives never exceeded 3%. Two relaxation frequencies appear for a mixture that is primarily nitrogen, although one is so strong that it overwhelms the other. Only one relaxation frequency resulted for the attenuation of the gas mixture that was primarily methane.

The important role of small additives is also evident in the acoustic attenuation resulting from the relaxation processes. In the case where nitrogen is the primary component, the attenuation is strongly dependent upon the small concentration of the additives. The influence of additives on the attenuation in a gas mixture that is primarily methane is much smaller. Nevertheless, the attenuation due to relaxation processes is much larger than that due to classical attenuation at low frequencies.

The analysis in this paper makes it clear that the theory of vibrational relaxation can be extended to multi-component mixtures to calculate the attenuation of sound. The weakest portion of the model is the estimation of the transition probabilities in SSH–Tanczos model where the quasi-classical approximation was used for the Lennard-Jones potential at relatively low temperatures. More realistic collision potential functions and the use of molecular dynamics to determine transition probabilities could significantly improve the model. Nevertheless, the model provides a theoretical means to investigate the influence of various conditions including temperature, pressure, and constituent gas concentrations on relaxational attenuation.

ACKNOWLEDGMENTS

The work was supported by DOE under subcontract to Commercial Electronics, Inc., Broken Arrow, OK and by the Ford Motor Company. We thank Dr. Jaco Visser and Scott Phillips for several helpful discussions.

[1] C. Zener, Phys. Rev. **37**, 556–569 (1931).
[2] L. Landau and E. Teller, Phys. Z. Sowjetunion **10**, 34–43 (1936).
[3] L. Landau, Phys. Z. Sowjetunion **1**, 88–98 (1932).
[4] L. Landau, Phys. Z. Sowjetunion **2**, 46–51 (1932).
[5] H. O. Kneser, J. Acoust. Soc. Am. **5**, 122–126 (1933).
[6] R. N. Schwartz, Z. I. Slawsky, and K. F. Herzfeld, J. Chem. Phys. **20**, 1591–1599 (1952).
[7] F. I. Tanczos, J. Chem. Phys. **25**, 439–447 (1956).
[8] K. F. Herzfeld and T. H. Litovitz, *Absorption and Dispersion of Ultrasonic Waves* (Academic, New York, 1959).
[9] G. M. Burnett and A. M. North, *Transfer and Storage of Energy by Molecules* (J. W. Arrowsmith Ltd., Bristol, 1969), Vol. 2.
[10] J. D. Lambert, *Vibrational and Rotational Relaxation in Gases* (Clarendon, Oxford, 1977).
[11] H. J. Bauer, ''Phenomenological theory of the relaxation phenomena in gases,'' in *Physical Acoustics*, edited by W. P. Mason (Academic, New York, 1965), Vol. 2, Pt. A, Chap. 2, Sec. E, pp. 47–131.
[12] A. J. Zuckerwar and K. W. Miller, J. Acoust. Soc. Am. **84**, 970–977 (1988).
[13] M. C. Henderson and K. F. Herzfeld, J. Acoust. Soc. Am. **37**, 986–988 (1965).
[14] H. E. Bass, L. C. Sutherland, J. Piercy, and L. Evans, ''Absorption of Sound by the Atmosphere,'' in *Physical Acoustics*, edited by W. P. Mason and R. N. Thurston (Academic, Orlando, 1984), Vol. XVII, pp. 145–232.
[15] L. W. Townsend and W. E. Meador, J. Acoust. Soc. Am. **99**, 920–925 (1996).

[16] A. J. Zuckerwar and W. A. Griffin, J. Acoust. Soc. Am. **69**, 150–154 (1981).
[17] M. C. Henderson, K. F. Herzfeld, J. Bry, R. Coakley, and G. Carriere, J. Acoust. Soc. Am. **45**, 109–114 (1969).
[18] M. N. Kogan, *Rarefield Gas Dynamics* (Plenum, New York, 1969).
[19] G. Herzberg, *Infrared and Raman Spectra of Polyatomic Molecules* (Van Nostrand, New York, 1945).
[20] J. P. Holman, *Thermodynamics* (McGraw-Hill, New York, 1988), p. 684.
[21] A. Eucken and S. Aybar, Z. Phys. Chem. Abt. B **46**, 195–211 (1940).
[22] P. D. Edmonds and J. Lamb, Proc. Phys. Soc. London **72**, 940–948 (1958).
[23] J. G. Parker and R. H. Swope, J. Chem. Phys. **43**, 4427–4434 (1965).
[24] J. C. Gravitt, C. N. Whetstone, and R. T. Langemann, J. Chem. Phys. **44**, 70–72 (1965).
[25] T. L. Cottrel and M. A. Day, J. Chem. Phys. **43**, 1433–1434 (1965).
[26] PPDS2 for Windows (National Engineering Laboratory, England, 1998).

Elastic constants determination of anisotropic materials from phase velocities of acoustic waves generated and detected by lasers

F. Reverdy and B. Audoin[a)]
Université Bordeaux 1, Laboratoire de Mécanique Physique, UMR 5469, 351 Cours de la Libération, 33405 Talence cedex, France

(Received 8 February 2000; revised 7 August 2000; revised 15 November 2000; accepted 31 January 2001)

This paper describes a method which allows the synthesis of acoustic plane waves by means of laser ultrasonic techniques. When a laser line source radiates at the surface of an anisotropic material, intricate effects are observed on signals obtained by transmission experiments. The material characterization from such a set of signals is not a trivial task since it requires correctly selecting the data necessary for the elastic constants determination, by distinguishing them from all the parasitic phenomena. In the present paper, it is shown that a judicious sum of signals obtained for various positions of the line source at the surface of a sample leads to the synthesis of an acoustic plane wave for which the odd phenomena disappear. Moreover, by applying a constant delay between the various signals, acoustic plane waves can be synthesized with varied refracted angles. One of the advantages of this technique is to offer convenient access to the phase velocity, providing an easier determination of the elastic constants. This method is successively applied to recover the stiffness coefficients of a silicon crystal and of a composite material. The good agreement between the so-recovered elastic constants and values given by other processes shows the contribution of such a method in the field of material characterization.
[DOI: 10.1121/1.1358298]

PACS numbers: 43.35.Cg [SGK]

I. INTRODUCTION

Ultrasonic techniques are customarily used to measure the elastic properties of anisotropic materials. The stiffness coefficients can be determined by processing the velocities measured in several directions through the sample. The traditional characterization techniques by which transducers are in contact with the specimen require cutting the sample in a number of determined directions at least equal to the number of unknown stiffness coefficients.[1] For instance, for orthotropic material six constants can be determined by separately investigating samples cut into three planes of symmetry. The three remaining coefficients are obtained with the aid of measurements in off-axis directions. This technique is inherently destructive; moreover, some composite materials cannot be built in sufficient thickness to offer a convenient set of samples with correctly oriented facets. On thin samples, only two sides of the specimen are experimentally accessible, and ultrasounds must be refracted in several directions to measure a sufficient number of velocities. Other methods deal with acoustic resonant spectroscopy to determine the material properties. However, the processing of these frequencies requires the accurate knowledge of the shape and the mass of the sample.[2,3]

Most of the nondestructive techniques use diagonally incident plane waves propagating from a coupling medium into the tested material to generate acoustic waves through the sample.[4] A closed form connecting the phase velocities of the refracted plane waves with the stiffness coefficients is then provided by the Christoffel equation.[5]

However, the inspecting methods quoted above require either a contact with the sample or a coupling medium, and thus they are not fitted for the purpose of thin material characterization at elevated temperatures. Since many of the new composite materials are intended to be used in extreme conditions of temperature, the interest in ultrasonic contactless techniques has increased. Laser ultrasonic techniques are particularly well suited to perform both generation and detection of ultrasounds without any contact with the specimen.[6] For these reasons, many works were carried out to describe the radiation of a laser source into materials.[7,8] These works have shown that when a point-like or a line laser source is used, the velocity measurements lead to group velocity data. The main difficulties arising from the identification of material properties are to measure the group velocities accurately by processing experimental waveforms.[9]

Studies dealing with the radiation of a laser source have shown that the directivity pattern of the generated acoustic beam exhibits some lobes in determined directions.[10] In order to control the amplitude of the acoustic field and the direction in which the energy is radiated, numerous schemes have been proposed involving the addition of an ensemble of acoustic waves generated by discrete or continuous arrays of laser sources. The first works have dealt with the sound generation in liquid by moving a laser beam over the surface of the liquid.[11,12] It was demonstrated that the underwater ultrasounds generated by the moving laser source were very directive. Then, arrays of laser sources were applied to im-

a)Electronic mail: audoin@lmp.u-bordeaux.fr

prove the beamsteering of laser generated ultrasounds in solids. First of all, a fiberoptic phased array was used to focus ultrasounds in specific directions of aluminum sample. It was shown that the peak amplitude of the acoustic waves could be substantially increased.[13] Afterwards, instead of splitting the laser beam, a multiple-beams YAG laser was used to increase the acoustic power and to obtain a narrow and steerable ultrasonic beam in duraluminum and steel plates.[14] Recently, Lamb waves generation using array of laser sources was investigated since an adjustment of the spacing between the various sources of the array allows the generation of narrow-band ultrasounds and thus the selection of specific Lamb wave modes.[15,16]

In this paper, an alternative approach to the point-like or line laser source is proposed in order to overcome the difficulties encountered in the retrieval of the elastic coefficients when using such sources. The method consists of synthesizing the waveform that would have been recorded if a plane wave had been generated into the composite plate by means of an array of laser line sources. The purpose is to compare the reliability of elastic constant determination from experimentally measured group and phase velocity data. In a recent paper, Degtyar and Rokhlin have demonstrated that for the same accuracy of the phase and group velocity data, the same accuracy in the elastic constants can be obtained.[17] Nevertheless, these conclusions were drawn on synthetic velocities and thus intricate effects attached to the radiation of a line- or a point source into an anisotropic material were not taken into account. In this paper, the inversion problem is solved starting with experimentally recorded waveforms.

The first part deals with the problems connected to the recovering of the stiffness coefficients when a laser line source is focused at the surface of an anisotropic material. The extrapolation to an array of sources as well as the conditions of plane-wave synthesis are then presented. Following, the application of the method to the recovering of material properties is performed by considering either an elastic or a viscoelastic material. Finally, the difference between the so-identified coefficients and the values obtained by either the literature or the method based on group velocities shows improvements brought by the proposed characterization process. The increase in the field of implementation of material properties investigation using laser ultrasonic techniques is especially enlightening.

II. LINE SOURCE IN ANISOTROPIC MEDIA

The waveforms generated through a plate when a laser line source is used are first presented, and the application to material characterization is analyzed second. Two materials are chosen for the discussion. First, a silicon crystal is considered since literature gives accurate values of its stiffness coefficients. Then, the characterization of a highly anisotropic and viscoelastic composite material is performed, and the consequences of the absorption on material inspection are analyzed.

First, let us describe the experimental setup with convenient details.[8] The ultrasounds are generated with the help of a Nd:YAG laser, the pulse burst duration of which is about 20 ns. The laser beam is focused by a set of lens to form a line illumination whose length and thickness are about 4 cm and 0.1 mm, respectively. This line can be moved in a direction normal to it along the interface of the sample by means of a system composed of mobile mirrors. The power density of the laser pulses is such that ablation occurs at the surface of the sample, vaporizing a small amount of material. The corresponding acoustic source is assumed to be a force normal to the loaded surface.[18] The ultrasounds generated by the line source are detected on the opposite interface of the plate by means of a laser probe, which measures the normal displacements of the surface.[19] This probe is a Mach–Zehnder interferometer, where the bandwidth extends from 20 kHz to 45 MHz. The direction of observation is then defined by the angle θ between the source–receiver direction and the plate normal. The particular value $\theta=0°$ is related with the epicenter position. A schematic view of the experimental setup is shown in Fig. 1.

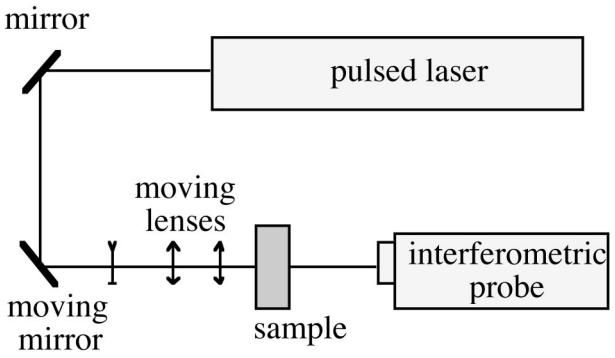


FIG. 1. Experimental setup.

The spatial profile of the acoustic source can be regarded as an infinite line along one direction and a delta function across its width. This source generates transient divergent waves for which the propagation is governed by group velocity curves. Using a geometrical construction, the ray theory gives the directions of the energy flux.[9]

Let us consider a silicon sample with thickness e and density ρ of 5 mm and 2.33 kg/dm^3, respectively. The surfaces of the plate have been cut in the plane (2,2,0) of the crystal. The normal to the interfaces of the plate, denoted axis X_1, is along the crystallographic direction [1,1,0], while directions X_2 and X_3 belonging to the sample surface correspond to crystallographic axes [−1,1,0] and [0,0,1], respectively. The orientation of the crystallographic axes are represented in Ref. 20. Despite the cubic symmetry of silicon, the particular orientation of the sample cut makes the wave propagation equations in the sample axes similar to those considered when propagation occurs in a material showing an orthorhombic symmetry. The components of the stiffness tensor in the sample axes can be calculated from those given by literature in the crystal axes for silicon. The so-calculated components are reported in Table II as reference values. The line source lies in the X_3 direction, thus, the plane (X_1,X_2) is investigated. In such a principal plane of an anisotropic material, three waves may be generated. However, since the source is a line, the shear mode whose polarization is along axis X_3 is not generated. Referring to ray theory, the group velocities of the two remaining modes are calculated and

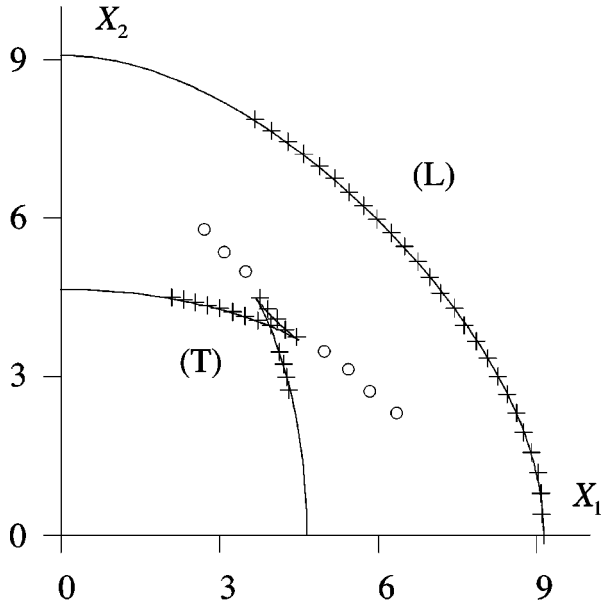


FIG. 2. Group velocities (km/s) recovered by processing experimental signals measured in plane (X_1, X_2) of a silicon sample. Data denoted by circles illustrate the effect of diffraction by the cusp edges.

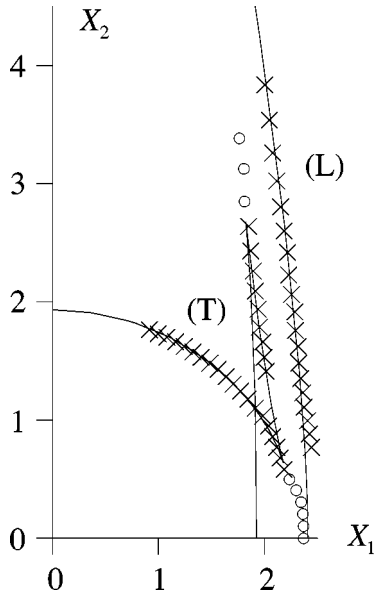


FIG. 3. Group velocities (km/s) in a composite sample obtained by processing experimental signals measured in plane (X_1, X_2). Data denoted by circles illustrate the effect of diffraction by the cusp edges.

they are denoted by solid lines in Fig. 2. The anisotropy of the medium yields to a characteristic shape in the sense that the group velocity curve of the quasishear mode contains cuspidal edges where the curve folds back on itself. The crosses and circles in Fig. 2 denote group velocities measured by processing experimental signals. One may observe beyond the predicted edges of the ray surface some features (circles in Fig. 2), which may be misconstrued. These data do not correspond to arrivals of waves predicted by the ray theory.

In order to explain the existence of such points, the finite wavelength of the acoustic waves should be considered. Indeed, diffraction occurs at the extremities of the cusp, and ultrasounds can be detected in directions for which no acoustic ray can be calculated. Maris has described this diffraction phenomenon in the geometrical shadow, extrapolating for the study of phonon focusing[21] the works of Pearcey performed in electromagnetism.[22] The phenomenon was also experimentally observed at ultrasonic frequencies using focused monochromatic acoustic sources.[23] Its observation when a broadband laser source is used and its effect on materials characterization is discussed in Ref. 8 and in the remainder of this section.

The retrieval of the stiffness coefficients of a composite material using laser ultrasonic techniques is now the focus of this section. This material is a unidirectional composite made of carbon fibers embedded into a polyimide resin. The sample is considered as an infinite plate where thickness e and density ρ are equal to 6.3 mm and 1.7 kg/dm^3, respectively. The direction of the fibers, axis X_2, and the plate normal, axis X_1, are principal directions. The remaining, axis X_3, is parallel to the interface of the plate. Because of that structure, we consider this medium to exhibit an orthorhombic symmetry.

In order to measure group velocities, an accurate estimation of the arrival times is necessary. However, owing to the viscoelactic behavior of this unidirectional material, the waveforms are spread in comparison with those observed in the case of an elastic material. Moreover, the proximity of the quasilongitudinal and quasishear group velocities in certain directions increases the possibility of modes overlapping.[8] With the aim of clearing up this difficulty, a time–frequency processing is used based on wavelets transform.[24] The arrival times of the various waves are then calculated by searching for the maxima of the modulus of the wavelet transform. The wavelet used in this work is the second derivative of Gaussian.

The line source, which lies along the X_3 direction, is displaced along the principal direction X_2. A set of transmitted signals is recorded and processed in order to measure the arrival times of the acoustic waves for various directions of observation. The group velocities deduced from these times are plotted in Fig. 3 as a polar function of angle θ. Again, they are denoted by either crosses or circles, the latter refers to diffraction effects. The velocities are then introduced into a double numerical process described in Ref. 7. Let us underline that the velocities related to diffraction effects are not taken into account by the recovering algorithm. The solving of the inverse problem provides the stiffness coefficients reported in Table I as well as their associated 90%-confidence intervals. The latter is not representative of an exact error, since systematic errors performed on the velocity measurement are not taken into account.[25] It is worth noting that the

TABLE I. Stiffness coefficients and 90%-confidence intervals (GPa) measured for the unidirectional composite with the optimization process based on group velocities and with the optimization process based on phase velocities.

	C_{11}	C_{22}	C_{12}	C_{66}
Group velocities	9.7±0.4	188±97	3.5±3.5	6.7±0.4
Phase velocities	10±0.4	186±16	4.8±2.6	6.6±0.3

arbitrariness as to where to cutoff wave arrival data that is presumed to lie beyond the cuspidal edges seriously affects the values of the recovered elastic constants, particularly C_{12}. This highly anisotropic and viscoelastic medium provides a limit case for the characterization method based on the optimization of group velocities.

So, to improve the measurement accuracy, an alternative approach is proposed in the remainder of this paper.

III. SYNTHESIS OF ACOUSTIC PLANE WAVES

When dealing with immersion experiments, the retrieval of the elastic constants is performed by measuring phase velocities of acoustic waves, which have propagated through a plate for various refracted angles. The purpose of this part is to build the waveforms, which lead to merely measuring the phase velocities, using laser ultrasonic techniques.

The method chosen by the authors consists of synthesizing the signal $s(t)$ that would have been obtained if an array of laser line sources had been used, and processing it to extract the phase velocities of the various generated plane waves. This array is built by displacing the line source along one interface of the sample with a constant step δx. For each position of the line source, the signal $s_i(t)$ obtained on the opposite interface is recorded. In this way, a set of $2N+1$ signals $s_i(t)$ is created. Then, a constant time delay δt is applied between two successive signals $s_i(t)$. Since the origin of times is chosen for the signal $s_0(t)$ obtained at the epicenter position, the shift in time is performed in an antisymmetric way with respect to the signal obtained for this particular position. Finally, the sum of all these shifted signals is made according to the following relation:

$$s(t)=\sum_{i=-N}^{N} s_i(t+i\,\delta t). \tag{1}$$

When dealing with refraction, Snell's law gives a unique, intimate relationship between the incident and the refracted waves in the sense that they must have the same projected wave number along the interface. This relation is of great importance to find the propagation directions of the refracted waves. Assuming that the ratio $\delta x/\delta t$ is equivalent to the projection along the free surface of the phase velocity of an incident plane wave, Snell's law provides the following relation:

$$\frac{\delta t}{\delta x}=\frac{\sin\varphi}{V_p}, \tag{2}$$

where V_p denotes the phase velocity of the refracted plane wave and φ stands for the refracted angle. For isotropic media, relation (2) allows one to determine the value of the refracted angle φ for each of the scattered waves. Conversely, for anisotropic materials these angles are calculated by introducing relation (2) into the Christoffel equation.

At this stage, it is natural to define an upper boundary for the parameter δx, below which the conditions of plane-wave synthesis are satisfied. The conditions of optimal sampling are analyzed by extrapolating the Shannon sampling theorem from time to space. We consider that the spatial sampling frequency k_s must satisfy the following relation:

$$k_s \geqslant 2k_m, \tag{3}$$

where k_m is the wave vector beyond which the displacement field can be considered equal to zero when it is calculated versus the wave vector k_2.

The wave vectors k_s and k_m are defined by relation (4) and (5)

$$k_s=\frac{2\pi}{\delta x}, \tag{4}$$

and

$$k_m=\frac{\omega_m}{V_p\sin(\varphi)}, \tag{5}$$

where ω_m is the central frequency of the wavelet used for the signal processing.

Thus, relation (3) leads to the following condition:

$$\delta x \leqslant \frac{\pi V_p\sin(\varphi)}{\omega_m}. \tag{6}$$

For δx values greater than the boundary defined by relation (6), the sum defined by relation (1) may lead to an apparent modulation of signal $s(t)$, significant of an undersampling phenomenon.

With the aim of characterizing materials, one should first investigate the forward problem of the synthesis of the displacement field $s(t)$ obtained when the array of line sources previously described is used. In a previous publication, a model was developed to represent the displacement field involved by the propagation of acoustic waves in anisotropic and viscoelastic materials.[8] For stiffness coefficients close to those of the unidirectional composite studied in this article, this tool allows the calculation of numerous signals $s_i(t)$. Afterwards, several signals $s(t)$ are built for various refracted angles by tuning the time delay δt.

When dealing with refraction problems, the concept of critical angles must be considered.[9] In our context, this notion should be replaced by the notion of critical slownesses. In a polar representation of phase slownesses, we consider a guide line perpendicular to the axis which denotes the slowness along the interface. The distance of this guide line from the origin equals the source slowness $\delta t/\delta x$. For a given source slowness, a refracted wave will propagate if the guide line intersects the slowness curve corresponding to that mode. For higher source slowness, the guide line moves further and passes positions of tangency with respect to the slowness curves of the various modes. Assuming that S_L is the position of tangency of the guide line with the slowness curve of the quasilongitudinal mode, slownesses greater than S_L are called critical slownesses. Figures 4(a) and (b) represent the displacement $s(t)$ calculated for two values of the source slowness $\delta t/\delta x$, lesser and greater than S_L, respectively. A set of 512 sources is used to synthesize the waveforms shown in Figs. 4(a) and (b). The δx distance between the different sources equals 0.1 mm to satisfy the criteria defined by relation (6).

The first establishment that can be made is that neither cuspidal contribution nor any diffraction effects appear in the synthesized waveforms. Figure 4(a) shows the arrivals of

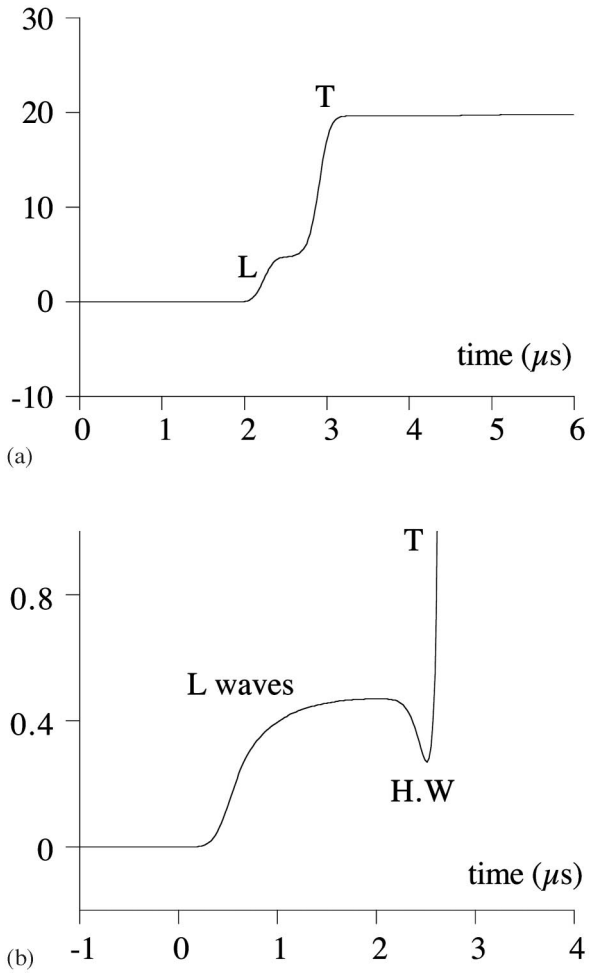


FIG. 4. Signal $s(t)$ (a.u.) built with a source slowness $\delta t/\delta x$ lesser (a) and greater (b) than the critical slowness S_L.

two ultrasonic waves, which can be identified as the quasilongitudinal (L) and the quasishear (T) plane waves. The arrivals of these bulk waves could be predicted since for such source slowness $\delta t/\delta x$, the guide line intersects the slowness curve of the two modes. Figure 4(b) is plotted in the case of a critical slowness ($\delta t/\delta x > S_L$); thus, the quasilongitudinal bulk plane wave is no longer generated. The vertical scale in Fig. 4(b) is adequately magnified to point out the arrival of acoustic waves at times lesser than the arrival time of the quasishear plane wave (T). In order to explain these precursors, the radiation of each source has to be considered separately. Despite that the sum defined by relation (1) does not lead to a quasilongitudinal plane wave, the individual quasilongitudinal waves radiating from each source are still detected. In such a configuration, the Mach number is lower than unity, i.e., each of these waves propagates faster along the interface than the source, so their wavefronts never intersect each other and thus cannot yield to the construction of a plane wave. Assuming that the stiffness coefficients and the position of each source are known, it is possible to calculate the group velocities of these waves and consequently their times of flight. The time, for which precursors first appear, is the arrival time of the quasilongitudinal wave radiated from the source which was first excited, i.e., it is the source most distant from the epicenter position among those which have been negatively shifted according to relation (1).

Moreover, other acoustic waves are generated by the various sources. Indeed, each source is able to generate head waves, which derive from mode conversion on the interface. Assuming a quasilongitudinal wave propagating along the interface, the intersection point between its wavefront and the free surface can be considered as a new source that radiates a quasishear wave by diffraction. This latter propagates through the material in a direction defined by critical angle θ_c fixed by Huyghens's principle.[26] The head waves originated from the sources for which the angle θ is equal or greater than θ_c are the only ones detected by the probe. Knowing the group velocity of the quasishear wave at the angle θ_c and the group velocity of the quasilongitudinal wave along the interface, it is possible to calculate the theoretical arrival time of the head waves. By reporting the time of the first head wave in Fig. 4(b), it is easy to show that its contribution to the waveform $s(t)$ is detected at the correct time close to 2.5 μs.

The precursors are present regardless of the distance δx satisfying the criterion defined by relation (6). Indeed, if the step size is reduced, precursors remain in the waveforms. If harmonic sources were considered, precursors would converge to an inhomogeneous surface wave. Since the source is purely impulsive, the shape of the waveforms remains unchanged even if smaller parameters δx are used. A difficulty arises for the characterization of a material for which the slowness S_L is *a priori* unknown, since the distinction between the quasilongitudinal plane waves and the precursors seems to be a difficult task. A criterion will be defined in the following section, providing an efficient mean for the selection of the data.

IV. APPLICATION TO MATERIALS CHARACTERIZATION

The identification of the stiffness coefficients of an elastic and of a viscoelastic media is now performed in order to show the application field of the method. Since this material characterization method is based on plane-wave propagation, the first stage is to show that the phase velocities of these waves are merely accessible.

The group velocity V_g of a plane wave progressing through an anisotropic plate is such that its projection along the direction of the wave vector is equal to the phase velocity[5]

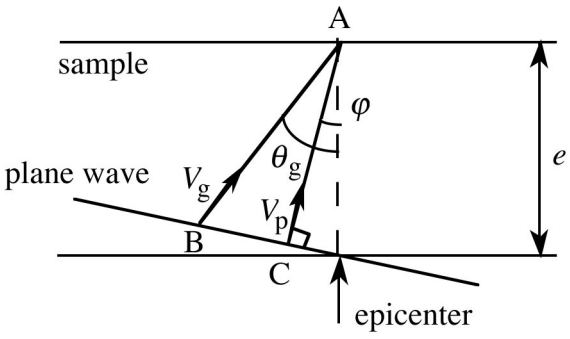


FIG. 5. Paths of energy (BA) and of the equiphase plane (CA).

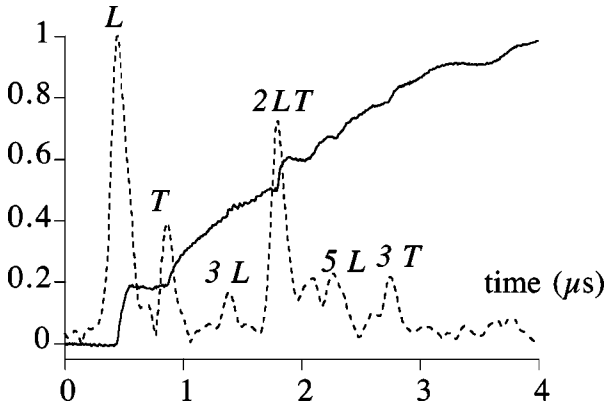


FIG. 6. Signal $s(t)$ (a.u.) built with a source slowness $\delta t/\delta x$ along the silicon sample interface equal to 0.075 ms/m (continuous line) and modulus of the convolution of signal $s(t)$ with a wavelet (dashed line).

$$V_g = V_p/\cos(\theta_g - \varphi), \tag{7}$$

where θ_g is the direction of the group velocity of the considered plane wave.

As a consequence, the time necessary for a wave to propagate along the distance BA at group velocity, in Fig. 5, is equal to the time necessary for the equiphase plane to propagate through the plate along the path CA. Thus, a velocity measurement along the distance CA leads to the phase velocity of the plane wave.

A. Elastic medium

The silicon crystal sample described in Sec. II is the first material to be characterized by the method presented in this paper. A set of $N=78$ signals $s_i(t)$ is experimentally recorded by moving the laser line source in direction X_2 with a constant step $\delta x = 0.2$ mm. The latter respects the criterion defined by relation (6). Angle θ varies from 0° to 72°. The signal $s_i(t)$ obtained for each position of the laser source arises from an average in order to provide the best reproducibility of the acoustic source, and to prevent any random variations of the laser intensity. Once all signals $s_i(t)$ are recorded, a set of signals $s(t)$ is synthesized for various values of the delay δt to obtain velocity data for various refracted angles. As an illustration, the signal $s(t)$ synthesized

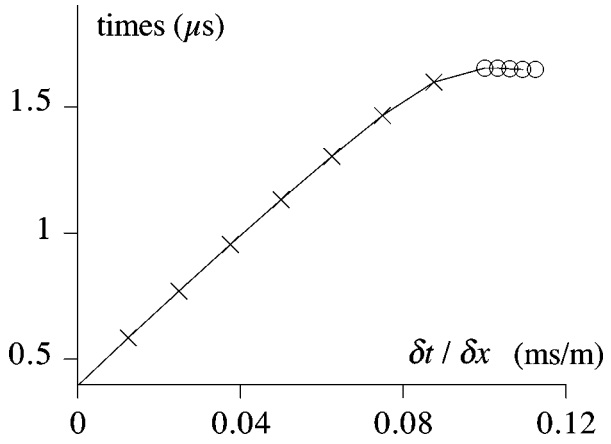


FIG. 7. Times of flight T of the quasilongitudinal plane waves (crosses) and precursors (circles) versus the source slowness $\delta t/\delta x$ along the silicon sample interface.

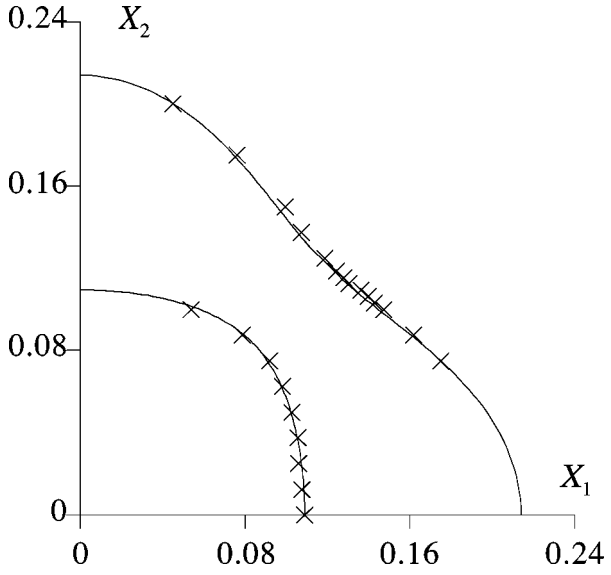


FIG. 8. Slownesses (ms/m) recovered by processing the synthesized signals $s(t)$ obtained from experiments on a silicon sample.

with a source slowness $\delta t/\delta x$ equal to 0.075 ms/m as well as the modulus of the convolution of this signal with a wavelet, whose central frequency is 7.5 MHz, are represented in Fig. 6. The signal processing mentioned in Sec. II provides the measurement of the phase velocities of both modes generated for each signal $s(t)$. A discerning choice should be made for signals $s(t)$ synthesized with a source slowness $\delta t/\delta x$ close to the slowness S_L. Indeed, since this latest value is not exactly known, a criterion should be defined to select the value $\delta t/\delta x$ from which the precursor of the quasishear plane wave is no longer considered as a quasilongitudinal plane wave. When the ratio $\delta t/\delta x$ becomes greater than the critical value S_L, the precursors result from the successive arrivals of the quasilongitudinal waves, radiated from each source, which have forwarded the plate at the group velocity. In Sec. III, we concluded that the time for which precursors first appear is connected with the time of flight of the quasilongitudinal wave radiated by the source the most distant from the epicenter position among those which have been excited at negative times. Let τ be this time of flight. Since this source was excited at time $-N\delta t$, the arrival time t of the precursors measured by processing the signals is $t=\tau-N\delta t$. Therefore, for waves synthesized with source slowness greater than the critical value, the sum $T=t+N\delta t$ is a constant equal to τ, unlike the case of the quasilongitudinal plane waves. This sum is shown in Fig. 7 after processing signals synthesized for source slowness close to the critical value S_L. This plot enables one to extract the data related with the arrival time of the above-mentioned precursors. These data are denoted by circles in Fig. 7. Data marked with crosses are retained to calculate the phase velocities of the quasilongitudinal mode. These velocities, together with velocities of the quasishear mode, are processed by making use of a numerical algorithm described in Ref. 27 to identify the stiffness coefficients. This process minimizes the $L2$ norm of the difference between the measured phase velocities and their counterpart calculated by solving the Christoffel equation. The slownesses deduced from the ex-

TABLE II. Components of the stiffness tensor and 90%-confidence intervals (GPa) calculated for the silicon sample with the optimization process based on the phase velocities and reference stiffness coefficients, both in the sample axes.

	C_{11}	C_{22}	C_{12}	C_{66}
Literature	194.36	194.36	35.24	50.9
Phase velocities	194.3±1.2	196.9±4.1	37.6±2.4	49.8±1.1

perimental velocities are overlaid in Fig. 8 with the slowness curve calculated from the stiffness coefficient given by the literature. At this stage, it is interesting to note that the presence of the head waves does not lead to significant additional data since the amplitude of these waves is relatively low with respect to the amplitude of the quasishear plane wave. The good agreement between the experimental data and the theoretical slownesses points out that phase velocities were measured instead of group velocities. The values of the silicon sample stiffness tensor components measured in the sample axes are reported in Table II as well as their 90%-confidence intervals. Results are in very good agreement with corresponding reference values. The low confidence intervals associated with each of the coefficients show the reliability of the method for elastic medium. It should be noted that smaller parameters δx do not lead to better recovered values.

B. Composite material

In this section, we discuss the identification of the stiffness coefficients of the unidirectional composite described in Sec. II, by using the method based on plane-wave synthesis. For the characterization of this material, $N=89$ signals $s_i(t)$ are recorded by moving the line source along direction X_2. The distance δx between two successive sources is equal to 0.25 mm, ensuring the conditions of optimal sampling defined by relation (6). Otherwise, the same stages in the plane-wave synthesis, as in the case of the silicon plate, are performed. Figure 9 represents the superposition of the slownesses deduced by processing experimental data with the slowness calculated by means of the Christoffel equation. The stiffness coefficients obtained by processing the measured phase velocities are reported in Table I. These coefficients were used to plot the solid lines in Fig. 9.

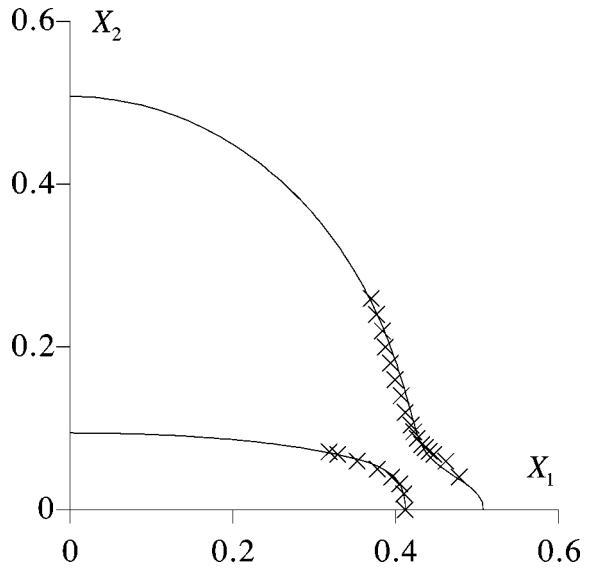


FIG. 9. Slownesses (ms/m) recovered by processing the synthesized signals $s(t)$ obtained from experiments on a unidirectional composite material.

The confidence intervals associated with the coefficients are smaller, showing that the reliability of the method is improved. However, although the confidence interval attached to coefficient C_{12} has decreased, the relative uncertainty is still large. With respect to Fig. 9, it should be noted that there is a lack of information for large refracted angles of the quasilongitudinal mode, unlike the case of the silicon crystal. The criterion described was applied to make the selection of the quasilongitudinal data. Nevertheless, owing to the high anisotropy of the material, the times-of-flight variations of the quasilongitudinal plane wave τ and of those of the precursors versus the source slowness are similar, i.e., the curve tends slowly toward a constant value. Thus, the application of the criterion previously quoted is not a trivial task. The selection of these data is of high importance for the magnitude of the C_{12} confidence interval.

V. CONCLUSION

An array of laser line sources was used to synthesize plane waves. The signals $s_i(t)$ deriving from each line of the array were summed in order to build the waveform obtained when an incident plane wave is refracted by an interface. A shift in time applied between two successive signals allows one to synthesize acoustic plane waves with various refracted angles. This method overcomes the difficulties encountered when dealing with the problems of retrieval of elastic constants by means of laser ultrasonic techniques. Indeed, it was found that the cuspidal contributions as well as the diffraction effects at the edges of the cusp disappear on the synthesized signals. Moreover, the processing of these waveforms leads to the measurement of the phase velocities. The method was applied to a silicon crystal sample and to a composite material. It was demonstrated that the components of the stiffness tensor can be calculated with the aid of the optimization scheme based on phase velocities. The confidence intervals associated with the recovered values show the reliability of the method. The characterization of the unidirectional composite material described in this paper illustrates the extension of the application field of the laser ultrasonic techniques. Other uses can be suggested for the future as the inspection of defects in materials, such as self-aperture microscopy. Indeed, the delay δt can be adjusted separately for each position of the line sources to focus acoustic energy on a defect inside a sample.

[1] R. D. Kriz and W. W. Stinchcomb, ''Elastic moduli of transversally isotropic graphite fibers and their composites,'' Exp. Mech. **19**, 41–49 (1979).
[2] H. Ledbetter, C. Fortunko, and P. Heyliger, ''Orthotropic elastic constants of a boron–aluminum fiber-reinforced composite: An acoustic-resonance-spectroscopy study,'' J. Appl. Phys. **78**, 1542–1546 (1995).
[3] J. Maynard, ''Resonant ultrasound spectroscopy,'' Phys. Today **49**(1), 26–31 (1996).

[4] M. F. Markham, ''Measurement of the elastic constants of fibre composites by ultrasonic,'' Composites **1**, 145–149 (1970).
[5] A. K. Nayfeh, *Wave Propagation in Layered Anisotropic Media* (North-Holland, Amsterdam, 1995).
[6] C. B. Scruby and L. E. Drain, *Laser Ultrasonics Techniques and Applications* (Bristol, Philadelphia and New York, 1990).
[7] M. Deschamps and C. Bescond, ''Inversion of elastic constants from ultrasound group velocity measurements,'' Ultrasonics **33**, 205–211 (1995).
[8] S. Guilbaud and B. Audoin, ''Measurement of the stiffness coefficients of a viscoelastic composite material with laser-generated and detected ultrasound,'' J. Acoust. Soc. Am. **105**, 2226–2235 (1999).
[9] B. A. Auld, *Acoustic Fields and Waves in Solids* (Krieger, Malabar, FL, 1990).
[10] C. Corbel, F. Guillois, and D. Royer, ''Laser-generated elastic waves in carbon–epoxy composite,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **40**, 710–716 (1993).
[11] L. M. Lyamshev and L. V. Sedov, ''Generation of sound by moving pulsed optoacoustic source,'' Sov. Phys. Acoust. **25**, 510–514 (1979).
[12] Y. H. Berthelot and I. Busch-Vishniac, ''Thermoacoustic radiation of sound by moving laser source,'' J. Acoust. Soc. Am. **81**, 317–327 (1987).
[13] J. A. Vogel, A. J. A. Bruinsman, and A. J. Berkhout, ''Beamsteering of laser generated ultrasound,'' Ultrasonics International 87 Conference Proceedings, pp. 141–521 (1987).
[14] M. H. Noroy, D. Royer, and M. Fink, ''The laser-generated ultrasonic phased array: Analysis and experiments,'' J. Acoust. Soc. Am. **94**, 1934–1943 (1993).
[15] R. C. Addison, Jr. and A. D. W. McKie, ''Laser-based Ultrasound Arrays for Generation and Detection of Narrow-band, Single Mode Lamb Waves,'' IEEE Ultrasonics Symposium Proceedings, pp. 1201–1204 (1994).
[16] R. D. Costley, Y. H. Berthelot, and L. J. Jacobs, ''Fresnel arrays for the study of Lamb waves in laser ultrasonics,'' J. Nondestruct. Eval. **13**, 33–42 (1994).
[17] A. D. Degtyar and S. I. Rokhlin, ''Comparison of elastic constant determination in anisotropic materials from ultrasonic group and phase velocity data,'' J. Acoust. Soc. Am. **102**, 3458–3466 (1992).
[18] J. D. Aussel, A. Le Brun, and J. C. Baboux, ''Generating acoustic waves by laser: Theoretical and experimental study of the emission source,'' Ultrasonics **26**, 245–255 (1988).
[19] D. Royer and E. Dieulesaint, ''Optical detection of sub-angström transient mechanical displacements,'' IEEE Ultrasonics Symposium Proceedings, pp. 527–530 (1986).
[20] B. Audoin, C. Bescond, and M. Deschamps, ''Measurement of stiffness coefficients of anisotropic materials from pointlike generation and detection of acoustic waves,'' J. Appl. Phys. **80**(7), 3760–3771 (1996).
[21] H. J. Maris, ''Effect of finite phonon wavelength on phonon focusing,'' Phys. Rev. B **28**(12), 7033–7037 (1983).
[22] T. Pearcey, ''The structure of an electromagnetic field in the neighbourhood of a Caustic,'' Philos. Mag. **37**(264), 311–317 (1946).
[23] K. Y. Kim, K. C. Bretz, A. G. Every, and W. Sasche, ''Ultrasonic imaging of the group velocity surface about the cubic axis in silicon,'' J. Appl. Phys. **79**(4), 1857–1863 (1996).
[24] A. Grossman, M. Holscneideir, and R. Kronland-Martinet, *Inverse Problems* (Academic, New York, 1987), p. 289.
[25] B. Audoin, S. Baste, and B. Castagnède, ''Evaluation of the confidence interval associated with elastic constants identified from ultrasonic velocity measurements,'' C. R. Acad. Sci. Paris, t. 312, Série II, 679–686 (1991).
[26] B. Castagnède, M. Deschamps, E. Mottay, and A. Mourad, ''Laser impact generation of ultrasound in composite materials,'' Acta Acust. **2**, 83–93 (1994).
[27] J. Roux, B. Hosten, B. Castagnède, and M. Deschamps, ''Caractérisation mécanique des solides par spectro-interférométrie ultrasonore,'' Rev. Phys. Appl. **20**(6), 351–358 (1985).

The design, fabrication, and measured acoustic performance of a 1–3 piezoelectric composite Navy calibration standard transducer

Kim C. Benjamin[a] and Sheridan Petrie
Naval Undersea Warfare Center Division Newport, Newport, Rhode Island 02841

(Received 10 April 2000; accepted for publication 3 January 2001)

The design, fabrication, and acoustic calibration of a new 1–3 piezoelectric composite-based U.S. Navy standard (USRD-F82) are presented. The F82 dual array/parametric mode projector may be used as a reciprocal linear transducer, or may be used to exploit the nonlinear properties of the water to produce highly directional acoustic beams (4 to 3 deg) at relatively low frequencies (5 to 50 kHz, respectively). As a result of its wide bandwidth, a broad range of primary as well as secondary frequencies of operation is possible. In the linear mode of operation the transducer provides two separate arrays to be addressed topside for either transmit or receive applications. The two circular apertures are centered on the acoustic axis and have active diameters of 22.8 cm (9 in.) and 5.1 cm (2 in.). The smaller array aperture could be used to obtain broader acoustic beams at relatively high frequencies. Due to the absence of air-filled pressure release components, the transducer will operate over most ocean pressures and temperatures. A general description of the 1–3 piezoelectric composite-based transducer configuration and measured performance is presented. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1358889]

PACS numbers: 43.38.Ar, 43.30.Yj [SLE]

I. INTRODUCTION

In the fall of 1995, the Naval Underwater Sound Reference Division (USRD) of the Naval Undersea Warfare Center (NUWC) Division Newport received funding to develop a replacement for the Navy's general purpose dual-aperture reciprocal transducer standard F33 (see Fig. 1), with a similar design constructed of the relatively new material ''injection-molded 1–3 piezoelectric composite.''[1] This effort represented the Navy's first use of 1–3 piezoelectric composite as a transmit material in the sonar frequency range of 20 to 300 kHz. Earlier acoustic measurements taken at USRD[2,3] on various ceramic volume fractions, and back-fill materials, were evaluated in this study prior to selecting the 1–3 piezoelectric composite configuration for the new F82. Those data indicated that some of the 6.3-mm ($\frac{1}{4}$-in.)-thick 1–3 piezoelectric composite substrates worked well in ''*both*'' the transmit and the receive mode. In fact, some of the rigid epoxy back-fill matrix materials evaluated showed the transmitting voltage response (TVR) of the composite to be very stable over typical ocean temperatures and pressures.

The USRD F33 calibration standard, along with its design variations F27 and F30, have remained the workhorses of the USRD Transducer Standards Loan Program. Most of these standard units remain within service for well over 10 years with minimal change in performance. As seen in Fig. 1, the F33 is constructed from several individual components. The transducer has a rubber molded acoustic window and is oil filled under vacuum.

Within the F33's active 20.3-cm (8-in.) quasicircular aperture, two separate arrays (64 elements, mass-loaded rectangular blocks of piezoelectric ceramic and 12 elements, mass-loaded circular disks of piezoelectric ceramic) allow the transducer to provide a broad spatial response over a wide operational frequency band. For example, either the inner, or both, arrays may be addressed over a given frequency range to obtain a relatively broad, or narrow, acoustic beam pattern, respectively.

The F82 development involved evaluating new material technologies experimentally. The objective was to combine a minimum number of components into a simple transducer assembly with emphasis on cost, ease of fabrication, and improved performance (over the F33). The F82 transducer was designed empirically, rather than using complicated computer models that rely on first principles. Although the latter approach often results in an ''optimal design,'' it is prone to prescribing hardware configurations that may be difficult or impossible to realize repeatedly in a cost-effective manner. In addition, prototypes are needed to validate modeled predictions, ultimately increasing development cost and schedule. The next section qualitatively describes the advantages associated with moving to a 1–3 piezoelectric composite-based transducer design.

II. PROTOTYPE TRANSDUCER FABRICATION ISSUES

The design and fabrication of a reliable broadband underwater transducer having spatial and temporal acoustic fidelity over a wide range of operational environmental conditions, 3–32 °C (37–90 °F) and up to 6895 kPa (1000 psi), was the objective of the F82 transducer development effort. The active material, acoustic backing, and adhesives were chosen on the premise that their dynamic moduli and loss factors would remain relatively stable over the anticipated operational environmental conditions (i.e., temperature and hydrostatic pressure). In moving to a 1–3 piezoelectric com-

[a]Electronic mail: benjaminkc@npt.nuwc.navy.mil

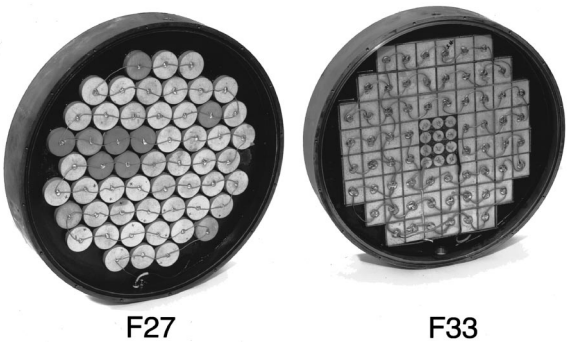


FIG. 1. USRD F27 (left) and F33 (right) calibration transducer standards.

posite transducer design, several advantages were realized. These are elaborated below and where applicable, comparison to the current F33 is made.

A. Temperature and pressure stability

Figure 2 shows the acoustic attenuation of the backing material as a function of frequency. The current F33 standard utilizes Corprene as a ''pressure-release'' component, and therefore is limited to maximum pressures of only 3447 kPa (500 psi).[4]

In contrast, the new F82 uses an ''absorptive'' backing, and a rigid thermoplastic epoxy, behind and around the piezoelectric ceramic rods, respectively, thereby allowing operational pressures in excess of 6895 kPa (1000 psi).

In addition to pressure stability, the viscoelastic components of the F82 show little change over the temperature range of interest, 3 to 32 °C (37 to 90 °F). The temperature

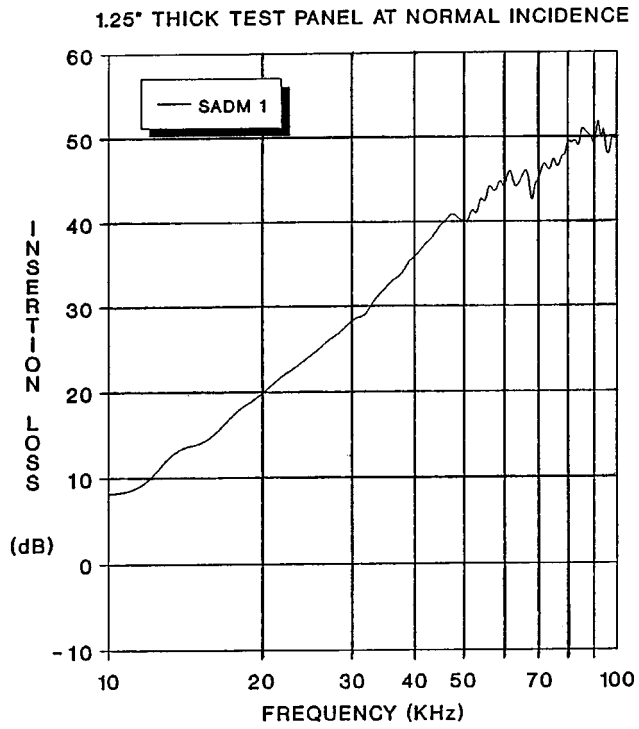


FIG. 2. Acoustic absorption of selected F82 backing material as a function of frequency (data courtesy of Syntech Materials, Inc.).

FIG. 3. Injection-molded 1–3 piezoelectric composite transducer standard (F82).

and hydrostatic pressure dependence of the F82 standard is presented in Sec. IV. The F82 transducer standard is shown in Fig. 3.

B. Low cost of fabrication

As seen in Fig. 4, the F82 is a very simple design to fabricate. Relative to its conventional predecessors, it involves significantly fewer individual components. This is always desirable from a reproducibility perspective. Each part has a specific inherent mechanical and/or electrical tolerance that combines randomly in any given assembly. The fewer the number of individual parts, the less that can go wrong in the fabrication process. The electroplated 1–3 piezoelectric composite, shown at left in Fig. 4, is patterned to realize two

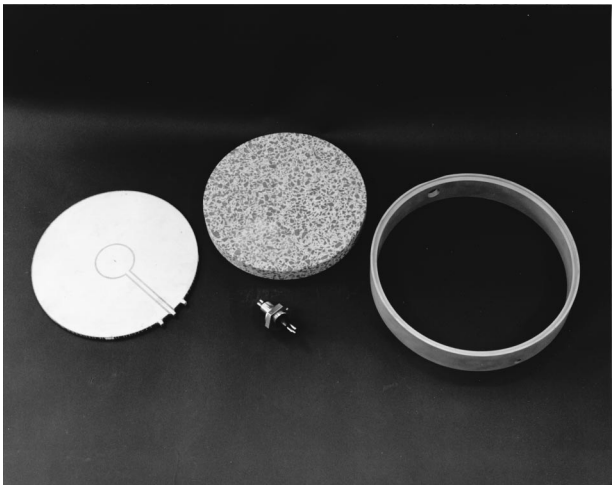


FIG. 4. Four discrete components make up the new F82 transducer standard.

array apertures. The plated tabs that protrude outside of the active aperture facilitate the wire attachment.

C. Available in large sheets

The active material, which combines Navy Type VI (PZT-5H) piezoelectric ceramic rods with a thermoplastic epoxy/microballoon mixture, is available in sheet form up to 30 in. in diameter.[5] This eliminates the need for assembling several discrete blocks of ceramic into an array and then connecting them electrically in parallel to cover a relatively large radiating aperture. The active panels are readily plated with copper to form either single- or multiple-element electrodes. The copper plating also allows the realization of complex shading patterns for controlling the surface velocity of a given aperture.

D. Rugged and shock resistant

The epoxy back-fill material, employed in this particular 1–3 piezoelectric composite configuration, is key to its inherent pressure stability. In addition, explosive shock measurements, made at the USRD conical shock tube facility, indicate that the active material described above is shock hardened.[6]

E. Broadband transmit capability

In general, the moderate damping of the viscoelastic back-fill material of the 1–3 piezoelectric composite results in mechanical quality factors (Q_m) that are typically low. Free of laterally coupled dilatational and shear modes, the transmitting voltage response of the large panel is relatively smooth, nominally changing 12 dB/octave below the half-wave resonance frequency, and minus 6 dB/octave above it.

The parametric mode generation of sound using the 1–3 piezoelectric composite was first demonstrated using the F82 prototype.[7] The drive stability and power density levels achieved by the Navy Type VI ceramic in a 1–3 piezoelectric composite configuration was surprising. Ongoing work, investigating thermal heating effects in thicker (25.4 mm) Navy Type VI composite configurations at high drive, indicate that very high input powers can be used at low ($<2\%$) duty cycles. The F82 was taken to sea and used in the parametric mode for sub-bottom profiling. For that application, the low duty cycle was not a limitation due to the short pulse lengths used, and depth of the water column.

III. DESCRIPTION OF SELECTED DESIGN

The selected transducer design consists of an electroplated 22.8-cm diameter $\times$0.63-cm-thick (9 in.$\times$0.25 in.) disk of injection-molded 1–3 piezoelectric composite. The configuration combines Navy Type VI (PZT-5H) piezoelectric ceramic (25% volume fraction), with a rigid Shore-D epoxy/microballoon thermoplastic matrix. The active substrate is bonded to a 2.5-cm (1.0 in.)-thick layer of SYNTECH/SADM-1 absorptive backing of the same circular diameter, using a structural epoxy. The size of the inner and outer plated regions is commensurate with those of the inner and outer array apertures of the F33. The active subassembly is encapsulated within a stainless-steel (or bronze) housing ring using room-temperature curable polyurethane. A connector receptacle, threaded into the housing ring, allows the transducer to be detached from the cable.

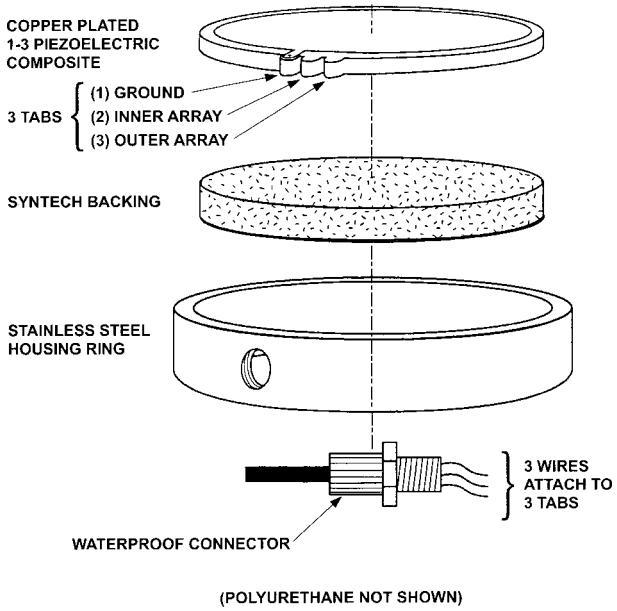


FIG. 5. F82 transducer fabrication concept.

Figure 5 illustrates the F82 transducer assembly. The three wires coming from the connector are attached to the three plated tabs on the 1–3 piezoelectric composite panel using conductive epoxy. The conductive epoxy connection is preferred over a solder joint connection, since the heat from the soldering process causes the copper to expand relative to the epoxy. Shear stresses at the plated interface produce a delamination between the copper and composite. Each tab has plating on one side only. The ground plane, which is common to both arrays, is outward facing to minimize stray capacitance across the polyurethane acoustic window to water.

IV. MEASURED RESULTS

This section presents measured results for the F82 transducer standard. Both linear and parametric mode data collected on the first built prototype are shown and discussed.

A. Linear measurements

The new F82 prototype was acoustically calibrated as a function of temperature (3 and 32 °C) and hydrostatic pressure (up to 6895 kPa) in the USRD Acoustic Pressure Tank Facility. The calibration measurements included the transducer's transmit and receive responses, as well as beam patterns at selected frequencies. The transducer showed little change in performance over this range of typical ocean environmental conditions. Figure 6 shows the F82 transducer's TVR as a function of the two temperatures indicated. Figure 7 shows the F82 transducer's TVR as a function of hydrostatic pressure up to 6895 kPa (1000 psi). As one can see, the response is essentially insensitive to hydrostatic pressure within the range indicated.

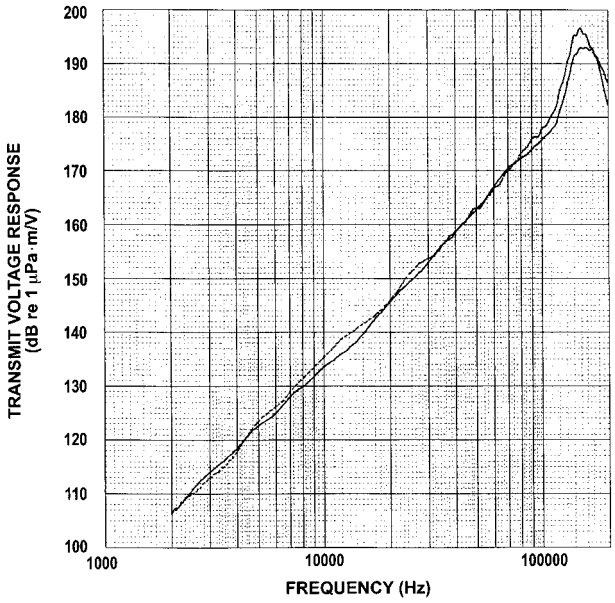


FIG. 6. F82 transmitting voltage response as a function of water temperature (3 and 32 °C).

One also notes from Fig. 7 that the transmit voltage response of the new F82 exceeds that of its predecessor. This is also true of the receiving voltage response that is shown in Fig. 8. The differences are attributed to the use of Navy Type VI piezoelectric ceramic in the 1–3 piezoelectric composite versus Navy Type I, or lead titanate that is used in the F33 transducer standard. On transmit the type VI ceramic affords a higher piezoelectric ''*d*'' constant, which relates induced strain to the applied field. On receive it offers a higher relative dielectric constant, which by virtue of its higher capacitance reduces the loss in sensitivity associated with the transducer cable. Figure 9 shows the measured beam patterns for the inner region and both array section(s) at 20 kHz. One notes a significant difference in beamwidth depending on the

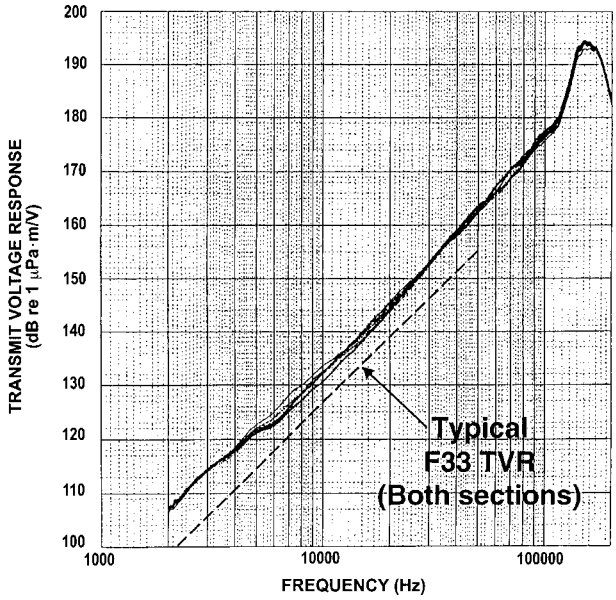


FIG. 7. Comparison of F82 and F33 transmitting voltage response curves (F82—Ambient to 6895 kPa) (F33—Ambient pressure only).

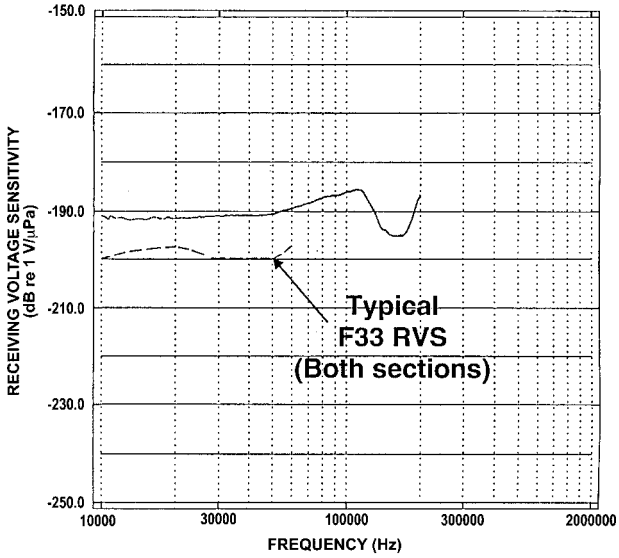


FIG. 8. Comparison of F82 and F33 receiving voltage sensitivity response (Ambient pressure only).

way the unit is operated. The backlobe structure of the beam patterns is a function of operating frequency and hydrostatic pressure, decreasing with frequency and increasing with hydrostatic pressure.

B. Parametric mode measurements

The use of 1–3 piezoelectric composite for the parametric mode generation of sound was first successfully demonstrated by the Navy, at NUWC's Dodge Pond Facility in Niantic, CT. The use of Navy Type VI ceramic, and the 1–3 piezoelectric composite's low efficiency were expected to preclude the material as a candidate for parametric mode sound generation. However, these results indicate that it is an excellent choice. In fact, subsequent parametric mode research by the Navy has indicated that the material provides better overall performance than similar solid ceramic designs. Figure 10 shows the transmitted difference frequency beam patterns for the F82 when excited in the parametric mode. The primary frequencies, approximately 200 kHz, are slightly above the 1–3 piezoelectric composite panel's half-wave thickness mode resonance. The secondary source levels (L_s) at 5, 10, 30, and 50 kHz are indicated on their respective plots in dB *re*: 1 μPa at 1 m. The 3-dB beamwidths ranged from 3 to 4 degrees.

Figure 11 shows the transmitting voltage response of the F82 with a standard length of transducer cable. One notes a resonance frequency of 165 kHz, and a mechanical quality factor of approximately 2.

During the experimental calibrations it was noted that the 1–3 piezoelectric composite has a significantly higher capacitance per unit area than its solid ceramic counterpart and is therefore not affected as much by differences in cable length. Typical solid ceramic parametric mode transducer designs require special narrow-band tuning circuits to mitigate the adverse transmission line effects associated with using long transducer cables at relatively high primary frequencies. This feature therefore makes conventional ceramic-

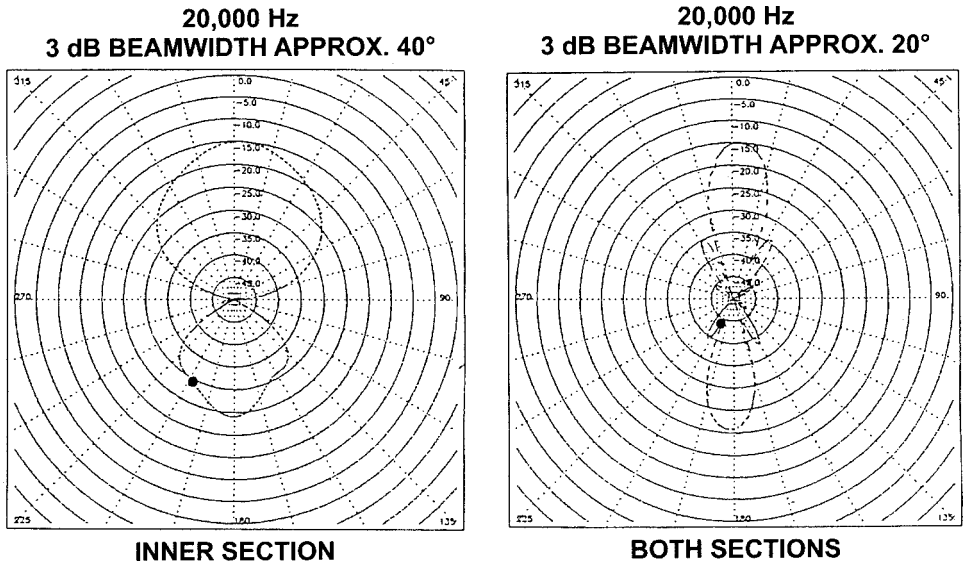


FIG. 9. Comparison of the F82 beam patterns at 20 kHz: (left) inner array; (right) both arrays.

based parametric mode transducers inherently narrow band. No such tuning was required for the F82 or any other subsequent 1–3 piezoelectric composite-based parametric mode designs.

V. CONCLUSIONS

With the larger-spaced lattice of posts and greater substrate thickness, injection-molded 1–3 piezoelectric compos-

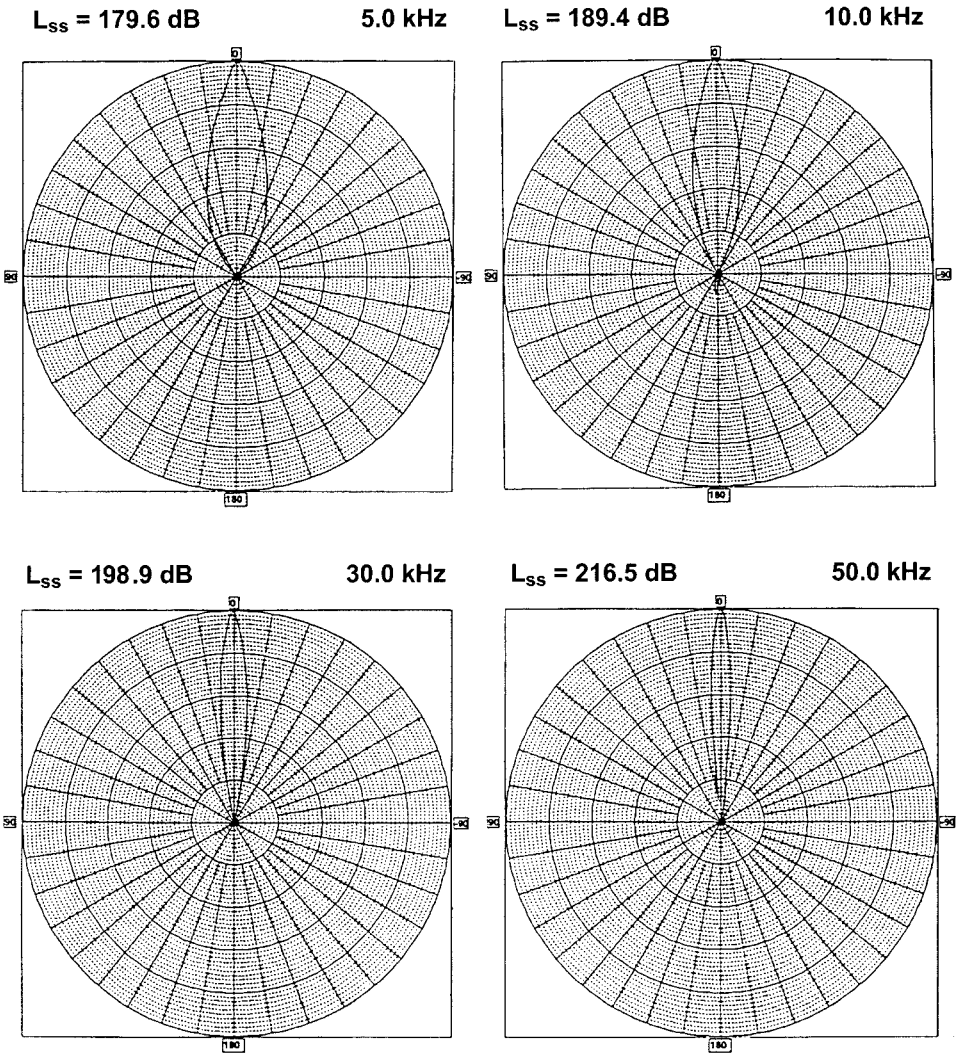


FIG. 10. Parametric mode difference frequency beam patterns for the F82 navy transducer standard.

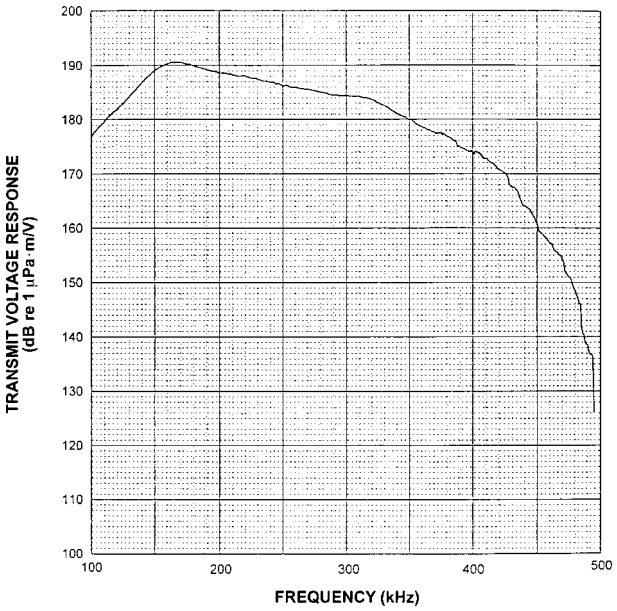


FIG. 11. High-frequency transmitting voltage response of F82 with 30 m of cable.

ite panels bridge the gap between biomedical ultrasonic dice-and-fill 1–3 piezoelectric composite and conventional sonar ceramic constructs. This new transducer material is finding several uses in modern Navy transducer applications. In particular, the construction and measured results for a new 1–3 piezoelectric composite-based Navy calibration transducer standard has been presented. The design incorporates many of the advantages offered by this unique transduction material.

Three distinct advantages over its solid ceramic predecessor are given below:

(i) Improved performance due to the use of Navy Type VI piezoelectric ceramic;
(ii) Temperature and hydrostatic pressure stability; and
(iii) Greatly reduced fabrication time.

As the first demonstration of a 1–3 piezoelectric composite-based transducer for parametric mode generation, the F82 has provided valuable experimental data needed for future development work in the area of parametric mode sonar. The 1–3 piezoelectric composite configuration described in this paper (i.e., Navy Type VI ceramic/rigid epoxy), enables the fabrication of broadband parametric mode arrays that were previously very complicated and costly to build.

ACKNOWLEDGMENTS

This work was sponsored in part by the Office of Naval Research (Code 321), the Naval Research Laboratory (Code 7431), and the Space and Naval Warfare Systems Center (Code D744). The author also wishes to express his sincere thanks to William L. Konrad and Mark B. Moffett for their valuable assistance with the parametric mode experiments.

[1] L. Bowen and K. French, ''Fabrication of piezoelectric/polymer composites by injection molding,'' Proceedings of the 8th IEEE International Symposium on Applications of Ferroelectrics, 1992, pp. 160–164.
[2] USRD Calibration Memorandum No. 0727, ''Measurements on piezocomposite transducers serials 4-48, 4-50, 4-51, 4-10P4, 4-11P4, and 2D-1,'' NUWC–USRD APTF Test Facility for Acoustic Calibration, Naval Undersea Warfare Center, Orlando, FL (April 1996).
[3] R. Y. Ting and T. R. Howarth, ''Piezoelectric 1–3 composite materials for underwater acoustical applications,'' *UDT 1994 Conference Proceedings* (Nexus Business Communications, Kent, UK, 1994), pp. 150–154.
[4] R. J. Bobber, *Underwater Electroacoustic Measurements* (NRL, Washington, D.C., 1969).
[5] K. C. Benjamin, A. L. Van Buren, and S. Petrie, ''The development of affordable constant beamwidth transducers using injection molded 1–3 piezoelectric composite,'' J. Acoust. Soc. Am. **103**, 2755(A)(1998).
[6] T. R. Howarth and R. Gentilman, ''Underwater explosive shock and evaluation of large area 1–3 piezocomposite sensors and actuator structures,'' Proceedings of the 6th International Conference on Adaptive Structures, Key West, FL (November 1995).
[7] K. C. Benjamin, M. B. Moffett, W. L. Konrad, and K. E. Jenne, ''On the use of injection molded 1–3 piezocomposite for parametric mode sonar projectors,'' J. Acoust. Soc. Am. **102**, 3106(A) (1997).

Maximum entropy approach for modeling random uncertainties in transient elastodynamics

Christian Soize[a)]
Structural Dynamics and Coupled Systems Department, ONERA, BP 72, 92322 Chatillon Cedex, France

(Received 5 July 2000; revised 8 February 2001; accepted 9 February 2001)

A new approach is presented for analyzing random uncertainties in dynamical systems. This approach consists of modeling random uncertainties by a nonparametric model allowing transient responses of mechanical systems submitted to impulsive loads to be predicted in the context of linear structural dynamics. The information used does not require the description of the local parameters of the mechanical model. The probability model is deduced from the use of the entropy optimization principle, whose available information is constituted of the algebraic properties related to the generalized mass, damping, and stiffness matrices which have to be positive-definite symmetric matrices, and the knowledge of these matrices for the mean reduced matrix model. An explicit construction and representation of the probability model have been obtained and are very well suited to algebraic calculus and to Monte Carlo numerical simulation in order to compute the transient responses of structures submitted to impulsive loads. The fundamental properties related to the convergence of the stochastic solution with respect to the dimension of the random reduced matrix model are analyzed. Finally, an example is presented. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1360716]

PACS numbers: 43.40.Dx, 43.40.Hb, 43.40.Kd [CBB]

I. INTRODUCTION

This paper deals with predicting the transient responses of structures submitted to impulsive loads in linear structural dynamics. The theory presented below can be extended without any difficulties to structural-acoustic problems such as a structure coupled with an internal acoustic cavity. In general, this kind of prediction is relatively difficult because the structural models have to be adapted to large, medium and small vibrational wavelengths which correspond to the low-, medium-, and high-frequency ranges.

Here, we are interested in the case where the impulsive load under consideration has an energy which is almost entirely distributed over a broad, low-frequency band and for which prediction of the impulsive load response can be obtained with a reduced matrix model constructed using the generalized coordinates of the mode-superposition method associated with the structural modes corresponding to the n lowest eigenfrequencies of the structure. It should be noted that, for a complex structure, only a numerical approximation of the first structural modes can be calculated using a large finite-element model of the structure. The low-frequency case considered in this paper is important for many applications, and details concerning such a case can be found in the literature on structural dynamics and vibrations (see Refs. 1–8).

Under the above assumptions and for a complex structure, dimension n of the reduced matrix model generally has to be high (several dozen or hundred structural modes may be necessary to predict transient responses). However, it is known that the higher the eigenfrequency of a structural mode, the lower its accuracy because the uncertainties in the model increase (in linear structural dynamics and vibrations, the effects of uncertainties on the model increase with the frequency and it should be kept in mind that the mechanical model and the finite-element model of a complex structure tend to be less reliable in predicting the higher structural modes). This is why random uncertainties in the mechanical model have to be taken into account. This is a fundamental problem in structural dynamics and in structural acoustics when the mechanical model has to be adapted to predict a transient response for which not only the low-frequency band is mainly concerned, but also the upper part of this low-frequency band and possibly the medium-frequency band have to be taken into account.

Random uncertainties in linear structural dynamics and structural acoustics are usually modeled using parametric models. This means that (1) the uncertain parameters (scalars, vectors, or fields) occurring in the boundary value problem (geometrical parameters; boundary conditions; mass density; mechanical parameters of constitutive equations; structural complexity, interface and junction modeling, etc.) have to be identified; (2) appropriate probabilistic models of these uncertain parameters have to be constructed, and (3) functions mapping the domains of uncertain parameters into the mass, damping, and stiffness operators have to be constructed. Concerning details related to such a parametric approach, we refer the reader to Refs. 9–15 for general developments, to Refs. 16–21 for general aspects related to stochastic finite elements and to Refs. 22–27 for other aspects related to this kind of parametric models of random uncertainties in the context of developments written in stochastic dynamics and parametric stochastic excitations.

In this paper we present a new approach, which we will call a nonparametric approach, for constructing a model of random uncertainties in linear structural dynamics in order to

a)Electronic mail: soize@univ-mlv.fr

predict the transient response of complex structures submitted to impulsive loads (as indicated above, this approach can be directly extended to structural-acoustic problems). This nonparametric model of random uncertainties does not require identifying the uncertain parameters in the boundary value problem as described above for the parametric approach, but is based on the use of recent research (see Refs. 28 and 29) in which the construction of a probability model for symmetric positive-definite real random matrices using the entropy optimization principle has been introduced and developed. These results will allow the direct construction of a probabilistic model of the reduced matrix model deduced from the variational formulation of the boundary value problem to be obtained, for which the only information used in this construction is the available information constituted of the mean reduced matrix model, the existence of second-order moments of inverses of the random matrices, and some algebraic properties relative to the positive definiteness of these random matrices. It should be noted that these properties have to be taken into account in order to obtain a mechanical system with random uncertainties, which models a dynamical system. For instance, if there are uncertainties on the generalized mass matrix, the probability distribution has to be such that this random matrix be positive definite. If not, the probability model would be wrong because the generalized mass matrix of any dynamical system has to be positive definite.

In Refs. 28 and 29, we presented the calculation of the matrix-valued frequency response functions for discretized linear dynamical systems with random uncertainties. Unfortunately, convergence results were not obtained yet and consequently, a parameter of the probability model was not clearly defined for a designer. In this paper, an explicit construction of the probabilistic reduced matrix model of finite dimension n is given and its convergence is studied as n approaches infinity. In such a probabilistic theory, it seems absolutely fundamental to prove the convergence. It is not self-evident that convergence properties exist in such a construction. In addition, it should be noted that Eqs. (65)–(68) have been deduced from the convergence analysis carried out. Thanks to this new analysis presented in this paper, we have obtained a new consistent and coherent theory in which all the parameters are clearly defined. In Sec. II, the mean boundary value problem is introduced and its variational formulation is given in order to construct the mean reduced matrix model, which is carried out in Sec. III, using the mode-superposition method. Section IV is devoted to construction of the nonparametric model of random uncertainties for the reduced matrix model. In this section, we introduce the available information which is directly used for constructing the probabilistic model of random uncertainties. In Sec. V, we give a summary of the main results established in Refs. 28 and 29 concerning the probability model for symmetric positive-definite real random matrices and we complete this construction in order to obtain a consistent probabilistic model useful for studying convergence as dimension n approaches infinity. The nonparametric model of random uncertainties for the reduced matrix model constructed using Secs. IV and V, is presented in Sec. VI. The convergence properties of this nonparametric model of random uncertainties as dimension n approaches infinity are given in Sec. VII. The convergence properties prove that the construction proposed is consistent. Finally, an example is presented in Sec. VIII.

II. MEAN BOUNDARY VALUE PROBLEM FOR MEAN TRANSIENT RESPONSE AND ITS VARIATIONAL FORMULATION

A. Definition of the mean boundary value problem

We consider the linear transient response of a three-dimensional damped fixed structure around a static equilibrium configuration considered as a natural state without prestresses, submitted to an impulsive load. The mean mechanical model is described by the following mean boundary value problem. Let Ω be the bounded open domain of $\mathbb{R}^3$ occupied by the mean structure at static equilibrium and made of viscoelastic material without memory. Let $\partial\Omega = \Gamma_0 \cup \Gamma$ be the boundary such that $\Gamma_0 \cap \Gamma = \varnothing$ and let $\mathbf{n}$ be its outward unit normal. Let $\mathbf{u}=(u_1,u_2,u_3)$ be the displacement field at each point $\mathbf{x}=(x_1,x_2,x_3)$ in Cartesian coordinates. On part Γ_0 of the boundary, the structure is fixed ($\mathbf{u}=\mathbf{0}$) while on part Γ it is free. There are external prescribed impulsive volumetric and surface force fields applied to Ω and Γ, written as $\{\mathbf{g}_{\rm vol}(\mathbf{x},t), t \geqslant 0\}$ and $\{\mathbf{g}_{\rm surf}(\mathbf{x},t), t \geqslant 0\}$, respectively. Let T be a positive real number. The mean transient response $\{\mathbf{u}(\mathbf{x},t), \mathbf{x}\in\Omega, t\in[0,T]\}$ is the solution of the following mean boundary value problem:

$$\rho \ddot{u}_i - \frac{\partial \sigma_{ij}}{\partial x_j} = g_{{\rm vol},i} \quad \text{in } \Omega, \quad t\in[0,T], \tag{1}$$

$$\sigma_{ij} n_j = g_{{\rm surf},i} \quad \text{on } \Gamma, \quad t\in[0,T], \tag{2}$$

$$u_i = 0 \quad \text{on } \Gamma_0, \quad t\in[0,T], \tag{3}$$

for $i=1,2,3$, with zero initial conditions at time $t=0$,

$$\mathbf{u}(\mathbf{x},0)=0, \quad \dot{\mathbf{u}}(\mathbf{x},0)=0, \quad \forall \mathbf{x}\in\Omega, \tag{4}$$

in which $\dot{\mathbf{u}}$ and $\ddot{\mathbf{u}}$ mean $\partial\mathbf{u}/\partial t$ and $\partial^2\mathbf{u}/\partial t^2$, respectively, where the index summation convention is used and where $\rho(\mathbf{x})>0$ is the mass density of the mean model (which is assumed to be a bounded function on Ω). For a viscoelastic material without memory, the stress tensor σ_{ij} of the mean model is written as

$$\sigma_{ij} = a_{ijkh}(\mathbf{x})\varepsilon_{kh}(\mathbf{u}) + b_{ijkh}(\mathbf{x})\varepsilon_{kh}(\dot{\mathbf{u}}), \tag{5}$$

in which $\varepsilon_{kh}(\mathbf{u}) = (\partial u_k/\partial x_h + \partial u_h/\partial x_k)/2$ is the linearized strain tensor. The mechanical coefficients of the mean model $a_{ijkh}(\mathbf{x})$ and $b_{ijkh}(\mathbf{x})$ are real, depend on $\mathbf{x}$, and verify the usual properties of symmetry and positiveness.[30–32,8]

B. Variational formulation of the mean boundary value problem

The variational formulation of the mean boundary value problem is absolutely necessary to construct the mean reduced matrix model in the general case. In addition, in order to prove the convergence properties of the stochastic transient response of the dynamical system with random uncertainties as the dimension of the reduction approaches infinity,

we need to introduce the set $\mathbb{V}$ of admissible displacement fields. For the mathematical notations used in this section, we refer the reader to Refs. 33–35, and for the general methodology for constructing a variational formulation of a boundary value problem, we refer the reader to Refs. 8, 33, 36–37.

1. Set of admissible displacement fields

We introduce the real Hilbert space $\mathbb{H}=\{\mathbf{u}=(u_1,u_2,u_3), u_j \in L^2(\Omega)\}$ equipped with the inner product

$$(\mathbf{u},\mathbf{v})_{\mathbb{H}} = \int_\Omega \mathbf{u}(\mathbf{x})\cdot\mathbf{v}(\mathbf{x})\,d\mathbf{x}, \tag{6}$$

and the associated norm

$$\|\mathbf{u}\|_{\mathbb{H}} = (\mathbf{u},\mathbf{u})_{\mathbb{H}}^{1/2}, \tag{7}$$

in which $\mathbf{u}\cdot\mathbf{v}=\Sigma_{j=1}^{3} u_j v_j$ and where $L^2(\Omega)$ denotes the set of all the square integrable functions from Ω into $\mathbb{R}$. Let $\mathbb{V}\subset\mathbb{H}$ be the Hilbert space representing the set of admissible displacement fields with values in $\mathbb{R}^3$ such that

$$\mathbb{V}=\{\mathbf{u}\in\mathbb{H}, \quad \partial\mathbf{u}/\partial x_j \in \mathbb{H}, \quad \mathbf{u}=\mathbf{0} \text{ on } \Gamma_0\}, \tag{8}$$

equipped with the inner product

$$(\mathbf{u},\mathbf{v})_{\mathbb{V}} = (\mathbf{u},\mathbf{v})_{\mathbb{H}} + \sum_{j=1}^{3}\left(\frac{\partial\mathbf{u}}{\partial x_j},\frac{\partial\mathbf{v}}{\partial x_j}\right)_{\mathbb{H}}, \tag{9}$$

and the associated norm

$$\|\mathbf{u}\|_{\mathbb{V}} = (\mathbf{u},\mathbf{u})_{\mathbb{V}}^{1/2}. \tag{10}$$

2. Linear form representing the prescribed external forces

For all fixed t, it is assumed that prescribed external forces $\mathbf{x}\mapsto\mathbf{g}_{\text{vol}}(\mathbf{x},t)$ and $\mathbf{x}\mapsto\mathbf{g}_{\text{surf}}(\mathbf{x},t)$ are square integrable functions on Ω and Γ, respectively. For all fixed t, we introduce the linear form $\mathbf{v}\mapsto g(\mathbf{v};t)$ on $\mathbb{V}$ representing the prescribed external forces and defined by

$$g(\mathbf{v};t) = \int_\Gamma \mathbf{g}_{\text{surf}}(\mathbf{x},t)\cdot\mathbf{v}(\mathbf{x})\,ds(\mathbf{x}) + \int_\Omega \mathbf{g}_{\text{vol}}(\mathbf{x},t)\cdot\mathbf{v}(\mathbf{x})\,d\mathbf{x}, \tag{11}$$

in which ds is the surface element.

3. Variational formulation

Below, for t fixed in $[0,T]$, the mapping $\mathbf{x}\mapsto\mathbf{u}(\mathbf{x},t)$ is denoted $\mathbf{u}(t)$. The variational formulation of the mean boundary value problem defined by Eqs. (1)–(3) consists of finding a function $t\mapsto\mathbf{u}(t)$ with values in $\mathbb{V}$ such that

$$\underline{m}(\ddot{\mathbf{u}},\mathbf{v}) + \underline{d}(\dot{\mathbf{u}},\mathbf{v}) + \underline{k}(\mathbf{u},\mathbf{v}) = g(\mathbf{v};t), \quad \forall\mathbf{v}\in\mathbb{V}, \quad \forall t\in[0,T], \tag{12}$$

with the initial conditions defined by Eq. (4). Bilinear form $\underline{m}(\mathbf{u},\mathbf{v})$ (mass term) is defined by

$$\underline{m}(\mathbf{u},\mathbf{v}) = \int_\Omega \rho(\mathbf{x})\mathbf{u}(\mathbf{x})\cdot\mathbf{v}(\mathbf{x})\,d\mathbf{x}, \tag{13}$$

and is symmetric, positive definite, continuous on $\mathbb{H}\times\mathbb{H}$, and such that

$$\underline{m}(\mathbf{u},\mathbf{u}) \geqslant c_{\underline{m}}\|\mathbf{u}\|_{\mathbb{H}}^2, \tag{14}$$

in which $c_{\underline{m}}$ is a positive real constant. The properties of mechanical coefficients $b_{ijkh}(\mathbf{x})$ and $a_{ijkh}(\mathbf{x})$ are such that bilinear forms $\underline{d}(\mathbf{u},\mathbf{v})$ (damping term) and $\underline{k}(\mathbf{u},\mathbf{v})$ (stiffness term) which are defined by

$$\underline{d}(\mathbf{u},\mathbf{v}) = \int_\Omega b_{ijkh}(\mathbf{x})\varepsilon_{kh}(\mathbf{u})\varepsilon_{ij}(\mathbf{v})\,d\mathbf{x}, \tag{15}$$

$$\underline{k}(\mathbf{u},\mathbf{v}) = \int_\Omega a_{ijkh}(\mathbf{x})\varepsilon_{kh}(\mathbf{u})\varepsilon_{ij}(\mathbf{v})\,d\mathbf{x}, \tag{16}$$

are symmetric, positive definite, continuous on $\mathbb{V}\times\mathbb{V}$, and are such that

$$\underline{d}(\mathbf{u},\mathbf{u}) \geqslant c_{\underline{d}}\|\mathbf{u}\|_{\mathbb{V}}^2, \quad \underline{k}(\mathbf{u},\mathbf{u}) \geqslant c_{\underline{k}}\|\mathbf{u}\|_{\mathbb{V}}^2, \tag{17}$$

in which $c_{\underline{d}}$ and $c_{\underline{k}}$ are positive real constants. Below, Eqs. (14) and (17) will be used to prove the convergence properties of the stochastic transient response. From the usual reference books (see, for instance, Ref. 33), we deduce that the problem defined by Eq. (12) with Eq. (4) has a unique solution $t\mapsto\mathbf{u}(t)$ with values in $\mathbb{V}$, which we refer to as the transient response of the mean model of the structure submitted to impulsive loads. Note that underlined quantities refer to the mean mechanical model.

4. Response ratio and dynamic magnification factor of the mean model

The complete history of the transient response and its Fourier transform are of most interest to analyze structural-dynamic and structural-acoustic systems subjected to impulsive loads. Nevertheless, the main objective of this paper being to present a new theory for modeling random uncertainties in such systems, the presentation is limited to the calculation of the maximum response produced by the impulsive load rather than the complete history. Consequently, we introduce the usual dynamic magnification factor which is also of interest to the designer and which allows the random uncertainties modeling to be easily analyzed. It should be noted that this choice for presenting the results of the example does not suppress the generality of the theory presented.

The elastic energy at time t, associated with the transient response $\mathbf{u}(t)$ of the mean model, is equal to $\frac{1}{2}\underline{k}(\mathbf{u}(t),\mathbf{u}(t))$. The quasistatic response $\{\mathbf{x}\mapsto\mathbf{u}_{\text{stat}}(\mathbf{x},t)\}$ at time t of the mean model, also denoted $\mathbf{u}_{\text{stat}}(t)$ and associated with Eq. (12), is the unique solution of the following quasistatic problem:

$$\underline{k}(\mathbf{u}_{\text{stat}}(t),\mathbf{v}) = g(\mathbf{v};t), \quad \forall\mathbf{v}\in\mathbb{V}, \quad \forall t\in[0,T], \tag{18}$$

and its maximum $\mathbf{u}_S(\mathbf{x})$ is such that

$$\mathbf{u}_S(\mathbf{x}) = \max_{t\geqslant 0}\mathbf{u}_{\text{stat}}(\mathbf{x},t). \tag{19}$$

The maximum of the elastic energy associated with the maximum of the quasistatic response of the mean model is then equal to $\frac{1}{2}\underline{k}(\mathbf{u}_S,\mathbf{u}_S)$. Finally, the response ratio $\underline{r}(t)$ at time t, associated with the elastic energy of transient re-

sponse $\mathbf{u}(t)$ of the mean model, is defined by

$$\underline{r}(t)=\sqrt{\frac{\frac{1}{2}\underline{k}(\mathbf{u}(t),\mathbf{u}(t))}{\frac{1}{2}\underline{k}(\mathbf{u}_S,\mathbf{u}_S)}}, \tag{20}$$

and the dynamic magnification factor $\underline{b}$ of the mean model (associated with the elastic energy) can then be defined by

$$\underline{b}=\max_{t\in[0,T]}\underline{r}(t). \tag{21}$$

III. MEAN REDUCED MATRIX MODEL

A. Spectral problem associated with the mean model

We consider the following spectral problem associated with Eq. (12), corresponding to the construction of the structural modes of the associated conservative mean structural model and consisting of finding $\underline{\lambda}$ and $\boldsymbol{\varphi}$ in $\mathbb{V}$ such that

$$\underline{k}(\boldsymbol{\varphi},\mathbf{v})=\underline{\lambda}\underline{m}(\boldsymbol{\varphi},\mathbf{v}),\quad \forall\mathbf{v}\in\mathbb{V}. \tag{22}$$

Taking into account the properties of bilinear forms $\underline{m}$ and $\underline{k}$, there exists an increasing sequence of positive eigenvalues with finite multiplicity $0<\underline{\lambda}_1\leqslant\underline{\lambda}_2\cdots\leqslant\underline{\lambda}_\alpha\leqslant\cdots$. Let $\boldsymbol{\varphi}_\alpha$ be the eigenfunction associated with the eigenvalue $\underline{\lambda}_\alpha$. Then, eigenfunctions $\{\boldsymbol{\varphi}_\alpha,\alpha\geqslant 1\}$ form a complete set in $\mathbb{V}$ which means that an arbitrary function $\mathbf{u}$ belonging to $\mathbb{V}$ can be expanded as $\mathbf{u}=\Sigma_{\alpha=1}^{+\infty}q_\alpha\boldsymbol{\varphi}_\alpha$, in which $\{q_\alpha\}_\alpha$ is a sequence of real numbers. The eigenfunctions satisfy the following orthogonality conditions:

$$\underline{m}(\boldsymbol{\varphi}_\alpha,\boldsymbol{\varphi}_\beta)=\underline{\mu}_\alpha\delta_{\alpha\beta},\quad \underline{k}(\boldsymbol{\varphi}_\alpha,\boldsymbol{\varphi}_\beta)=\underline{\mu}_\alpha\underline{\omega}_\alpha^2\delta_{\alpha\beta}, \tag{23}$$

in which $\underline{\omega}_\alpha=\sqrt{\underline{\lambda}_\alpha}$ is the eigenfrequency of the mean model associated with structural mode $\boldsymbol{\varphi}_\alpha$ whose normalization $\underline{\mu}_\alpha>0$ is the mean generalized mass and where $\delta_{\alpha\alpha}=1$ and $\delta_{\alpha\beta}=0$ for $\alpha\neq\beta$. Consequently, the eigenfrequencies of the associated conservative mean structural model are such that

$$0<\underline{\omega}_1\leqslant\underline{\omega}_2\cdots\leqslant\underline{\omega}_\alpha\leqslant\cdots. \tag{24}$$

B. Derivation of the mean reduced matrix model

The mean reduced matrix model is obtained using the Ritz–Galerkin projection of the variational formulation of the mean boundary value problem on the subspace $\mathbb{V}_n$ of $\mathbb{V}$ spanned by the structural modes $\{\boldsymbol{\varphi}_1,\ldots,\boldsymbol{\varphi}_n\}$ of the mean structural model, which corresponds to the n lowest eigenfrequencies $\{\underline{\omega}_1,\ldots,\underline{\omega}_n\}$. Let $t\mapsto\mathbf{u}(t)$ from $\mathbb{R}^+$ into $\mathbb{V}$ be the unique solution of Eq. (12) with the initial conditions defined by Eq. (4), and let $\mathbf{u}_n(t)$ be the projection of $\mathbf{u}(t)$ on $\mathbb{V}_n$ (in structural dynamics, this corresponds to the usual mode-superposition method). Consequently, the mean reduced matrix model related to structural modes $\{\boldsymbol{\varphi}_1,\ldots,\boldsymbol{\varphi}_n\}$ of the mean structural model is then written as

$$\mathbf{u}_n(\mathbf{x},t)=\sum_{\alpha=1}^{n}q_\alpha^n(t)\boldsymbol{\varphi}_\alpha(\mathbf{x}), \tag{25}$$

in which the $\mathbb{R}^n$-valued vector $\mathbf{q}^n(t)=(q_1^n(t),\ldots,q_n^n(t))$ of the generalized coordinates is the solution of the time reduced problem

$$[\underline{M}_n]\ddot{\mathbf{q}}^n(t)+[\underline{D}_n]\dot{\mathbf{q}}^n(t)+[\underline{K}_n]\mathbf{q}^n(t)=\mathbf{F}^n(t),\quad t\in[0,T], \tag{26}$$

with the initial conditions

$$\mathbf{q}^n(0)=0,\quad \dot{\mathbf{q}}^n(0)=0, \tag{27}$$

in which the generalized force $\mathbf{F}^n(t)$ is the n real vector $(F_1(t),\ldots,F_n(t))$ such that

$$F_\alpha^n(t)=g(\boldsymbol{\varphi}_\alpha;t),\quad \alpha=1,...,n. \tag{28}$$

The mean generalized mass, damping, and stiffness matrices $[\underline{M}_n]$, $[\underline{D}_n]$, and $[\underline{K}_n]$ are positive-definite symmetric $(n\times n)$ real matrices such that

$$\begin{aligned}&[\underline{M}_n]_{\alpha\beta}=\underline{\mu}_\alpha\delta_{\alpha\beta},\quad [\underline{D}_n]_{\alpha\beta}=\underline{d}(\boldsymbol{\varphi}_\beta,\boldsymbol{\varphi}_\alpha),\\ &[\underline{K}_n]_{\alpha\beta}=\underline{\mu}_\alpha\underline{\omega}_\alpha^2\delta_{\alpha\beta}.\end{aligned} \tag{29}$$

C. Response ratio and dynamic magnification factor of the mean reduced matrix model

From Eqs. (23), (25), and (29), we deduce that the elastic energy $\frac{1}{2}\underline{k}(\mathbf{u}_n(t),\mathbf{u}_n(t))$ at time t, associated with transient response $\mathbf{u}_n(t)$ of the mean model, is equal to $\frac{1}{2}\langle[\underline{K}_n]\mathbf{q}^n(t),\mathbf{q}^n(t)\rangle$ in which $\langle\mathbf{y},\mathbf{x}\rangle=y_1x_1+\cdots+y_nx_n$ is the Euclidean inner product of $\mathbb{R}^n$. Consequently, the response ratio $\underline{r}_n(t)$ at time t, associated with the elastic energy of transient response $\mathbf{u}_n(t)$ (with values in $\mathbb{V}_n$) of the mean model and defined by Eq. (20), is such that

$$\underline{r}_n(t)=\sqrt{\frac{\frac{1}{2}\langle[\underline{K}_n]\mathbf{q}^n(t),\mathbf{q}^n(t)\rangle}{\frac{1}{2}\underline{k}(\mathbf{u}_S,\mathbf{u}_S)}}, \tag{30}$$

and the dynamic magnification factor $\underline{b}_n$ of the mean model (associated with the elastic energy) and defined by Eq. (21), is such that

$$\underline{b}_n=\max_{t\in[0,T]}\underline{r}_n(t). \tag{31}$$

We have $\lim_{n\to+\infty}\underline{r}_n(t)=\underline{r}(t)$ and $\lim_{n\to+\infty}\underline{b}_n=\underline{b}$ in which $\underline{r}(t)$ and $\underline{b}$ are defined by Eqs. (20) and (21), respectively.

IV. PRINCIPLE OF CONSTRUCTION OF A NONPARAMETRIC MODEL OF RANDOM UNCERTAINTIES FOR THE REDUCED MATRIX MODEL

In this section, we introduce the principle of construction of a nonparametric model of random uncertainties and we define the available information for the construction of the nonparametric model.

A. Principle of construction

It should be noted that if the boundary value problem defined by Eqs. (1)–(3) corresponds to an exact mechanical model of the structure under consideration, there are no uncertainties in the model, which is then sure. However, in continuum mechanics, the exact boundary value problem cannot be written for a given complex structure due to uncertainties in the data (geometry, boundary conditions, constitutive equation, structural complexity, etc.) and the most advanced deterministic model which it is possible to construct can only be considered as a mean mechanical model

leading to the notion of the mean boundary value problem introduced above. In addition, Eq. (12), which corresponds to the variational formulation of this mean boundary value problem, does not constitute available information for predicting the transient response $\{\mathbf{u}(\mathbf{x},t), \mathbf{x}\in\Omega, t\in[0,T]\}$ of the mean structure subjected to any impulsive loads for the following reason. For a complex structure, such a mean boundary value problem defined by Eqs. (1)–(3) is not able to predict the transient response due to impulsive loads whose energy is distributed over a very broad frequency band, i.e., over the low-, medium-, and high-frequency ranges (for instance, if there is energy in the medium-frequency range, more advanced probabilistic mechanical models such as the fuzzy structure theory have to be used to take into account the role played by the structural complexity[8]); the most that this kind of deterministic mean boundary value problem is able to predict is the transient response due to impulsive loads whose energy is mainly distributed over a broad low-frequency range for which the mean reduced matrix model defined by Eqs. (25)–(27) is suitable and allows the transient response to be predicted with good accuracy. This means that the variational formulation of the mean boundary value problem does not constitute available information for constructing the nonparametric model of random uncertainties. However, the mean reduced matrix model defined by Eqs. (25)–(27) (with n not too large) does constitute the available information for constructing the transient response of the mean model, then constructing the probability model of random uncertainties. This probabilistic model is a nonparametric model of random uncertainties because the sources of random uncertainties in the mechanical model which are due to uncertain mechanical parameters such as geometrical parameters, boundary conditions, junction stiffness, mass density, Young's modulus, etc., are not directly modeled by random variables or stochastic fields. These random uncertain geometrical and mechanical parameters mean that the generalized mass, damping, and stiffness matrices of the reduced matrix model are random matrices. In addition, it should be noted that such a nonparametric approach also allows the model uncertainties to be taken into account, whereas parametric approaches do not allow it.

Consequently, the main idea for this nonparametric construction is that not only are all the parametric random uncertainties of the mechanical model taken into account but also the model uncertainties, which lead to a random set of coupled generalized linear oscillators in the space of generalized coordinates, represented by the system of random generalized matrices of the random reduced matrix model. The problem is then to construct the probability distribution of this set of generalized oscillators, i.e., the probability distribution of the random generalized matrices of the random reduced matrix model, using only the available information. The available information is constituted of the mean generalized matrices of the mean reduced matrix model, the positive definiteness of the random generalized matrices, and the existence of second-order moments of inverses of these random generalized matrices. The nonparametric model of random uncertainties which is proposed consists of introducing a direct construction of a probabilistic model of these random generalized matrices.

B. Random reduced matrix model

In this paper, $\mathbb{M}_n(\mathbb{R})$, $\mathbb{M}_n^S(\mathbb{R})$, and $\mathbb{M}_n^+(\mathbb{R})$ are the set of all the $(n\times n)$ real matrices, the set of all the symmetric $(n\times n)$ real matrices, and the set of all the positive-definite symmetric $(n\times n)$ real matrices, respectively; we have $\mathbb{M}_n^+(\mathbb{R})\subset\mathbb{M}_n^S(\mathbb{R})\subset\mathbb{M}_n(\mathbb{R})$; all the random variables are defined on a probability space $(\mathcal{A},\mathcal{T},\mathcal{P})$.

Using the construction principle presented in Sec. IV A, the random reduced matrix model associated with the mean reduced matrix model introduced in Sec. III B is written as

$$\mathbf{U}_n(\mathbf{x},t)=\sum_{\alpha=1}^{n} Q_\alpha^n(t)\boldsymbol{\varphi}_\alpha(\mathbf{x}), \tag{32}$$

in which $\mathbf{Q}^n(t)=(Q_1^n(t),\ldots,Q_n^n(t))$ is the $\mathbb{R}^n$-valued random variable such that

$$[\mathbf{M}_n]\ddot{\mathbf{Q}}^n(t)+[\mathbf{D}_n]\dot{\mathbf{Q}}^n(t)+[\mathbf{K}_n]\mathbf{Q}^n(t)=\mathbf{F}^n(t), \quad t\geqslant 0, \tag{33}$$

with the initial conditions

$$\mathbf{Q}^n(0)=0, \quad \dot{\mathbf{Q}}^n(0)=0 \ \text{ a.s.}, \tag{34}$$

in which generalized force $\mathbf{F}^n(t)$ is the $\mathbb{R}^n$-valued vector defined by Eq. (28) and where $[\mathbf{M}_n]$, $[\mathbf{D}_n]$, and $[\mathbf{K}_n]$ are the random generalized mass, damping, and stiffness matrices, respectively, with values in $\mathbb{M}_n^+(\mathbb{R})$. It should be noted that the mathematical property related to the positiveness of the random matrices is absolutely fundamental and is required so that the second-order differential equation in time corresponds effectively to a dynamical system.

C. Random response ratio and random dynamic magnification factor

For the elastic energy of the model with random uncertainties and due to Eqs. (30)–(32), the random response ratio $R_n(t)$ at time t, associated with stochastic transient response $\mathbf{U}_n(t)$ with values in $\mathbb{V}_n$, is written as

$$R_n(t)=\sqrt{\frac{\frac{1}{2}\langle[\underline{K}_n]\mathbf{Q}^n(t),\mathbf{Q}^n(t)\rangle}{\frac{1}{2}\underline{k}(\mathbf{u}_S,\mathbf{u}_S)}}. \tag{35}$$

The random dynamic magnification factor B_n is then defined by

$$B_n=\max_{t\in[0,T]} R_n(t). \tag{36}$$

D. Available information for the construction of the nonparametric model

We have to define the available information which is useful for constructing the probabilistic model. The basic information available are the mean reduced matrix model, the positive definiteness of the random generalized matrices, and the existence of second-order moments of inverses of these random generalized matrices. The mean reduced matrix model is constituted of mean generalized mass, damping,

and stiffness matrices $[\underline{M}_n]$, $[\underline{D}_n]$, and $[\underline{K}_n]$ defined in Sec. III B and which belong to $\mathbb{M}_n^+(\mathbb{R})$. Random generalized mass, damping, and stiffness matrices $[\mathbf{M}_n]$, $[\mathbf{D}_n]$, and $[\mathbf{K}_n]$ are second-order random variables with values in $\mathbb{M}_n^+(\mathbb{R})$ (the fundamental algebraic property relative to the positive definiteness of the random generalized matrices) such that

$$E\{[\mathbf{M}_n]\}=[\underline{M}_n], \quad E\{[\mathbf{D}_n]\}=[\underline{D}_n], \quad E\{[\mathbf{K}_n]\}=[\underline{K}_n]. \tag{37}$$

In addition, in order to obtain a consistent probabilistic model and in particular, to obtain convergence properties of stochastic transient response $\{\mathbf{U}_n(\mathbf{x},t),\mathbf{x}\in\Omega,t\in[0,T]\}$ when dimension n approaches infinity, we need to introduce information relative to the existence of moments of random variables $[\mathbf{M}_n]^{-1}$, $[\mathbf{D}_n]^{-1}$, and $[\mathbf{K}_n]^{-1}$ (such as second-order moments). It should be noted that since random matrices $[\mathbf{M}_n]$, $[\mathbf{D}_n]$, and $[\mathbf{K}_n]$ are almost surely positive definite, the inverse matrices almost surely exist, but the existence of moments does not follow. We therefore introduce the following constraints:

$$E\{\|[\mathbf{M}_n]^{-1}\|_F^{\gamma}\}<+\infty, \quad E\{\|[\mathbf{D}_n]^{-1}\|_F^{\gamma}\}<+\infty, \quad E\{\|[\mathbf{K}_n]^{-1}\|_F^{\gamma}\}<+\infty, \tag{38}$$

in which $\gamma\geqslant 1$ is a positive integer and

$$\|[A]\|_F=(\mathrm{tr}\{[A][A]^T\})^{1/2}, \tag{39}$$

is the Frobenius norm of matrix $[A]\in\mathbb{M}_n(\mathbb{R})$ where tr is the trace of the matrices and $[A]^T$ is the transpose of matrix $[A]$. We then have to construct a probability model for symmetric positive-definite real random matrices $[\mathbf{M}_n]$, $[\mathbf{D}_n]$, and $[\mathbf{K}_n]$ with the available information defined by Eqs. (37) and (38). This construction is presented in Sec. V.

V. PROBABILITY MODEL FOR SYMMETRIC POSITIVE-DEFINITE REAL RANDOM MATRICES

In a part of this section, we recall the main results established in Refs. 28 and 29 concerning the construction of a probability model for random matrices with values in $\mathbb{M}_n^+(\mathbb{R})$ using the entropy optimization principle, which allows only the available information to be used. It should be noted that the results obtained and presented below differ from the known results concerning Gaussian and circular ensembles for random matrices (orthogonal, symplectic, unitary, and antisymmetric Hermitian ensembles) which have been extensively studied in the literature (see, for instance, Refs. 38–43). In another part of this section, we complete the construction given in Refs. 28 and 29 in order to obtain a consistent probabilistic model which allows the convergence properties to be studied when dimension n approaches infinity. In particular, we give an explicit expression of parameter λ_{A_n} as a function of scalar parameter δ_A which is independent of dimension n of random matrix $[\mathbf{A}_n]$ and which allows the dispersion of random matrix $[\mathbf{A}_n]$ to be given.

A. Probability density function on the space of positive-definite symmetric real matrices and characteristic function

Let $[\mathbf{A}_n]$ be a random matrix with values in $\mathbb{M}_n^+(\mathbb{R})$, defined on probability space $(\mathcal{A},\mathcal{T},\mathcal{P})$, whose probability distribution

$$P_{[\mathbf{A}_n]}=p_{[\mathbf{A}_n]}([A_n])\,\tilde{d}A_n, \tag{40}$$

is defined by a probability density function $[A_n]\to p_{[\mathbf{A}_n]}([A_n])$ from $\mathbb{M}_n^+(\mathbb{R})$ into $\mathbb{R}^+$ with respect to the measure (volume element) $\tilde{d}A_n$ on $\mathbb{M}_n^S(\mathbb{R})$ defined by[28,29]

$$\tilde{d}A_n=2^{n(n-1)/4}\Pi_{1\leqslant i\leqslant j\leqslant n}d[A_n]_{ij}. \tag{41}$$

This probability density function is such that

$$\int_{\mathbb{M}_n^+(\mathbb{R})}p_{[\mathbf{A}_n]}([A_n])\,\tilde{d}A_n=1. \tag{42}$$

For all $[\Theta_n]$ in $\mathbb{M}_n^S(\mathbb{R})$, the characteristic function of random matrix $[\mathbf{A}_n]$ with values in $\mathbb{M}_n^+(\mathbb{R})\subset\mathbb{M}_n^S(\mathbb{R})$ is defined by $\Phi_{[\mathbf{A}_n]}([\Theta_n])=E\{\exp(i\langle\langle[\Theta_n],[\mathbf{A}_n]\rangle\rangle)\}$ in which $\langle\langle[\Theta_n],[\mathbf{A}_n]\rangle\rangle=\mathrm{tr}\{[\Theta_n][\mathbf{A}_n]^T\}=\mathrm{tr}\{[\Theta_n][\mathbf{A}_n]\}$. We then have

$$\Phi_{[\mathbf{A}_n]}([\Theta_n])=\int_{\mathbb{M}_n^+(\mathbb{R})}\exp(i\,\mathrm{tr}\{[\Theta_n][A_n]\})\times p_{[\mathbf{A}_n]}([A_n])\,\tilde{d}A_n. \tag{43}$$

B. Available information for construction of the probability model

We are interested in constructing the probability distribution of a second-order random variable $[\mathbf{A}_n]$ with values in $\mathbb{M}_n^+(\mathbb{R})$ for which the available information is the mean value $[\underline{A}_n]$ of random matrix $[\mathbf{A}_n]$

$$E\{[\mathbf{A}_n]\}=\int_{\mathbb{M}_n^+(\mathbb{R})}[A_n]p_{[\mathbf{A}_n]}([A_n])\,\tilde{d}A_n=[\underline{A}_n], \tag{44}$$

in which E denotes the mathematical expectation and where mean value $[\underline{A}_n]$ is given in $\mathbb{M}_n^+(\mathbb{R})$. In addition, we assume that random matrix $[\mathbf{A}_n]$ is such that

$$E\{\ln(\det[\mathbf{A}_n])\}=v \quad \text{with} \quad |v|<+\infty. \tag{45}$$

Below, we prove that the constraint defined by Eq. (45) allows us to demonstrate the existence of moments related to the inverse random matrix $[\mathbf{A}_n]^{-1}$

$$E\{\|[\mathbf{A}_n]^{-1}\|_F^{\gamma}\}<+\infty, \tag{46}$$

in which $\gamma\geqslant 1$ is a positive integer. Consequently, from Eqs. (42), (44), and (45), we deduce that the constraints imposed for the construction of the probability model of random matrix $[\mathbf{A}_n]$ with values in $\mathbb{M}_n^+(\mathbb{R})$ are

$$\int_{\mathbb{M}_n^+(\mathbb{R})}p_{[\mathbf{A}_n]}([A_n])\,\tilde{d}A_n=1, \tag{47}$$

$$\int_{\mathbb{M}_n^+(\mathbb{R})}[A_n]p_{[\mathbf{A}_n]}([A_n])\,\tilde{d}A_n=[\underline{A}_n]\in\mathbb{M}_n^+(\mathbb{R}), \tag{48}$$

$$\int_{\mathbb{M}_n^+(\mathbb{R})} \ln(\det[A_n])\, p_{[\mathbf{A}_n]}([A_n])\, \tilde{d}A_n = v \quad \text{with} \quad |v| < +\infty. \tag{49}$$

C. Probability model using the maximum entropy principle

By introducing the measure of entropy[44] (uncertainty) and the maximum entropy principle[45,46] to construct the probability model of random matrix $[\mathbf{A}_n]$ with values in $\mathbb{M}_n^+(\mathbb{R})$ based only on the use of the available information defined by Eqs. (47)–(49), we proved that, for $\lambda_{A_n} > 0$ and $[\Theta_n] \in \mathbb{M}_n^S(\mathbb{R})$, the probability density function $p_{[\mathbf{A}_n]}([A_n])$ and characteristic function $\Phi_{[\mathbf{A}_n]}([\Theta_n])$ of positive-definite random matrix $[\mathbf{A}_n]$ are written as[28,29]

$$p_{[\mathbf{A}_n]}([A_n]) = \mathbb{1}_{\mathbb{M}_n^+(\mathbb{R})}([A_n]) \times c_{A_n} \times (\det[A_n])^{\lambda_{A_n}-1} \times \exp\left(-\frac{(n-1+2\lambda_{A_n})}{2} \operatorname{tr}\{[\underline{A}_n]^{-1}[A_n]\} \right), \tag{50}$$

$$\Phi_{[\mathbf{A}_n]}([\Theta_n]) = \left\{ \det\left([I_n] - \frac{2i}{(n-1+2\lambda_{A_n})} [\underline{A}_n] \times [\Theta_n] \right) \right\}^{-(n-1+2\lambda_{A_n})/2}, \tag{51}$$

in which det is the determinant of the matrices, $[I_n]$ is the $(n \times n)$ identity matrix, and $\mathbb{1}_{\mathbb{M}_n^+(\mathbb{R})}([A_n])$ is equal to 1 if $[A_n] \in \mathbb{M}_n^+(\mathbb{R})$ and is equal to zero if $[A_n] \notin \mathbb{M}_n^+(\mathbb{R})$. When λ_{A_n} is an integer, the probability distribution defined by Eq. (50) or (51) is a Wishart distribution.[47,48] If λ_{A_n} is not an integer, the probability distribution defined by Eq. (50) or (51) is not a Wishart distribution. In Eq. (50), positive constant c_{A_n} is written as

$$c_{A_n} = \frac{(2\pi)^{-n(n-1)/4} \left(\frac{n-1+2\lambda_{A_n}}{2} \right)^{n(n-1+2\lambda_{A_n})/2}}{\left\{ \prod_{j=1}^{n} \Gamma\left(\frac{n-j+2\lambda_{A_n}}{2} \right) \right\} (\det[\underline{A}_n])^{(n-1+2\lambda_{A_n})/2}}, \tag{52}$$

where $\Gamma(z)$ is the gamma function defined for $\Re e z > 0$ by $\Gamma(z) = \int_0^{+\infty} t^{z-1} e^{-t} dt$. The range of parameter λ_{A_n} satisfying Eq. (46) in which $\gamma \geqslant 1$ is a positive integer has to be determined. It is proved (see Appendix A) that

$$\lambda_{A_n} > \gamma + 1 \Rightarrow E\{\|[\mathbf{A}_n]^{-1}\|_F^{\gamma}\} < +\infty, \quad \gamma \geqslant 1. \tag{53}$$

In addition, it can easily be proved that

$$\lambda_{A_n} > 0 \Rightarrow E\{\|[\mathbf{A}_n]\|_F^{\eta}\} < +\infty, \quad \eta > 0. \tag{54}$$

Equation (54) means that for $\lambda_{A_n} > 0$, all the moments of random matrix $[\mathbf{A}_n]$ exist (η is any positive integer). The covariance $C_{jk,j'k'}^{A_n} = E\{([\mathbf{A}_n]_{jk} - [\underline{A}_n]_{jk})([\mathbf{A}_n]_{j'k'} - [\underline{A}_n]_{j'k'})\}$ of random variables $[\mathbf{A}_n]_{jk}$ and $[\mathbf{A}_n]_{j'k'}$ is written as

$$C_{jk,j'k'}^{A_n} = (n-1+2\lambda_{A_n})^{-1} \{ [\underline{A}_n]_{j'k} [\underline{A}_n]_{jk'} + [\underline{A}_n]_{jj'} [\underline{A}_n]_{kk'} \}, \tag{55}$$

and the variance $V_{jk}^{A_n} = C_{jk,jk}^{A_n}$ of random variable $[\mathbf{A}_n]_{jk}$ is such that $V_{jk}^{A_n} = (n-1+2\lambda_{A_n})^{-1} \times \{ [\underline{A}_n]_{jk}^2 + [\underline{A}_n]_{jj} [\underline{A}_n]_{kk} \}$.

D. Dispersion parameter δ_A of random matrix $[\mathbf{A}_n]$

Since $[\underline{A}_n]$ is a positive-definite real matrix, there is an upper triangular matrix $[\underline{L}_{A_n}]$ in $\mathbb{M}_n(\mathbb{R})$ such that

$$[\underline{A}_n] = [\underline{L}_{A_n}]^T [\underline{L}_{A_n}], \tag{56}$$

which corresponds to the Cholesky factorization of matrix $[\underline{A}_n]$. Considering Eq. (56), random matrix $[\mathbf{A}_n]$ can be written as

$$[\mathbf{A}_n] = [\underline{L}_{A_n}]^T [\mathbf{G}_{A_n}] [\underline{L}_{A_n}], \tag{57}$$

in which matrix $[\mathbf{G}_{A_n}]$ is a random variable with values in $\mathbb{M}_n^+(\mathbb{R})$. From Eqs. (44) and (57), we deduce that the mean value $[\underline{G}_{A_n}]$ of random matrix $[\mathbf{G}_{A_n}]$ is such that

$$[\underline{G}_{A_n}] = E\{[\mathbf{G}_{A_n}]\} = [I_n]. \tag{58}$$

The probability density function $p_{[\mathbf{G}_{A_n}]}([G_n])$ with respect to measure $\tilde{d}G_n$ on $\mathbb{M}_n^S(\mathbb{R})$ of random matrix $[\mathbf{G}_{A_n}]$ with values in $\mathbb{M}_n^+(\mathbb{R})$ is given by Eqs. (50) and (52) in which $[\underline{A}_n]$ has to be replaced by $[I_n]$. We then have

$$p_{[\mathbf{G}_{A_n}]}([G_n]) = \mathbb{1}_{\mathbb{M}_n^+(\mathbb{R})}([G_n]) \times C_{G_{A_n}} \times (\det[G_n])^{\lambda_{A_n}-1} \times \exp\left(-\frac{(n-1+2\lambda_{A_n})}{2} \operatorname{tr}[G_n] \right), \tag{59}$$

in which positive constant $C_{G_{A_n}}$ is such that

$$C_{G_{A_n}} = \frac{(2\pi)^{-n(n-1)/4} \left(\frac{n-1+2\lambda_{A_n}}{2} \right)^{n(n-1+2\lambda_{A_n})/2}}{\left\{ \prod_{j=1}^{n} \Gamma\left(\frac{n-j+2\lambda_{A_n}}{2} \right) \right\}}. \tag{60}$$

From Eqs. (55) and (57), we deduce that the covariance $C_{jk,j'k'}^{G_n}$ of random variables $[\mathbf{G}_{A_n}]_{jk}$ and $[\mathbf{G}_{A_n}]_{j'k'}$, defined by $C_{jk,j'k'}^{G_n} = E\{([\mathbf{G}_{A_n}]_{jk} - [\underline{G}_{A_n}]_{jk})([\mathbf{G}_{A_n}]_{j'k'} - [\underline{G}_{A_n}]_{j'k'})\}$, is written as

$$C_{jk,j'k'}^{G_n} = (n-1+2\lambda_{A_n})^{-1} \{ [\underline{G}_{A_n}]_{j'k} [\underline{G}_{A_n}]_{jk'} + [\underline{G}_{A_n}]_{jj'} [\underline{G}_{A_n}]_{kk'} \}. \tag{61}$$

Since $[\underline{G}_{A_n}] = [I_n]$, the variance $V_{jk}^{G_n} = C_{jk,jk}^{G_n}$ of random variable $[\mathbf{G}_{A_n}]_{jk}$ is such that

$$V_{jk}^{G_n} = (n-1+2\lambda_{A_n})^{-1} (1+\delta_{jk}), \tag{62}$$

in which $\delta_{jk} = 0$ if $j \neq k$ and $\delta_{jj} = 1$. Let $\delta_A > 0$ be defined by

$$\delta_A=\left\{\frac{E\{\|[\mathbf{G}_{A_n}]-[G_{A_n}]\|_F^2\}}{\|[G_{A_n}]\|_F^2}\right\}^{1/2}. \tag{63}$$

Equation (62) yields $E\{\|[\mathbf{G}_{A_n}]-[G_{A_n}]\|_F^2\}=\Sigma_j\Sigma_k V_{jk}^{G_n}=n(n+1)(n-1+2\lambda_{A_n})^{-1}$ and since $\|[G_{A_n}]\|_F^2=\|[I_n]\|_F^2=n$, we deduce that

$$\delta_A=\left\{\frac{n+1}{n-1+2\lambda_{A_n}}\right\}^{1/2}, \tag{64}$$

and consequently,

$$\lambda_{A_n}=l_A(n), \tag{65}$$

in which $n\to l_A(n)$ is the mapping defined on the set $\mathbb{N}^*$ of all positive integers such that

$$l_A(n)=\frac{1-\delta_A^2}{2\,\delta_A^2}n+\frac{1+\delta_A^2}{2\,\delta_A^2}. \tag{66}$$

From Eqs. (53) and (64), we deduce that parameter δ_A has to be such that

$$0<\delta_A<\sqrt{\frac{n+1}{n+1+2\,\gamma}}<1, \quad \gamma\geqslant 1, \quad \forall n\geqslant 1. \tag{67}$$

Equation (67) shows that γ has to be chosen as small as possible in order to increase the domain of possible values for δ_A. From convergence considerations when $n\to+\infty$ (see Sec. VII) and from Eq. (67), we deduce that $\gamma=2$ is the optimal value. Let $n_0\geqslant 1$ be a fixed integer. Taking the value $\gamma=2$, we then deduce that, if parameter δ_A satisfies

$$0<\delta_A<\sqrt{\frac{n_0+1}{n_0+5}}, \tag{68}$$

then $\forall n\geqslant n_0$, we have $\lambda_{A_n}=l_A(n)>\gamma+1=3$, and consequently Eq. (53) holds. These equations will be used as follows. The lower bound n_0 of positive integer n is fixed. Then, the dispersion of the probability model is fixed by giving parameter δ_A, independent of n, a value such that Eq. (68) is satisfied. For each value of integer $n\geqslant n_0$, parameter $\lambda_{A_n}=l_A(n)$ is then calculated by using Eq. (66). Consequently, $\lambda_{A_n}=l_A(n)$ appears as a function of n.

E. Algebraic representation of random matrix $[\mathbf{A}_n]$ when λ_{A_n} is an integer

When $\lambda_{A_n}=l_A(n)$ is a positive integer, we introduce the positive integer m_A such that

$$m_A(n)=n-1+2l_A(n). \tag{69}$$

Substituting Eq. (66) in the right-hand side of Eq. (69) yields

$$m_A(n)=(n+1)/\delta_A^2. \tag{70}$$

Since $m_A(n)$ is a positive integer, it can be verified that the probability distribution defined by Eq. (50) or (51) is a Wishart distribution[47,48] and that the random matrix $[\mathbf{A}_n]$ can be written as[28,29]

$$[\mathbf{A}_n]=\frac{1}{m_A(n)}\sum_{j=1}^{m_A(n)}([\underline{L}_{A_n}]^T\mathbf{X}_j)([\underline{L}_{A_n}]^T\mathbf{X}_j)^T, \tag{71}$$

in which $[\underline{L}_{A_n}]$ is the upper triangular matrix defined by Eq. (56) and where $\mathbf{X}_1,\ldots,\mathbf{X}_{m_A(n)}$ are independent random vectors, each vector $\mathbf{X}_j$ being an $\mathbb{R}^n$-valued second-order Gaussian random variable, centered, and whose covariance matrix is $[C_{\mathbf{X}_j}]=E\{\mathbf{X}_j\mathbf{X}_j^T\}=[I_n]$. Consequently, Eq. (71) gives an efficient procedure for algebraic calculations and the Monte Carlo numerical simulation of random matrix $[\mathbf{A}_n]$.

F. Algebraic representation of random matrix $[\mathbf{A}_n]$ when λ_{A_n} is not an integer

Let us now assume that $\lambda_{A_n}=l_A(n)$, given by Eq. (65), is a positive real number (the particular case for which λ_{A_n} is a positive integer is presented above in Sec. V E). Since $[\mathbf{G}_{A_n}]$ defined by Eq. (57) is a random matrix with values in $\mathbb{M}_n^+(\mathbb{R})$, the Cholesky factorization allows us to write

$$[\mathbf{G}_{A_n}]=[\mathbf{L}_{A_n}]^T[\mathbf{L}_{A_n}] \quad \text{a.s.}, \tag{72}$$

in which $[\mathbf{L}_{A_n}]$ is an upper triangular random matrix with values in $\mathbb{M}_n(\mathbb{R})$. The following results, which allow a procedure for the Monte Carlo simulation of random matrix $[\mathbf{A}_n]$ to be defined, are proved.[28,29]

(1) Random variables $\{[\mathbf{L}_{A_n}]_{jj'}, j\leqslant j'\}$ are independent.

(2) For $j<j'$, real-valued random variable $[\mathbf{L}_{A_n}]_{jj'}$ can be written as $[\mathbf{L}_{A_n}]_{jj'}=2^{-1/2}[\widetilde{\mathbf{L}}_{A_n}]_{jj'}$, in which $[\widetilde{\mathbf{L}}_{A_n}]_{jj'}$ is a real-valued Gaussian random variable with zero mean and variance given by

$$v=2(n-1+2l_A(n))^{-1}. \tag{73}$$

(3) For $j=j'$, positive-valued random variable $[\mathbf{L}_{A_n}]_{jj}$ can be written as $[\mathbf{L}_{A_n}]_{jj}=\sqrt{v\,\mathbf{Y}_j}$ in which v is given by Eq. (73) and where $\mathbf{Y}_j$ is a positive-valued gamma random variable whose probability density function with respect to dy is given by

$$\Gamma_j(y)=\frac{\mathbb{1}_{[0,+\infty[}(y)}{\Gamma\left(\dfrac{n-j+2l_A(n)}{2}\right)}y^{[(n-j+2l_A(n))/2]-1}e^{-y}. \tag{74}$$

(4) We have $[\mathbf{G}_{A_n}]=[\mathbf{L}_{A_n}]^T[\mathbf{L}_{A_n}]$ and $[\mathbf{A}_n]=[\underline{L}_{A_n}]^T[\mathbf{G}_{A_n}]\times[\underline{L}_{A_n}]$.

G. Probability model of a set of positive-definite symmetric real random matrices

Let us consider ν random matrices $[\mathbf{A}_n^1],\ldots,[\mathbf{A}_n^\nu]$ with values in $\mathbb{M}_n^+(\mathbb{R})$ such that for each j in $\{1,\ldots,\nu\}$, the probability density function of random matrix $[\mathbf{A}_n^j]$ satisfies Eqs. (47)–(49). This means that only the mean values of the random matrices are known. Applying the maximum entropy principle, it can be proved that the probability density function $([A_n^1],\ldots,[A_n^\nu])\mapsto p_{[\mathbf{A}_n^1],\ldots,[\mathbf{A}_n^\nu]}([A_n^1],\ldots,[A_n^\nu])$ from $\mathbb{M}_n^+(\mathbb{R})\times\cdots\times\mathbb{M}_n^+(\mathbb{R})$ into $\mathbb{R}^+$ with respect to the measure (volume element) $\widetilde{d}A_n^1\times\cdots\times\widetilde{d}A_n^\nu$ on $\mathbb{M}_n^S(\mathbb{R})\times\cdots\times\mathbb{M}_n^S(\mathbb{R})$ is written as

$$p_{[\mathbf{A}_n^1],\ldots,[\mathbf{A}_n^\nu]}([A_n^1],\ldots,[A_n^\nu]) = p_{[\mathbf{A}_n^1]}([A_n^1]) \times \cdots \times p_{[\mathbf{A}_n^\nu]}([A_n^\nu]), \tag{75}$$

which means that $[\mathbf{A}_n^1],\ldots,[\mathbf{A}_n^\nu]$ are independent random matrices.

VI. NONPARAMETRIC MODEL OF RANDOM UNCERTAINTIES

In this section we complete the construction of the probability model introduced in Sec. IV using the developments of Sec. V.

A. Probability model of the reduced matrix model

Let $n_0 \geq 1$ be a fixed integer and $n \geq n_0$. We apply the results of Sec. V to the set of positive-definite symmetric real random matrices $\{[\mathbf{M}_n],[\mathbf{D}_n],[\mathbf{K}_n]\}$ defined in Sec. IV, for which the available information is described by Eqs. (37)–(38). As indicated in Sec. V D, we take $\gamma=2$ and the levels of dispersion of random matrices $[\mathbf{M}_n]$, $[\mathbf{D}_n]$, and $[\mathbf{K}_n]$ are controlled by parameters δ_M, δ_D, and δ_K, respectively, which are independent of n and are chosen such that [see Eq. (68)]

$$0 < \delta_M,\ \delta_D,\ \delta_K < \sqrt{\frac{n_0+1}{n_0+5}}. \tag{76}$$

Parameters λ_M, λ_D, and λ_K are defined by Eq. (65)

$$\lambda_M = l_M(n), \quad \lambda_D = l_D(n), \quad \lambda_K = l_K(n), \tag{77}$$

in which $l_M(n)$, $l_D(n)$, and $l_K(n)$ are given by Eq. (66)

$$l_M(n) = \frac{1-\delta_M^2}{2\delta_M^2} n + \frac{1+\delta_M^2}{2\delta_M^2}, \tag{78}$$

$$l_D(n) = \frac{1-\delta_D^2}{2\delta_D^2} n + \frac{1+\delta_D^2}{2\delta_D^2}, \tag{79}$$

$$l_K(n) = \frac{1-\delta_K^2}{2\delta_K^2} n + \frac{1+\delta_K^2}{2\delta_K^2}. \tag{80}$$

From Sec. V G, we deduce that random matrices $[\mathbf{M}_n]$, $[\mathbf{D}_n]$, and $[\mathbf{K}_n]$ are independent random variables with values in $\mathbb{M}_n^+(\mathbb{R})$ and their probability density functions $p_{[\mathbf{M}_n]}([M_n])$, $p_{[\mathbf{D}_n]}([D_n])$, and $p_{[\mathbf{K}_n]}([K_n])$ with respect to the measures (volume elements) $\tilde{d}M_n$, $\tilde{d}D_n$, and $\tilde{d}K_n$ on $\mathbb{M}_n^S(\mathbb{R})$ are given by Eqs. (50) and (52), and their characteristic functions by Eq. (51).

B. Construction of the stochastic transient response

For fixed positive integer $n \geq n_0$, we have to construct stochastic processes $\{\mathbf{U}_n(\mathbf{x},t), \mathbf{x}\in\Omega, t\in[0,T]\}$ defined by Eqs. (32)–(34) and $\{R_n(t), t\geq 0\}$ defined by Eq. (35), and random variable B_n defined by Eq. (36). Below, we present a formulation which is adapted to Monte Carlo numerical simulation.

For given matrices $[M_n]$, $[D_n]$, $[K_n]$ in $\mathbb{M}_n^+(\mathbb{R})$, let $t \mapsto \mathbf{q}_\mathbf{F}^n(t;[M_n],[D_n],[K_n])$ be the solution from $\mathbb{R}^+$ into $\mathbb{R}^n$ of the deterministic second-order differential equation

$$[M_n]\ddot{\mathbf{q}}_\mathbf{F}^n(t) + [D_n]\dot{\mathbf{q}}_\mathbf{F}^n(t) + [K_n]\mathbf{q}_\mathbf{F}^n(t) = \mathbf{F}^n(t), \quad t \geq 0, \tag{81}$$

with the initial conditions

$$\mathbf{q}_\mathbf{F}^n(0) = 0, \quad \dot{\mathbf{q}}_\mathbf{F}^n(0) = 0. \tag{82}$$

We deduce that stochastic process $\{\mathbf{Q}^n(t), t\geq 0\}$ which is the solution of the stochastic dynamical problem defined by Eqs. (33)–(34), can be written as

$$\mathbf{Q}^n(t) = \mathbf{q}_\mathbf{F}^n(t;[\mathbf{M}_n],[\mathbf{D}_n],[\mathbf{K}_n]). \tag{83}$$

It should be noted that $\mathbf{Q}^n(t)$ can usually be written as

$$\mathbf{Q}^n(t) = \int_0^t [h_n(t-\tau)]\mathbf{F}^n(\tau)\,d\tau, \tag{84}$$

in which $t \mapsto [h_n(t)]$ is the matrix-valued impulse response function of the linear filter associated with second-order differential Eq. (81). If $\mathbf{q} \mapsto [\theta_n(\mathbf{q})]$ is a mapping from $\mathbb{R}^n$ into the set $\mathbb{M}_{\nu_1,\nu_2}(\mathbb{R})$ of all the $(\nu_1 \times \nu_2)$ real matrices, we have

$$E\{[\theta_n(\mathbf{Q}^n(t))]\} = \int_{\mathbb{M}_n^+(\mathbb{R})} \int_{\mathbb{M}_n^+(\mathbb{R})} \int_{\mathbb{M}_n^+(\mathbb{R})} [\theta_n(\mathbf{q}_\mathbf{F}^n(t;[M_n],[D_n],[K_n]))] \times p_{[\mathbf{M}_n]}([M_n]) \times p_{[\mathbf{D}_n]}([D_n]) \times p_{[\mathbf{K}_n]}([K_n])\,\tilde{d}M_n\,\tilde{d}D_n\,\tilde{d}K_n. \tag{85}$$

For instance, $R_n(t)$ defined by Eq. (35) can be written as $R_n(t) = [\theta_n(\mathbf{Q}^n(t))]$ with $\nu_1 = \nu_2 = 1$. Calculation of the stochastic transient response of the dynamical system with random uncertainties requires the numerical construction of mapping $t \mapsto \mathbf{q}_\mathbf{F}^n(t;[M_n],[D_n],[K_n])$ as the solution of the deterministic Eqs. (81)–(82). Since matrices $[M_n]$, $[D_n]$, and $[K_n]$ are full matrices (not diagonal) as samplings of random matrices $[\mathbf{M}_n]$, $[\mathbf{D}_n]$, and $[\mathbf{K}_n]$, Eq. (84) is not used but second-order differential Eq. (81) is solved directly using an unconditionally stable implicit step-by-step integration method (such as the Newmark integration scheme[2]) with initial conditions defined by Eq. (82). In addition, we have to calculate multiple integrals in a higher dimension [see Eq. (85)] for which a well-suited method consists of using a Monte Carlo calculation with or without variance reduction procedures.[49–55] This method is very efficient if there is a Monte Carlo simulation procedure for random matrices $[\mathbf{M}_n]$, $[\mathbf{D}_n]$, and $[\mathbf{K}_n]$ which is the case of the method presented in Secs. V E and V F. It should be noted that for many applications, integer n is sufficiently high that λ_M, λ_D, and λ_K can be considered as positive integers without introducing any significant limitation in the model. Applying Eqs. (70)–(71) to random matrices $[\mathbf{M}_n]$, $[\mathbf{D}_n]$, and $[\mathbf{K}_n]$ yields

$$[\mathbf{M}_n] = \frac{1}{m_M(n)} \sum_{j=1}^{m_M(n)} ([\underline{L}_{M_n}]^T \mathbf{X}_j)([\underline{L}_{M_n}]^T \mathbf{X}_j)^T, \tag{86}$$

$$[\mathbf{D}_n] = \frac{1}{m_D(n)} \sum_{j=1}^{m_D(n)} ([\underline{L}_{D_n}]^T \mathbf{Y}_j)([\underline{L}_{D_n}]^T \mathbf{Y}_j)^T, \tag{87}$$

$$[\mathbf{K}_n] = \frac{1}{m_K(n)} \sum_{j=1}^{m_K(n)} ([\underline{L}_{K_n}]^T \mathbf{Z}_j)([\underline{L}_{K_n}]^T \mathbf{Z}_j)^T, \tag{88}$$

in which

$$m_M(n)=\mathrm{Fix}\left(\frac{n+1}{\delta_M^2}\right),\quad m_D(n)=\mathrm{Fix}\left(\frac{n+1}{\delta_D^2}\right),$$
$$m_K(n)=\mathrm{Fix}\left(\frac{n+1}{\delta_K^2}\right), \tag{89}$$

where $\mathrm{Fix}(x)$ is equal to x when x is an integer and $\mathrm{Fix}(x)$ rounds down $x+1$ to the nearest integer when x is not an integer. In Eqs. (86)–(88), $[\underline{L}_{M_n}]$, $[\underline{L}_{D_n}]$, and $[\underline{L}_{K_n}]$ are upper triangular matrices in $\mathbb{M}_n(\mathbb{R})$ corresponding to the Cholesky factorization of symmetric positive-definite matrices $[\underline{M}_n]$, $[\underline{D}_n]$, and $[\underline{K}_n]$

$$[\underline{M}_n]=[\underline{L}_{M_n}]^T[\underline{L}_{M_n}],\quad [\underline{D}_n]=[\underline{L}_{D_n}]^T[\underline{L}_{D_n}],$$
$$[\underline{K}_n]=[\underline{L}_{K_n}]^T[\underline{L}_{K_n}]. \tag{90}$$

The set of all the components of vectors $\mathbf{X}_1,\ldots,\mathbf{X}_{m_M(n)}$, $\mathbf{Y}_1,\ldots,\mathbf{Y}_{m_D(n)}$, and $\mathbf{Z}_1,\ldots,\mathbf{Z}_{m_K(n)}$ with values in $\mathbb{R}^n$ is constituted of $m_M(n)\times n+m_D(n)\times n+m_K(n)\times n$ independent random variables, each of which is a real-valued second-order normalized Gaussian random variable (zero mean value and unit variance).

VII. CONVERGENCE PROPERTIES AS THE DIMENSION APPROACHES INFINITY

For each $n\geq n_0$ fixed, stochastic transient response $\{\mathbf{U}_n(\mathbf{x},t),\mathbf{x}\in\Omega,t\in[0,T]\}$ of the dynamical system with random uncertainties can be constructed using Secs. IV to VI. A major problem concerns the convergence properties of stochastic transient response $\{\mathbf{U}_n(\mathbf{x},t),\mathbf{x}\in\Omega,t\in[0,T]\}$ and related quantities as $n\to+\infty$ for the nonparametric probabilistic model proposed in Secs. IV to VI. This problem is studied below.

A. Introduction of norms useful for the convergence properties

As above, all the random variables are defined on probability space $(\mathcal{A},\mathcal{T},\mathcal{P})$. Let $\mathbf{Q}=(Q_1,\ldots,Q_n)$ be an $\mathbb{R}^n$-valued random variable. The norm $|||\mathbf{Q}|||$ of $\mathbf{Q}$ is defined by

$$|||\mathbf{Q}|||=\sqrt{E\{\|\mathbf{Q}\|^2\}}, \tag{91}$$

in which $\|\mathbf{Q}\|=(\Sigma_{\alpha=1}^n Q_\alpha^2)^{1/2}$ is the Euclidean norm of $\mathbb{R}^n$ and where E is the mathematical expectation. It should be noted that vector $\mathbf{Q}$ is a second-order random variable if $|||\mathbf{Q}|||<+\infty$. Let $\mathbf{U}$ be an $\mathbb{H}$-valued random variable (see Sec. II B). Its norm in $\mathbb{H}$ is $\|\mathbf{U}\|_{\mathbb{H}}$ and is a positive-valued random variable defined [see Eq. (7)] by

$$\|\mathbf{U}\|_{\mathbb{H}}=\left(\int_\Omega \mathbf{U}(\mathbf{x})\cdot\mathbf{U}(\mathbf{x})d\mathbf{x}\right)^{1/2}.$$

The norm $|||\mathbf{U}|||_{\mathbb{H}}$ of $\mathbf{U}$ is defined by

$$|||\mathbf{U}|||_{\mathbb{H}}=\sqrt{E\{\|\mathbf{U}\|_{\mathbb{H}}^2\}}. \tag{92}$$

Similarly, if $\mathbf{U}$ is a $\mathbb{V}$-valued random variable (see Sec. II B), its norm $\|\mathbf{U}\|_{\mathbb{V}}$ is the positive-valued random variable defined [see Eqs. (9)–(10)] by

$$\|\mathbf{U}\|_{\mathbb{V}}=\left(\|\mathbf{U}\|_{\mathbb{H}}^2+\sum_{j=1}^3\left\|\frac{\partial\mathbf{U}}{\partial x_j}\right\|_{\mathbb{H}}^2\right)^{1/2},$$

and the norm $|||\mathbf{U}|||_{\mathbb{V}}$ of $\mathbf{U}$ is defined by

$$|||\mathbf{U}|||_{\mathbb{V}}=\sqrt{E\{\|\mathbf{U}\|_{\mathbb{V}}^2\}}. \tag{93}$$

B. Prerequisite to the construction of basic inequalities

Below, the $\mathbb{M}_n^+(\mathbb{R})$-valued random matrix $[\mathbf{A}_n]$ denotes $\mathbb{M}_n^+(\mathbb{R})$-valued random matrices $[\mathbf{M}_n]$, $[\mathbf{D}_n]$, or $[\mathbf{K}_n]$ defined on probability space $(\mathcal{A},\mathcal{T},\mathcal{P})$ and introduced in Secs. IV and VI. Let $[\mathbf{G}_{A_n}]$ be the $\mathbb{M}_n^+(\mathbb{R})$-valued random matrix on probability space $(\mathcal{A},\mathcal{T},\mathcal{P})$ defined by Eq. (57), whose probability density function $p_{[\mathbf{G}_{A_n}]}([G_n])$ is given by Eq. (59). For ω fixed in $\mathcal{A}$, the norm of matrix $[\mathbf{G}_{A_n}(\omega)]^{-1}$ induced by the Euclidean norm of $\mathbb{R}^n$ is defined by

$$\|[\mathbf{G}_{A_n}(\omega)]^{-1}\|=\sup_{\mathbf{q}\in\mathbb{R}^n,\|\mathbf{q}\|=1}\|[\mathbf{G}_{A_n}(\omega)]^{-1}\mathbf{q}\|, \tag{94}$$

and can be written as

$$\|[\mathbf{G}_{A_n}(\omega)]^{-1}\|=\frac{1}{\widetilde{\Sigma}_1(\omega)}, \tag{95}$$

in which $\widetilde{\Sigma}_1(\omega)>0$ is the smallest eigenvalue of matrix $[\mathbf{G}_{A_n}(\omega)]\in\mathbb{M}_n^+(\mathbb{R})$ whose eigenvalues are such that $0<\widetilde{\Sigma}_1(\omega)\leq\widetilde{\Sigma}_2(\omega)\leq\cdots\leq\widetilde{\Sigma}_n(\omega)$. It should be noted that $\|[\mathbf{G}_{A_n}(\omega)]^{-1}\|\leq\|[\mathbf{G}_{A_n}(\omega)]^{-1}\|_F\leq\sqrt{n}\|[\mathbf{G}_{A_n}(\omega)]^{-1}\|$ in which the Frobenius norm is defined by Eq. (39). We then have

$$E\{\|[\mathbf{G}_{A_n}]^{-1}\|^2\}=E\{\widetilde{\Sigma}_1^{-2}\}. \tag{96}$$

In Appendix B, the following inequality is proved:

$$\forall n\geq n_0,\quad E\{\|[\mathbf{G}_{A_n}]^{-1}\|^2\}\leq C_{\delta_A}<+\infty, \tag{97}$$

in which $n_0\geq 1$ is the fixed integer defined in Sec. VI A and where C_{δ_A} is a positive finite constant which is independent of n but which depends on δ_A defined by Eq. (76). Equation (97) means that $n\mapsto E\{\|[\mathbf{G}_{A_n}]^{-1}\|^2\}$ is a bounded function from $\{n\geq n_0\}$ into $\mathbb{R}^+$. We have numerically verified Eq. (97) using a Monte Carlo numerical simulation based on Eqs. (95)–(96), Sec. V E and the usual estimator of the second-order moment of random variable $1/\widetilde{\Sigma}_1$. Figure 1 shows the graph of function $n\mapsto E\{\|[\mathbf{G}_{A_n}]^{-1}\|^2\}$ for $n\geq n_0=2$, $\delta_A=0.1$, 0.3, and 0.5, and corresponds to 100 samples in the Monte Carlo numerical simulation. These numerical results confirm Eq. (97), which is mathematically proved.

C. Basic inequalities derived from the random energy equation

Let us assume that prescribed external forces $(\mathbf{x},t)\mapsto\mathbf{g}_{\mathrm{vol}}(\mathbf{x},t)$ and $(\mathbf{x},t)\mapsto\mathbf{g}_{\mathrm{surf}}(\mathbf{x},t)$ are such that

$$\int_0^T\|\mathbf{f}(\tau)\|_{\mathbb{V}}^2\,d\tau<+\infty, \tag{98}$$

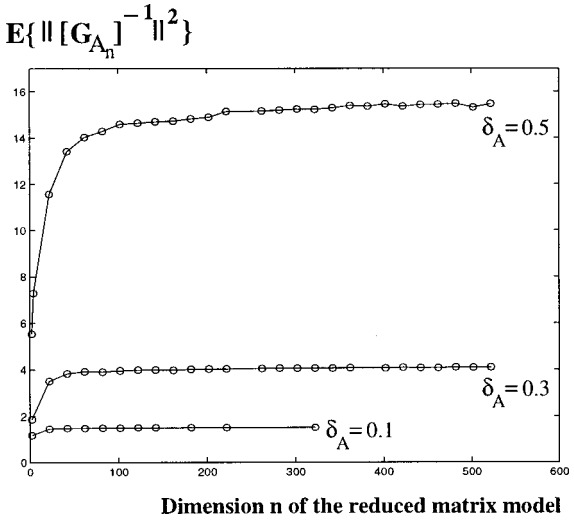


FIG. 1. Graph of function $n \mapsto E\{\|[\mathbf{G}_{A_n}]^{-1}\|^2\}$ for $\delta_A = 0.1$, 0.3, and 0.5.

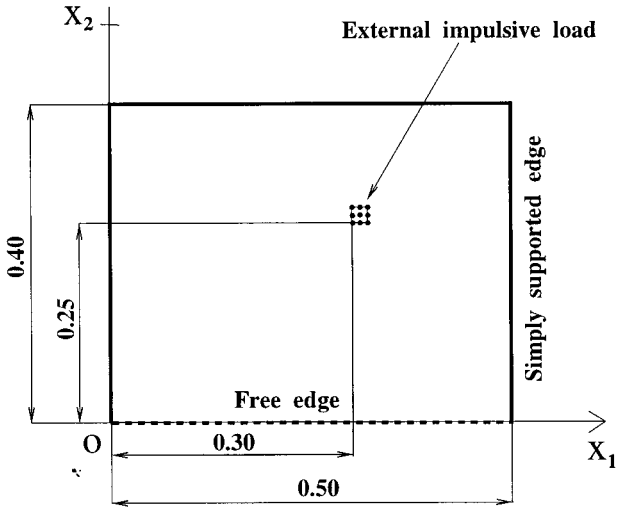


FIG. 2. Geometry of the mean structure.

in which $\mathbf{f}(\tau)$ and $\|\mathbf{f}(\tau)\|_{\mathbb{V}'}$ are defined by Eqs. (D3) and (D2) in Appendix D. From Appendix E, we deduce that

$$\forall n \geq n_0, \quad \forall t \in [0,T], \quad |||\mathbf{U}_n(t)|||_{\mathbb{V}}^2 \leq C_1 < +\infty, \tag{99}$$

$$\forall n \geq n_0, \quad \forall t \in [0,T], \quad |||\dot{\mathbf{U}}_n(t)|||_{\mathbb{H}}^2 \leq C_2 < +\infty, \tag{100}$$

in which C_1 and C_2 are positive constants which are independent of n and t but which depend on the prescribed external forces, parameters T, δ_M, δ_D, δ_K [see Eq. (76)], and which are written [see Eqs. (E22) and (E21)] as

$$C_1 = \left(\frac{C_{\delta_K}}{c_{\underline{k}}^2} + \frac{C_{\delta_D}}{c_{\underline{d}}^2} \right) \int_0^T \|\mathbf{f}(\tau)\|_{\mathbb{V}'}^2 \, d\tau, \tag{101}$$

$$C_2 = \left(\frac{C_{\delta_M}}{c_{\underline{m}}^2} + \frac{C_{\delta_D}}{c_{\underline{d}}^2} \right) \int_0^T \|\mathbf{f}(\tau)\|_{\mathbb{V}'}^2 \, d\tau, \tag{102}$$

where positive finite constants $c_{\underline{m}}$, $c_{\underline{d}}$, and $c_{\underline{k}}$ are defined by Eqs. (14) and (17) and where positive finite constants C_{δ_M}, C_{δ_D}, and C_{δ_K} are defined by Eq. (97).

Let $\mathbb{X}_{\mathbb{V}}$ and $\mathbb{X}_{\mathbb{H}}$ be the vector spaces of all the second-order random variables defined on probability space $(\mathcal{A},\mathcal{T},\mathcal{P})$ with values in $\mathbb{V}$ and $\mathbb{H}$, respectively. If random variables $\mathbf{U}=\{\mathbf{U}(\mathbf{x}),\mathbf{x}\in\Omega\}$ and $\mathbf{W}=\{\mathbf{W}(\mathbf{x}),\mathbf{x}\in\Omega\}$ belong to $\mathbb{X}_{\mathbb{V}}$ and $\mathbb{X}_{\mathbb{H}}$, respectively, then $|||\mathbf{U}|||_{\mathbb{V}}^2 = E\{\|\mathbf{U}\|_{\mathbb{V}}^2\} < +\infty$ and $|||\mathbf{W}|||_{\mathbb{H}}^2 = E\{\|\mathbf{W}\|_{\mathbb{H}}^2\} < +\infty$. Denoting $\{\mathbf{U}_n(\mathbf{x},t),\mathbf{x}\in\Omega\}$ by $\mathbf{U}_n(t)$, and since $[0,T]$ is a bounded interval, the inequality defined by Eq. (99) means that, for all fixed $n \geq n_0$, function $t \mapsto \mathbf{U}_n(t)$ belongs to the set $L^2([0,T],\mathbb{X}_{\mathbb{V}})$ of all the square integrable functions from $[0,T]$ into $\mathbb{X}_{\mathbb{V}}$ and that the sequence of functions $\{t \mapsto \mathbf{U}_n(t)\}_{n \geq n_0}$ belongs to a bounded set of $L^2([0,T],\mathbb{X}_{\mathbb{V}})$. Similarly, denoting $\{\dot{\mathbf{U}}_n(\mathbf{x},t),\mathbf{x}\in\Omega\}$ as $\dot{\mathbf{U}}_n(t)$, the inequality defined by Eq. (100) means that for all fixed $n \geq n_0$, function $t \mapsto \dot{\mathbf{U}}_n(t)$ belongs to the set $L^2([0,T],\mathbb{X}_{\mathbb{H}})$ of all the square integrable functions from $[0,T]$ into $\mathbb{X}_{\mathbb{H}}$ and that the sequence of functions $\{t \mapsto \dot{\mathbf{U}}_n(t)\}_{n \geq n_0}$ belongs to a bounded set of $L^2([0,T],\mathbb{X}_{\mathbb{H}})$. It should be noted that the above results hold if spaces $L^2([0,T],\mathbb{X}_{\mathbb{V}})$ and $L^2([0,T],\mathbb{X}_{\mathbb{H}})$ are replaced by the sets of all the bounded functions from $[0,T]$ into $\mathbb{X}_{\mathbb{V}}$ and $\mathbb{X}_{\mathbb{H}}$, respectively.

D. Convergence as dimension *n* approaches infinity

Since $\{t \mapsto \mathbf{U}_n(t)\}_n$ and $\{t \mapsto \dot{\mathbf{U}}_n(t)\}_n$ are bounded sequences in $L^2([0,T],\mathbb{X}_{\mathbb{V}})$ and $L^2([0,T],\mathbb{X}_{\mathbb{H}})$, respectively (see Sec. VII C), from sequences $\{t \mapsto \mathbf{U}_n(t)\}_n$ and $\{t \mapsto \dot{\mathbf{U}}_n(t)\}_n$ can be extracted[33] subsequences $\{t \mapsto \mathbf{U}_{n_k}(t)\}_k$ and $\{t \mapsto \dot{\mathbf{U}}_{n_k}(t)\}_k$, respectively, which are weakly convergent in $L^2([0,T],\mathbb{X}_{\mathbb{V}})$ and $L^2([0,T],\mathbb{X}_{\mathbb{H}})$, respectively, as k approaches infinity.

VIII. EXAMPLE

A. Definition of the mean model

The mean structure is constituted of a rectangular homogeneous and isotropic plate located in the plane (Ox_1, Ox_2) of a Cartesian coordinate system $(Ox_1x_2x_3)$, in bending mode (the out-of-plane displacement is x_3), with constant thickness 4×10^{-4} m, width 0.40 m, length 0.50 m, mass density 7800 kg/m^3, Young's modulus 2.1×10^{11} N/m^2, and Poisson ratio 0.29. This plate is simply supported on three edges and free on the fourth edge corresponding to $x_2 = 0$ (see Fig. 2). The spectral problem related to the mean reduced matrix model is analyzed using the finite-element method. The mean finite-element model is constituted of a regular rectangular mesh with a constant step size of 0.01 m in x_1 and x_2 (41 nodes in the width, 51 nodes in the length). Consequently, all the finite elements are the same and each one is a four-node square plate element. There are 2000 finite elements and $m = 6009$ degrees of freedom (DOF) (x_3 translations and x_1- and x_2 rotations). The eigenfrequencies $\underline{\omega}_\alpha = 2\pi\underline{\nu}_\alpha$ of the mean model (see Sec. III A) are $\underline{\nu}_1 = 1.94$, $\underline{\nu}_2 = 10.28$, $\underline{\nu}_3 = 15.47$,..., $\underline{\nu}_{23} = 150.89$, $\underline{\nu}_{24} = 155.82$, $\underline{\nu}_{25} = 167.44$,..., $\underline{\nu}_{37} = 247.89$, $\underline{\nu}_{38} = 253.01$,..., $\underline{\nu}_{80} = 527.29$,..., $\underline{\nu}_{120} = 817.65$ Hz. The finite-element discretization of the prescribed external forces yields an impulsive load vector denoted $\mathbf{F}(t) \in \mathbb{R}^m$ which is written as $\mathbf{F}(t) = e(t)\mathbf{Z}$. The spa-

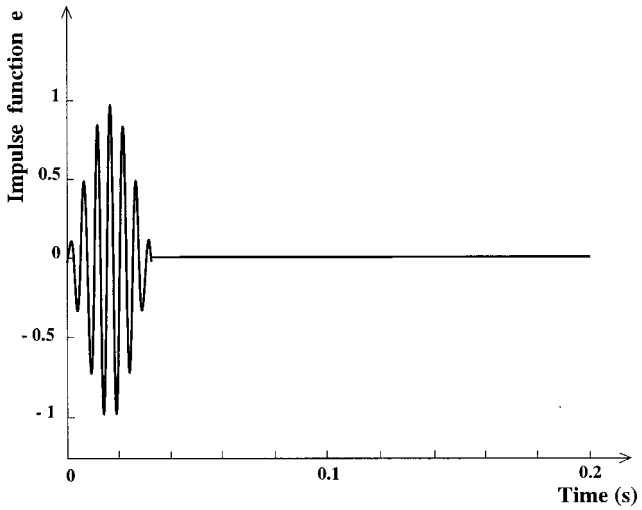


FIG. 3. Graph of wave impulse function $t \mapsto e(t)$.

tial part $\mathbf{Z}=(Z_1,\ldots,Z_m)\in\mathbb{R}^m$ is independent of time t and is such that $Z_j=0$ for all j in $\{1,...,m\}$ except for the nine DOFs corresponding to the nodes whose (x_1,x_2) coordinates are (0.30,0.25), (0.30,0.26), (0.30,0.27), (0.31,0.25), (0.31,0.26), (0.31,0.27), (0.32,0.25), (0.32,0.26), and (0.32,0.27), for which $Z_j=1$. Let t_1 be defined by $t_1=2\pi/\Delta\omega$, in which $\Delta\omega=2\pi\times 60$ rad/s and let $\Omega_c=2\pi\times 200$ rad/s. The impulse function $t\mapsto e(t)$ is a wave-type impulse function such that, for $t<0$ and $t>t_1$, $e(t)=0$, and for $0\leqslant t\leqslant t_1$

$$e(t)=\frac{1}{\Delta\omega(t-t_1)}\{\sin\{(\Omega_c+\Delta\omega/2)(t-t_1)\}-\sin\{(\Omega_c-\Delta\omega/2)(t-t_1)\}\}. \quad (103)$$

Figure 3 shows the graph of function $t\mapsto e(t)$ and Fig. 4 shows the graph of the modulus of its Fourier transform. It can be seen in Fig. 4 that the main part of the energy of impulse function e is distributed over the [150,250] Hz frequency band in which there are 15 structural modes of the mean model (one has $\underline{\nu}_{23}=150.89$ Hz and $\underline{\nu}_{37}=247.89$ Hz). It is assumed that the damping rate ξ of the mean model is 0.001 for frequencies around 200 Hz. The generalized damping matrix $[\underline{D}_n]$ of the mean reduced matrix model, defined by Eq. (29), is written as $[\underline{D}_n]=2\xi\Omega_{\text{ref}}[\underline{M}_n]$, in which $[\underline{M}_n]$ is the generalized mass matrix of the mean reduced matrix model, defined by Eq. (29), and where $\Omega_{\text{ref}}=2\pi\times 200$ rad/s.

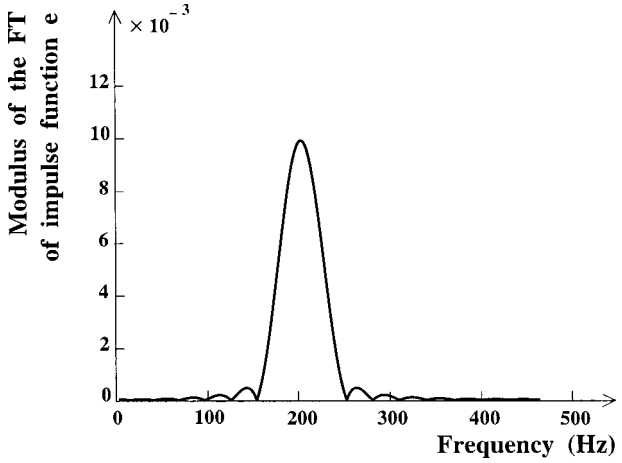


FIG. 4. Graph of the modulus of the Fourier transform of the wave impulse function.

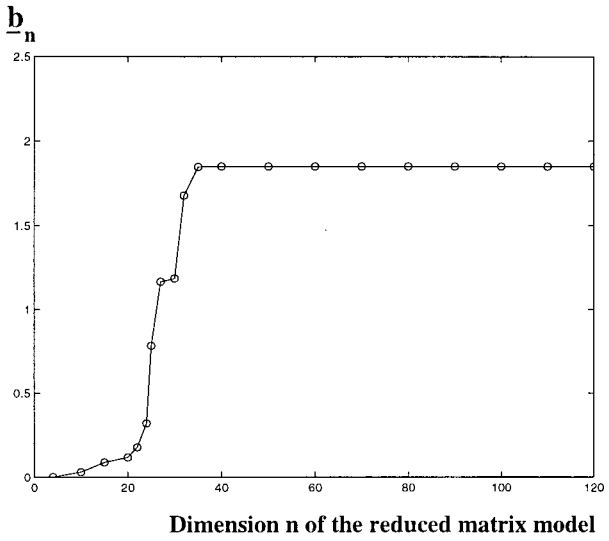


FIG. 5. Graph of the convergence of dynamic magnification factor $\underline{b}_n$ as function of n for the mean reduced matrix model.

B. Transient response of the mean model

The transient response of the mean model is calculated by solving the evolution problem defined by Eqs. (26)–(27) using an unconditionally stable implicit step-by-step integration method (Newmark integration scheme) with a time-step size $\Delta t=1/4000$ s. This time step corresponds to 10 time steps per period for the structural mode of the mean model whose eigenfrequency is $\underline{\nu}_{61}=402.24$ Hz. The finite-element approximation of the maximum $0.5\underline{k}(\mathbf{u}_S,\mathbf{u}_S)$ of the quasi-static response of the mean model (see Sec. II B) is equal to 2.5384. Figure 5 shows the convergence of the dynamic magnification factor $\underline{b}_n$ of the mean model, defined by Eq. (31), as dimension n of the mean reduced matrix model increases. From Fig. 5, it can be deduced that the transient response of the mean model is reasonably converged when $n\geqslant 40$. Figure 6 shows the graph of function $t\mapsto\underline{r}_n(t)$ for $n=40$, in which $\underline{r}_n(t)$ is the response ratio for the elastic energy calculated with the mean reduced matrix model and defined by Eq. (30). For $n=40$, the corresponding value of the dynamic magnification factor is $\underline{b}_n=1.85$.

C. Transient response of the model with random uncertainties

Let us choose $n_0=4$. Therefore, for the convergence analysis with respect to dimension n of the reduced matrix model with random uncertainties, we have to consider $n\geqslant n_0=4$. The dispersions of the generalized mass, damping, and stiffness random matrices of the reduced matrix model with random uncertainties, are controlled by parameters δ_M, δ_D, and δ_K introduced in Sec. VI A, which have to verify the constraints defined by Eq. (76)

$$0<\delta_M,\delta_D,\delta_K<0.7453. \quad (104)$$

The numerical simulations presented below correspond to the values

$$\delta_M = 0.3, \quad \delta_D = 0.3, \quad \delta_K = 0.3, \tag{105}$$

which verify Eq. (104). We are interested in the random response ratio $R_n(t)$ defined by Eq. (35) and the random dynamic magnification factor $B_n = \max_{t \geq 0} R_n(t)$ defined by Eq. (36). The transient response of the structure with random uncertainties is calculated using the Monte Carlo numerical simulation method. For given generalized mass, damping, and stiffness matrices, the evolution problem defined by Eqs. (81)–(82) is solved with the same Newmark integration scheme and with the same time-step size $\Delta t = 1/4000$ s.

The Monte Carlo numerical simulation is carried out with n_S samples, denoted $\theta_1, \ldots, \theta_{n_S}$, for which the samples $t \mapsto R_n(t;\theta_1), \ldots, t \mapsto R_n(t;\theta_{n_S})$ are numerically calculated. For fixed t, the mean value of random variable $R_n(t)$ is estimated by

$$E\{R_n(t)\} \simeq \frac{1}{n_S} \sum_{j=1}^{n_S} R_n(t;\theta_j). \tag{106}$$

The samples of random variable B_n are such that

$$B_n(\theta_j) = \max_{t \geq 0} R_n(t;\theta_j). \tag{107}$$

The mean value of random variable B_n is estimated by

$$E\{B_n\} \simeq \frac{1}{n_S} \sum_{j=1}^{n_S} B_n(\theta_j). \tag{108}$$

Finally, we introduce the function $t \mapsto R_{n,\max}(t;\boldsymbol{\theta}(n_S))$ and the real number $B_{n,\max}(\boldsymbol{\theta}(n_S))$ defined by

$$R_{n,\max}(t;\boldsymbol{\theta}(n_S)) = \max_{j=1,\ldots,n_S} R_n(t;\theta_j), \tag{109}$$

$$B_{n,\max}(\boldsymbol{\theta}(n_S)) = \max_{j=1,\ldots,n_S} B_n(\theta_j), \tag{110}$$

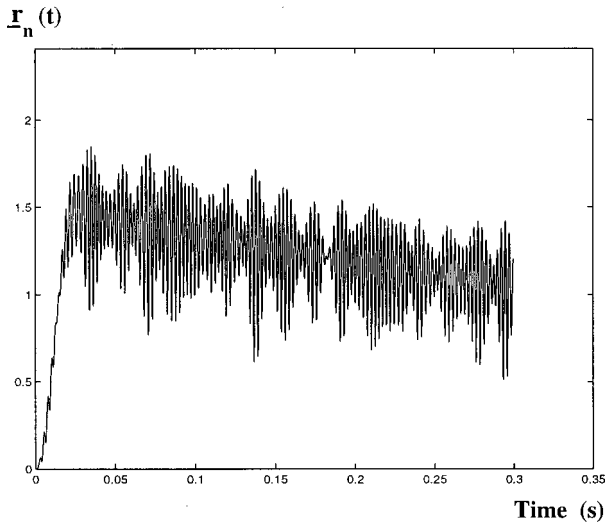


FIG. 6. Graph of function $t \mapsto \underline{r}_n(t)$ for $n = 40$ corresponding to the response ratio for the mean reduced matrix model.

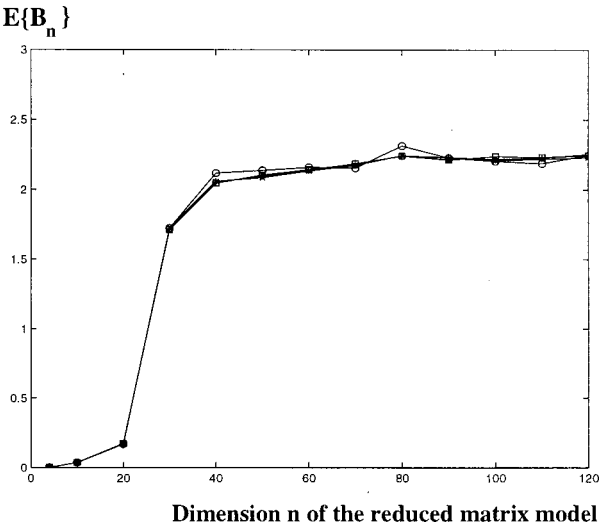


FIG. 7. Graph of function $n \mapsto E\{B_n\}$ (mathematical expectation of the random dynamic magnification factor) for $n_S = 50$ (circle symbol), 300 ($\times$-mark symbol), 600 (plus symbol), and 900 (square symbol).

in which $\boldsymbol{\theta}(n_S) = (\theta_1, \ldots, \theta_{n_S})$. Figure 7 shows the functions $n \mapsto E\{B_n\}$ calculated by Eq. (108) and Fig. 8 shows the function $n \mapsto B_{n,\max}(\boldsymbol{\theta}(n_S))$ calculated by Eq. (110) for $n_S = 50$, 300, 600, and 900. For n_S sufficiently high ($n_S \geq 300$) the Monte Carlo numerical method is reasonably converged and it can be seen that the nonparametric model proposed is convergent with respect to dimension n of the random reduced matrix model (see Sec. VII D). For $n = 120$ and $n_S = 900$, the value of $B_{n,\max}(\boldsymbol{\theta}(n_S))$ is 3.45. This value has to be compared to the value for the mean model, which is 1.85. Finally, Fig. 9 is relative to $n = 120$ and $n_S = 900$ and shows three curves: the lower thin solid line corresponds to the graph of function $t \mapsto \underline{r}_n(t)$, the thick solid line to the graph of function $t \mapsto E\{R_n(t)\}$ calculated by Eq. (106), and the upper thin solid line to the graph of function $t \mapsto R_{n,\max}(t;\boldsymbol{\theta}(n_S))$ defined by Eq. (109). This figure shows the sensitivity of the maximum transient response due to random uncertainties. The dynamic magnification factor in-

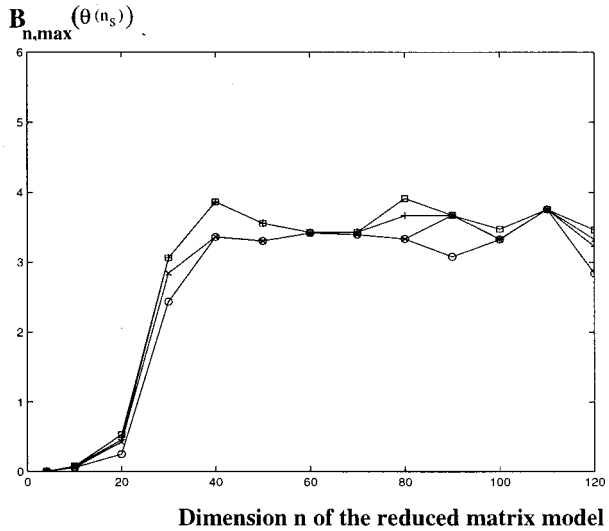


FIG. 8. Graph of function $n \mapsto B_{n,\max}(\boldsymbol{\theta}(n_S))$ (maximum of the random dynamic magnification factor) for $n_S = 50$ (circle symbol), 300 ($\times$-mark symbol), 600 (plus symbol), and 900 (square symbol).

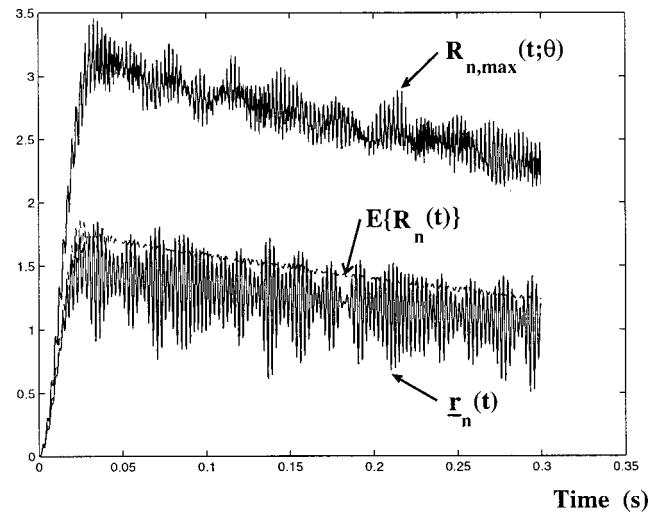


FIG. 9. Transient responses $t \mapsto \underline{r}_n(t)$ (lower thin solid line), $t \mapsto E\{R_n(t)\}$ (thick solid line) and $t \mapsto R_{n,\max}(t;\boldsymbol{\theta}(n_S))$ (upper thin solid line) for $n=120$ and $n_S=900$.

creases when the random uncertainties increase, and is greater than the deterministic dynamic amplification factor of the mean model.

IX. CONCLUSIONS

We have presented a new approach allowing the random uncertainties to be modeled by a nonparametric model for prediction of transient responses to impulsive loads in linear structural dynamics. This approach has been presented in the context of structural dynamics but can be extended without any difficulty to structural-acoustic problems such as a structure coupled with an internal acoustic cavity. The parametric approaches existing in literature are very useful when the number of uncertain parameters is small and when the probabilistic model can be constructed for the set of parameters considered. The nonparametric approach presented is useful when the number of uncertain parameters is high or when the probabilistic model is difficult to construct for the set of parameters considered. In addition, the parametric approaches do not allow the model uncertainties to be taken into account (because a parametric approach is associated with a fixed model exhibiting some parameters), whereas the nonparametric approach proposed allows to take into account the model uncertainties. For this nonparametric approach, the information used does not require the description of the local parameters of the mechanical model. The probability model is deduced from the use of the entropy optimization principle, whose available information is constituted of the fundamental algebraic properties related to the generalized mass, damping, and stiffness matrices which have to be positive-definite symmetric matrices for which the second-order moments of their inverse have to exist, and the knowledge of these matrices for the mean reduced matrix model. An explicit construction and representation of the probability model has been obtained and are very well suited to algebraic calculus and to Monte Carlo numerical simulation. The fundamental properties related to the convergence of the stochastic solution with respect to the dimension of the random reduced matrix model has been analyzed. This convergence analysis carried out has allowed the consistency of the theory proposed to be proved and the parameters of the probability distribution of the random generalized matrices to be clearly defined. With this nonparametric model, the probability distribution of each random generalized matrix of the random reduced matrix model depends only on two parameters: the mean generalized matrix associated with the mean mechanical model and corresponding to the design model, and a scalar parameter δ whose values have to be fixed by the designer in the interval [0,1] in order to give the dispersion level attached to the random generalized matrix. It seems clear that parameter δ should be a global parameter resulting from expertise, because the model uncertainties which can be taken into account with the nonparametric model cannot be quantified in terms of correlation between random variables. For instance, if there is no uncertainty for the stiffness model, then $\delta_K=0$. On the other hand, if it is assumed that the global uncertainty for the stiffness model is 10%, then δ_K has to be 0.1. Nevertheless, experiments are in progress to study the correlation which could exist between the dispersion of the random responses and parameters δ_M, δ_D, and δ_K associated with the random generalized matrices.

APPENDIX A: PROOF OF EQ. (53)

In this Appendix, we prove Eq. (53), i.e.,

$$\lambda_{A_n} > \gamma+1 \Rightarrow E\{\|[\mathbf{A}_n]^{-1}\|_F^{\gamma}\} < +\infty, \quad \gamma \geqslant 1, \tag{A1}$$

in which $\gamma \geqslant 1$ is a positive integer. Since $[\mathbf{A}_n]$ is a positive-definite random matrix, it can be written as $[\mathbf{A}_n]=[\mathbf{R}_n]\times[\boldsymbol{\Sigma}_n][\mathbf{R}_n]^T$, in which $[\mathbf{R}_n]$ is an orthogonal random matrix and $[\boldsymbol{\Sigma}_n]$ is a diagonal positive-definite random matrix whose diagonal elements are the random eigenvalues $\Sigma_1,\ldots,\Sigma_n$. We then have

$$\|[\mathbf{A}_n]^{-1}\|_F^2 = \frac{1}{\Sigma_1^2}+\cdots+\frac{1}{\Sigma_n^2}. \tag{A2}$$

Using the probability density function of random vector $\Sigma_1,\ldots,\Sigma_n$ constructed in Ref. 29 and reusing the proof given in Sec. 3.6 of this reference, it can be proved that $E\{\|[\mathbf{A}_n]^{-1}\|_F^{\gamma}\} < +\infty$ if and only if

$$\mathcal{I}_\varepsilon = \int_{\|\boldsymbol{\sigma}\|<\varepsilon} (\sigma_1\times\cdots\times\sigma_n)^{\lambda_A-1}\left(\frac{1}{\sigma_1^2}+\cdots+\frac{1}{\sigma_n^2}\right)^{\gamma/2} \times\{\Pi_{j<k}|\sigma_k-\sigma_j|\}d\boldsymbol{\sigma} < +\infty, \tag{A3}$$

in which $0<\varepsilon\leqslant 1$, $\boldsymbol{\sigma}=(\sigma_1,\ldots,\sigma_n)$ and $d\boldsymbol{\sigma}=d\sigma_1\times\ldots\times d\sigma_n$. We introduce polar coordinates r and $\boldsymbol{\theta}=(\theta_1,\ldots,\theta_{n-1})$ in $\mathbb{R}^n$ such that $\sigma_1=r\sin\theta_1$, $\sigma_2=r\cos\theta_1\sin\theta_2,\ldots,\sigma_{n-1}=r\cos\theta_1\cos\theta_2\ldots\cos\theta_{n-2}\sin\theta_{n-1}$, and $\sigma_n=r\cos\theta_1\cos\theta_2\ldots\cos\theta_{n-2}\cos\theta_{n-1}$, in which $-\pi/2<\theta_j\leqslant\pi/2$ for $j=1,\ldots,n-2$ and $-\pi<\theta_{n-1}\leqslant\pi$. We have $d\boldsymbol{\sigma}=r^{n-1}h_1(\boldsymbol{\theta})dr\,d\boldsymbol{\theta}$ in which $h_1(\boldsymbol{\theta})=|\cos^{n-2}\theta_1\times\cos^{n-3}\theta_3\ldots\cos\theta_{n-2}|$. If we assume that $\lambda_{A_n}>\gamma+1$ with $\gamma>0$, then Eq. (A3) holds if $\lambda_{A_n}>\gamma/n+(1-n)/2$ and consequently, λ_{A_n} has to be such that $\lambda_{A_n}>\max\{\gamma+1,\gamma/n+(1-n)/2\}$. Since $n\geqslant 1$ and $\gamma>0$, we deduce Eq. (A1).

APPENDIX B: PROOF OF EQ. (97)

In this Appendix, we prove Eq. (97). Let $\Sigma_1,\ldots,\Sigma_n$ be the positive-valued random eigenvalues of random matrix $[\mathbf{G}_{A_n}]$ with values in $\mathbb{M}_n^+(\mathbb{R})$. It is proved in Refs. 28 and 29 that the probability density function $p_{\mathbf{\Sigma}}(\boldsymbol{\sigma})$ with respect to $d\boldsymbol{\sigma}=d\sigma_1\times\ldots\times d\sigma_n$ of the random vector $\mathbf{\Sigma}=(\Sigma_1,\ldots,\Sigma_n)$ with values in $\mathcal{D}_n=]0,+\infty[^n\subset\mathbb{R}^n$ is written as

$$p_{\mathbf{\Sigma}}(\boldsymbol{\sigma})=\mathbb{1}_{\mathcal{D}_n}(\boldsymbol{\sigma})\times c_{\Sigma}\times(\sigma_1\times\cdots\times\sigma_n)^{\lambda_{A_n}-1}\times\{\Pi_{\alpha<\beta}|\sigma_\beta-\sigma_\alpha|\}e^{-1/2(n-1+2\lambda_{A_n})(\sigma_1+\cdots+\sigma_n)}, \tag{B1}$$

in which c_{Σ} is a constant of normalization defined by the equation $\int_{\mathcal{D}_n}p_{\mathbf{\Sigma}}(\boldsymbol{\sigma})d\boldsymbol{\sigma}=1$ and where $\lambda_{A_n}=l_A(n)$ is given by Eq. (66). Let $\tilde{\mathbf{\Sigma}}=(\tilde{\Sigma}_1,\ldots,\tilde{\Sigma}_n)$ be the order statistics of $\mathbf{\Sigma}=(\Sigma_1,\ldots,\Sigma_n)$ such that $0\leq\tilde{\Sigma}_1\leq\tilde{\Sigma}_2\leq\cdots\leq\tilde{\Sigma}_n$. Let $\mathcal{P}_n$ be the group of all the permutations τ of the first n positive integers $\{1,2,...,n\}$. Since $p_{\mathbf{\Sigma}}$ is a symmetric function in all the variables $\sigma_1,\ldots,\sigma_n$, that is to say, for any permutation τ in $\mathcal{P}_n$

$$p_{\mathbf{\Sigma}}(\sigma_{\tau(1)},\ldots,\sigma_{\tau(n)})=p_{\Sigma}(\sigma_1\ldots\sigma_n), \tag{B2}$$

then the probability density function $p_{\tilde{\mathbf{\Sigma}}}(\boldsymbol{\sigma})$ of order statistics $\tilde{\mathbf{\Sigma}}$ with respect to $\mathbf{d}\boldsymbol{\sigma}=d\sigma_1\times\ldots\times d\sigma_n$ is written as[52]

$$p_{\tilde{\mathbf{\Sigma}}}(\boldsymbol{\sigma})=\frac{\mathbb{1}_{\mathcal{S}_n}(\boldsymbol{\sigma})p_{\mathbf{\Sigma}}(\boldsymbol{\sigma})}{\int_{\mathcal{S}_n}p_{\mathbf{\Sigma}}(\boldsymbol{\sigma})d\boldsymbol{\sigma}}, \tag{B3}$$

in which $\mathcal{S}_n$ is the simplex defined by

$$\mathcal{S}_n=\{\boldsymbol{\sigma}=(\sigma_1,\ldots,\sigma_n)\in\mathbb{R}^n;\quad 0<\sigma_1<\cdots<\sigma_n<+\infty\}. \tag{B4}$$

Consequently, using Eqs. (96), (B3), (B1), and (65), (66) yields

$$E\{\|[\mathbf{G}_{A_n}]^{-1}\|^2\}=\frac{\int_{\mathcal{S}_n}\sigma_1^{-2}h(\boldsymbol{\sigma})d\boldsymbol{\sigma}}{\int_{\mathcal{S}_n}h(\boldsymbol{\sigma})d\boldsymbol{\sigma}}, \tag{B5}$$

in which

$$h(\boldsymbol{\sigma})=(\sigma_1\times\cdots\times\sigma_n)^{a(n+1)}\{\Pi_{\alpha<\beta}|\sigma_\beta-\sigma_\alpha|\}\times e^{-b(n+1)(\sigma_1+\cdots+\sigma_n)}, \tag{B6}$$

where

$$a=\frac{1-\delta_A^2}{2\,\delta_A^2}>0,\quad b=\frac{1}{2\,\delta_A^2}.$$

Let $\varepsilon>0$ be a positive real number independent of n. Since $h(\boldsymbol{\sigma})>0$ for $\boldsymbol{\sigma}\in\mathcal{S}_n$, we have

$$\int_{\varepsilon}^{+\infty}d\sigma_1\int_{\sigma_1}^{+\infty}d\sigma_2\cdots\int_{\sigma_{n-1}}^{+\infty}d\sigma_n\,\sigma_1^{-2}h(\boldsymbol{\sigma})\leq\frac{1}{\varepsilon^2}\int_{\mathcal{S}_n}h(\boldsymbol{\sigma})d\boldsymbol{\sigma}, \tag{B7}$$

and consequently, from Eqs. (B5) and (B7), we deduce that

$$E\{\|[\mathbf{G}_{A_n}]^{-1}\|^2\}\leq\frac{1}{\varepsilon^2}+\frac{\int_0^{\varepsilon}d\sigma_1\int_{\sigma_1}^{+\infty}d\sigma_2\cdots\int_{\sigma_{n-1}}^{+\infty}d\sigma_n\,\sigma_1^{-2}h(\boldsymbol{\sigma})}{\int_0^{+\infty}d\sigma_1\int_{\sigma_1}^{+\infty}d\sigma_2\cdots\int_{\sigma_{n-1}}^{+\infty}d\sigma_n\,\sigma_1^{-2}h(\boldsymbol{\sigma})}. \tag{B8}$$

Since the function $\boldsymbol{\sigma}\mapsto\sigma_1^{-2}h(\boldsymbol{\sigma})$ is symmetric in the variables $\sigma_2,\ldots,\sigma_n$, we have

$$\int_0^{\varepsilon}d\sigma_1\int_{\sigma_1}^{+\infty}d\sigma_2\cdots\int_{\sigma_{n-1}}^{+\infty}d\sigma_n\,\sigma_1^{-2}h(\boldsymbol{\sigma})=\frac{1}{(n-1)!}\int_0^{\varepsilon}d\sigma_1\int_{\sigma_1}^{+\infty}d\sigma_2\cdots\int_{\sigma_1}^{+\infty}d\sigma_n\,\sigma_1^{-2}h(\boldsymbol{\sigma})\leq\frac{1}{(n-1)!}\int_0^{\varepsilon}d\sigma_1\int_0^{+\infty}d\sigma_2\cdots\int_0^{+\infty}d\sigma_n\,\sigma_1^{-2}h(\boldsymbol{\sigma}), \tag{B9}$$

and since the function $\boldsymbol{\sigma}\mapsto h(\boldsymbol{\sigma})$ is symmetric in the variables $\sigma_1,\ldots,\sigma_n$, we have

$$\int_0^{+\infty}d\sigma_1\int_{\sigma_1}^{+\infty}d\sigma_2\cdots\int_{\sigma_{n-1}}^{+\infty}d\sigma_n\,h(\boldsymbol{\sigma})=\frac{1}{n!}\int_0^{+\infty}d\sigma_1\int_0^{+\infty}d\sigma_2\cdots\int_0^{+\infty}d\sigma_n\,h(\boldsymbol{\sigma}). \tag{B10}$$

We then deduce that

$$E\{\|[\mathbf{G}_{A_n}]^{-1}\|^2\}\leq\frac{1}{\varepsilon^2}+H_n(\varepsilon), \tag{B11}$$

with

$$H_n(\varepsilon)=\frac{n\int_0^{\varepsilon}d\sigma_1\int_0^{+\infty}d\sigma_2\cdots\int_0^{+\infty}d\sigma_n\,\sigma_1^{-2}h(\boldsymbol{\sigma})}{\int_0^{+\infty}d\sigma_1\int_0^{+\infty}d\sigma_2\cdots\int_0^{+\infty}d\sigma_n\,h(\boldsymbol{\sigma})}. \tag{B12}$$

Using a similar proof to the proof of Lemma 4.4, page 196 of Ref. 56, it can be proved that, for $\varepsilon>0$ taken sufficiently small and independent of n, we have

$$\lim_{n\to+\infty}H_n(\varepsilon)=0. \tag{B13}$$

From Eqs. (B11) and (B13), we deduce that $n\mapsto E\{\|[\mathbf{G}_{A_n}]^{-1}\|^2\}$ is a bounded function and consequently, Eq. (97) is proved.

APPENDIX C: INEQUALITIES FOR THE RANDOM INSTANTANEOUS KINETIC ENERGY, POTENTIAL ENERGY, AND DISSIPATED POWER

In this Appendix, we construct inequalities for the random instantaneous kinetic energy defined by $\frac{1}{2}\langle[\mathbf{M}_n]\dot{\mathbf{Q}}^n(t),\dot{\mathbf{Q}}^n(t)\rangle$, the random instantaneous potential energy defined by $\frac{1}{2}\langle[\mathbf{K}_n]\mathbf{Q}^n(t),\mathbf{Q}^n(t)\rangle$, and the random instantaneous dissipated power $\langle[\mathbf{D}_n]\dot{\mathbf{Q}}^n(\tau),\dot{\mathbf{Q}}^n(\tau)\rangle$.

Let $[\mathbf{A}_n]$ be a random matrix $[\mathbf{M}_n]$, $[\mathbf{D}_n]$, or $[\mathbf{K}_n]$ defined on probability space $(\mathcal{A},\mathcal{T},\mathcal{P})$ which is written [see Eq. (57)] as $[\mathbf{A}_n]=[\underline{L}_{A_n}]^T[\mathbf{G}_{A_n}][\underline{L}_{A_n}]$. Let $\mathbf{Q}^n$ be an $\mathbb{R}^n$-valued random vector defined on the same probability space $(\mathcal{A},\mathcal{T},\mathcal{P})$ and which is not independent of random matrix $[\mathbf{G}_{A_n}]$. Let $[\underline{A}_n]=E\{[\mathbf{A}_n]\}=[\underline{L}_{A_n}]^T[\underline{L}_{A_n}]\in\mathbb{M}_n^+(\mathbb{R})$. We then have

$$\langle [\underline{A}_n]\mathbf{Q}^n,\mathbf{Q}^n\rangle \leq \|[\mathbf{G}_{A_n}]^{-1}\|\langle [\mathbf{A}_n]\mathbf{Q}^n,\mathbf{Q}^n\rangle. \quad \text{(C1)}$$

To prove Eq. (C1), we write $\langle [\underline{A}_n]\mathbf{Q}^n,\mathbf{Q}^n\rangle = \langle \mathbf{S}^n,\mathbf{S}^n\rangle$ in which $\mathbf{S}^n = [\underline{L}_{A_n}]\mathbf{Q}^n$. Since $[\mathbf{G}_{A_n}]$ is a random matrix with values in $\mathbb{M}_n^+(\mathbb{R})$, we can write

$$\begin{aligned}\langle [\underline{A}_n]\mathbf{Q}^n,\mathbf{Q}^n\rangle &= \langle [\mathbf{G}_{A_n}]^{-1}[\mathbf{G}_{A_n}]^{1/2}\mathbf{S}^n,[\mathbf{G}_{A_n}]^{1/2}\mathbf{S}^n\rangle \\ &\leq \|[\mathbf{G}_{A_n}]^{-1}\|\times\|[\mathbf{G}_{A_n}]^{1/2}\mathbf{S}^n\|^2 \\ &= \|[\mathbf{G}_{A_n}]^{-1}\|\langle [\mathbf{G}_{A_n}]\mathbf{S}^n,\mathbf{S}^n\rangle \\ &= \|[\mathbf{G}_{A_n}]^{-1}\|\langle [\underline{L}_{A_n}]^T[\mathbf{G}_{A_n}][\underline{L}_{A_n}]\mathbf{Q}^n,\mathbf{Q}^n\rangle \\ &= \|[\mathbf{G}_{A_n}]^{-1}\|\langle [\mathbf{A}_n]\mathbf{Q}^n,\mathbf{Q}^n\rangle,\end{aligned}$$

which is the inequality defined by Eq. (C1).

Let t and τ be two fixed times in $[0,T]$. Let $\mathbf{U}_n(t)$ be the mapping $\{\mathbf{x}\mapsto\mathbf{U}_n(\mathbf{x},t)\}$ and $\dot{\mathbf{U}}_n(t)=\partial\mathbf{U}_n(t)/\partial t$. From Eqs. (32), (13), (15), (16) and using Eqs. (23) and (29), we deduce that

$$\underline{m}(\dot{\mathbf{U}}_n(t),\dot{\mathbf{U}}_n(t)) = \langle [\underline{M}_n]\dot{\mathbf{Q}}^n(t),\dot{\mathbf{Q}}^n(t)\rangle, \quad \text{(C2)}$$

$$\underline{k}(\mathbf{U}_n(t),\mathbf{U}_n(t)) = \langle [\underline{K}_n]\mathbf{Q}^n(t),\mathbf{Q}^n(t)\rangle, \quad \text{(C3)}$$

$$\underline{d}(\dot{\mathbf{U}}_n(\tau),\dot{\mathbf{U}}_n(\tau)) = \langle [\underline{D}_n]\dot{\mathbf{Q}}^n(\tau),\dot{\mathbf{Q}}^n(\tau)\rangle. \quad \text{(C4)}$$

Equations (14), (17) and Eqs. (C2)–(C4) yield

$$c_{\underline{m}}\|\dot{\mathbf{U}}_n(t)\|_{\mathbb{H}}^2 \leq \langle [\underline{M}_n]\dot{\mathbf{Q}}^n(t),\dot{\mathbf{Q}}^n(t)\rangle, \quad \text{(C5)}$$

$$c_{\underline{k}}\|\mathbf{U}_n(t)\|_{\mathbb{V}}^2 \leq \langle [\underline{K}_n]\mathbf{Q}^n(t),\mathbf{Q}^n(t)\rangle, \quad \text{(C6)}$$

$$c_{\underline{d}}\|\dot{\mathbf{U}}_n(\tau)\|_{\mathbb{V}}^2 \leq \langle [\underline{D}_n]\dot{\mathbf{Q}}^n(\tau),\dot{\mathbf{Q}}^n(\tau)\rangle, \quad \text{(C7)}$$

in which $c_{\underline{m}}$, $c_{\underline{k}}$, and $c_{\underline{d}}$ are positive finite constants independent of n. Applying Eq. (C1) to Eqs. (C5)–(C7), we deduce the following inequalities of positive-valued random variables:

$$\frac{c_{\underline{m}}}{\|[\mathbf{G}_{M_n}]^{-1}\|}\|\dot{\mathbf{U}}_n(t)\|_{\mathbb{H}}^2 \leq \langle [\mathbf{M}_n]\dot{\mathbf{Q}}^n(t),\dot{\mathbf{Q}}^n(t)\rangle, \quad \text{(C8)}$$

$$\frac{c_{\underline{k}}}{\|[\mathbf{G}_{K_n}]^{-1}\|}\|\mathbf{U}_n(t)\|_{\mathbb{V}}^2 \leq \langle [\mathbf{K}_n]\mathbf{Q}^n(t),\mathbf{Q}^n(t)\rangle, \quad \text{(C9)}$$

$$\frac{c_{\underline{d}}}{\|[\mathbf{G}_{D_n}]^{-1}\|}\|\dot{\mathbf{U}}_n(\tau)\|_{\mathbb{V}}^2 \leq \langle [\mathbf{D}_n]\dot{\mathbf{Q}}^n(\tau),\dot{\mathbf{Q}}^n(\tau)\rangle. \quad \text{(C10)}$$

APPENDIX D: INEQUALITY FOR THE RANDOM INSTANTANEOUS INPUT POWER

In this Appendix, we introduce a mapping $\mathbf{f}(t)$ representing the prescribed external forces and we deduce an inequality for the random instantaneous input power which is absolutely necessary to construct the energy inequality. It should be noted that this kind of inequality cannot be constructed without introducing the continuous dual space of $\mathbb{V}$ and we give the reason below.

Let $\mathbb{V}'$ be the continuous dual space of $\mathbb{V}$ (i.e., the set of all the continuous linear forms on vector space $\mathbb{V}$ defined in Sec. II B) and $\langle \mathbf{f},\mathbf{v}\rangle_{\mathbb{V}',\mathbb{V}}$ be the duality bracket between $\mathbf{f}\in\mathbb{V}'$ and $\mathbf{v}\in\mathbb{V}$ which is linear with respect to $\mathbf{f}$ and $\mathbf{v}$. For all $\mathbf{f}$ in $\mathbb{V}'$ and $\mathbf{v}$ in $\mathbb{V}$, we have

$$|\langle \mathbf{f},\mathbf{v}\rangle_{\mathbb{V}',\mathbb{V}}| \leq \|\mathbf{f}\|_{\mathbb{V}'}\|\mathbf{v}\|_{\mathbb{V}}, \quad \text{(D1)}$$

in which

$$\|\mathbf{f}\|_{\mathbb{V}'} = \sup_{\mathbf{v}\in\mathbb{V},\mathbf{v}\neq 0}\{|\langle \mathbf{f},\mathbf{v}\rangle_{\mathbb{V}',\mathbb{V}}|/\|\mathbf{v}\|_{\mathbb{V}}\}, \quad \text{(D2)}$$

is the norm on $\mathbb{V}'$. Since Hilbert space $\mathbb{H}$ (defined in Sec. II B) is identified to its continuous dual space $\mathbb{H}'$, we have $\mathbb{V}\subset\mathbb{H}\subset\mathbb{V}'$ and, if $\mathbf{f}$ is in $\mathbb{H}$, we have $\langle \mathbf{f},\mathbf{v}\rangle_{\mathbb{V}',\mathbb{V}} = (\mathbf{f},\mathbf{v})_{\mathbb{H}}$.

Since $g(\mathbf{v};t)$ defined by Eq. (11) is continuous on $\mathbb{V}$, there exists a unique element $\mathbf{f}(t)$ in $\mathbb{V}'$ such that

$$g(\mathbf{v};t) = \langle \mathbf{f}(t),\mathbf{v}\rangle_{\mathbb{V}',\mathbb{V}}, \quad \forall\mathbf{v}\in\mathbb{V}, \quad \text{(D3)}$$

and Eq. (D1) holds. It should be noted that if $\mathbf{g}_{\text{surf}}=0$, then $\mathbf{f}$ coincides with $\mathbf{g}_{\text{vol}}$ and consequently $\mathbf{f}(t)$ is in $\mathbb{H}$; in general, $\mathbf{g}_{\text{surf}}\neq 0$ and then $\mathbf{f}(t)$ is not in $\mathbb{H}$ but is in $\mathbb{V}'$. This is why we need vector space $\mathbb{V}'$ (it is wrong to consider a nonzero prescribed surface force field ($\mathbf{g}_{\text{surf}}\neq 0$) and to assume that $\mathbf{f}(t)$ is a square integrable function on domain Ω, i.e., that it belongs to $\mathbb{H}$).

Let $\mathbf{U}_n(t)$ be the mapping $\{\mathbf{x}\mapsto\mathbf{U}_n(\mathbf{x},t)\}$ and $\dot{\mathbf{U}}_n(t) = \partial\mathbf{U}_n(t)/\partial t$. From Eqs. (D3), (28), and (32), we deduce the expression of the random instantaneous input power

$$\langle \mathbf{F}^n(\tau),\dot{\mathbf{Q}}^n(\tau)\rangle = \langle \mathbf{f}(\tau),\dot{\mathbf{U}}_n(\tau)\rangle_{\mathbb{V}',\mathbb{V}}. \quad \text{(D4)}$$

Using Eqs. (D1) and (D4) yields

$$|\langle \mathbf{F}^n(\tau),\dot{\mathbf{Q}}^n(\tau)\rangle| \leq \|\mathbf{f}(\tau)\|_{\mathbb{V}'}\|\dot{\mathbf{U}}_n(\tau)\|_{\mathbb{V}}. \quad \text{(D5)}$$

APPENDIX E: BASIC INEQUALITIES DERIVED FROM THE RANDOM ENERGY EQUATION

In this Appendix, for any t fixed in $[0,T]$ and for any $n\geq n_0$, using the energy inequality, we prove basic inequalities relative to $\|\|\dot{\mathbf{U}}_n(t)\|\|_{\mathbb{H}}^2$ and $\|\|\mathbf{U}_n(t)\|\|_{\mathbb{V}}^2$.

Taking the inner product of the two members of Eq. (33) with $\dot{\mathbf{Q}}^n(t)$ yields

$$\begin{aligned}&\langle [\mathbf{M}_n]\ddot{\mathbf{Q}}^n(t),\dot{\mathbf{Q}}^n(t)\rangle + \langle [\mathbf{D}_n]\dot{\mathbf{Q}}^n(t),\dot{\mathbf{Q}}^n(t)\rangle \\ &\quad + \langle [\mathbf{K}_n]\mathbf{Q}^n(t),\dot{\mathbf{Q}}^n(t)\rangle = \langle \mathbf{F}^n(t),\dot{\mathbf{Q}}^n(t)\rangle. \quad \text{(E1)}\end{aligned}$$

Using the symmetry properties of random matrices $[\mathbf{M}_n]$, $[\mathbf{D}_n]$, and $[\mathbf{K}_n]$, Eq. (E1) is rewritten as

$$\begin{aligned}&\frac{1}{2}\frac{d}{dt}\{\langle [\mathbf{M}_n]\dot{\mathbf{Q}}^n(t),\dot{\mathbf{Q}}^n(t)\rangle + \langle [\mathbf{K}_n]\mathbf{Q}^n(t),\mathbf{Q}^n(t)\rangle\} \\ &\quad + \langle [\mathbf{D}_n]\dot{\mathbf{Q}}^n(t),\dot{\mathbf{Q}}^n(t)\rangle = \langle \mathbf{F}^n(t),\dot{\mathbf{Q}}^n(t)\rangle. \quad \text{(E2)}\end{aligned}$$

Integrating the two members of Eq. (E2) with respect to t over $[0,t]$ with $0\leq t\leq T$ and taking into account Eq. (34) yields the energy random equation

$$\langle [\mathbf{M}_n]\dot{\mathbf{Q}}^n(t),\dot{\mathbf{Q}}^n(t)\rangle + \langle [\mathbf{K}_n]\mathbf{Q}^n(t),\mathbf{Q}^n(t)\rangle + 2\int_0^t \langle [\mathbf{D}_n]\dot{\mathbf{Q}}^n(\tau),\dot{\mathbf{Q}}^n(\tau)\rangle d\tau = 2\int_0^t \langle \mathbf{F}^n(\tau),\dot{\mathbf{Q}}^n(\tau)\rangle d\tau, \tag{E3}$$

which is an equality of random variables. Since $[\mathbf{M}_n]$, $[\mathbf{D}_n]$, and $[\mathbf{K}_n]$ are random matrices with values in $\mathbb{M}_n^+(\mathbb{R})$, for any fixed t in $[0,T]$, we have $\langle [\mathbf{M}_n]\dot{\mathbf{Q}}^n(t),\dot{\mathbf{Q}}^n(t)\rangle \geqslant 0$, $\langle [\mathbf{K}_n]\mathbf{Q}^n(t),\mathbf{Q}^n(t)\rangle \geqslant 0$ and $\int_0^t \langle [\mathbf{D}_n]\dot{\mathbf{Q}}^n(\tau),\dot{\mathbf{Q}}^n(\tau)\rangle d\tau \geqslant 0$. From Eq. (E3), we then deduce that

$$\langle [\mathbf{M}_n]\dot{\mathbf{Q}}^n(t),\dot{\mathbf{Q}}^n(t)\rangle \leqslant 2\int_0^t |\langle \mathbf{F}^n(\tau),\dot{\mathbf{Q}}^n(\tau)\rangle| d\tau, \tag{E4}$$

$$\langle [\mathbf{K}_n]\mathbf{Q}^n(t),\mathbf{Q}^n(t)\rangle \leqslant 2\int_0^t |\langle \mathbf{F}^n(\tau),\dot{\mathbf{Q}}^n(\tau)\rangle| d\tau, \tag{E5}$$

$$\int_0^t \langle [\mathbf{D}_n]\dot{\mathbf{Q}}^n(\tau),\dot{\mathbf{Q}}^n(\tau)\rangle d\tau \leqslant \int_0^t |\langle \mathbf{F}^n(\tau),\dot{\mathbf{Q}}^n(\tau)\rangle| d\tau. \tag{E6}$$

Using the inequalities defined by Eqs. (C8)–(C10) and (D5), Eqs. (E4)–(E6) yield

$$c_{\underline{m}} \|\dot{\mathbf{U}}_n(t)\|_{\mathbb{H}}^2 \leqslant 2\int_0^t \|\mathbf{f}(\tau)\|_{\mathbb{V}'} \|[\mathbf{G}_{M_n}]^{-1}\| \|\dot{\mathbf{U}}_n(\tau)\|_{\mathbb{V}} d\tau, \tag{E7}$$

$$c_{\underline{k}} \|\dot{\mathbf{U}}_n(t)\|_{\mathbb{V}}^2 \leqslant 2\int_0^t \|\mathbf{f}(\tau)\|_{\mathbb{V}'} \|[\mathbf{G}_{K_n}]^{-1}\| \|\dot{\mathbf{U}}_n(\tau)\|_{\mathbb{V}} d\tau, \tag{E8}$$

$$c_{\underline{d}} \int_0^t \|\dot{\mathbf{U}}_n(\tau)\|_{\mathbb{V}}^2 d\tau \leqslant \int_0^t \|\mathbf{f}(\tau)\|_{\mathbb{V}'} \|[\mathbf{G}_{D_n}]^{-1}\| \|\dot{\mathbf{U}}_n(\tau)\|_{\mathbb{V}} d\tau. \tag{E9}$$

Since the left-hand sides and the right-hand sides of Eqs. (E7)–(E9) are positive-valued random variables, taking the mathematical expectation of the two members of inequalities (E7)–(E9) and using Eqs. (92) and (93) yields

$$c_{\underline{m}} |||\dot{\mathbf{U}}_n(t)|||_{\mathbb{H}}^2 \leqslant 2\int_0^t \|\mathbf{f}(\tau)\|_{\mathbb{V}'} E\{\|[\mathbf{G}_{M_n}]^{-1}\| \|\dot{\mathbf{U}}_n(\tau)\|_{\mathbb{V}}\} d\tau, \tag{E10}$$

$$c_{\underline{k}} |||\dot{\mathbf{U}}_n(t)|||_{\mathbb{V}}^2 \leqslant 2\int_0^t \|\mathbf{f}(\tau)\|_{\mathbb{V}'} E\{\|[\mathbf{G}_{K_n}]^{-1}\| \|\dot{\mathbf{U}}_n(\tau)\|_{\mathbb{V}}\} d\tau, \tag{E11}$$

$$c_{\underline{d}} \int_0^t |||\dot{\mathbf{U}}_n(\tau)|||_{\mathbb{V}}^2 d\tau \leqslant \int_0^t \|\mathbf{f}(\tau)\|_{\mathbb{V}'} E\{\|[\mathbf{G}_{D_n}]^{-1}\| \|\dot{\mathbf{U}}_n(\tau)\|_{\mathbb{V}}\} d\tau. \tag{E12}$$

Let $[\mathbf{G}_{A_n}]$ be $[\mathbf{G}_{M_n}]$, $[\mathbf{G}_{K_n}]$, or $[\mathbf{G}_{D_n}]$. Using the Holder inequality, we can write

$$E\{\|[\mathbf{G}_{A_n}]^{-1}\| \|\dot{\mathbf{U}}_n(\tau)\|_{\mathbb{V}}\} \leqslant \sqrt{E\{\|[\mathbf{G}_{A_n}]^{-1}\|^2\}}\, |||\dot{\mathbf{U}}_n(\tau)|||_{\mathbb{V}}. \tag{E13}$$

For $n \geqslant n_0$, Eqs. (97) and (E13) yield

$$E\{\|[\mathbf{G}_{A_n}]^{-1}\| \|\dot{\mathbf{U}}_n(\tau)\|_{\mathbb{V}}\} \leqslant C_{\delta_A}^{1/2} |||\dot{\mathbf{U}}_n(\tau)|||_{\mathbb{V}}, \tag{E14}$$

in which C_{δ_A} is a finite real constant which is independent of n. Substituting the left-hand side of Eq. (E14) into the right-hand side of Eqs. (E10)–(E12) yields

$$|||\dot{\mathbf{U}}_n(t)|||_{\mathbb{H}}^2 \leqslant \frac{2C_{\delta_M}^{1/2}}{c_{\underline{m}}} \int_0^t \|\mathbf{f}(\tau)\|_{\mathbb{V}'} |||\dot{\mathbf{U}}_n(\tau)|||_{\mathbb{V}}\} d\tau, \tag{E15}$$

$$|||\dot{\mathbf{U}}_n(t)|||_{\mathbb{V}}^2 \leqslant \frac{2C_{\delta_K}^{1/2}}{c_{\underline{k}}} \int_0^t \|\mathbf{f}(\tau)\|_{\mathbb{V}'} |||\dot{\mathbf{U}}_n(\tau)|||_{\mathbb{V}}\} d\tau, \tag{E16}$$

$$\int_0^t |||\dot{\mathbf{U}}_n(\tau)|||_{\mathbb{V}}^2 d\tau \leqslant \frac{C_{\delta_D}^{1/2}}{c_{\underline{d}}} \int_0^t \|\mathbf{f}(\tau)\|_{\mathbb{V}'} |||\dot{\mathbf{U}}_n(\tau)|||_{\mathbb{V}}\} d\tau. \tag{E17}$$

If a, b, and η are three positive real numbers, we have

$$ab \leqslant \frac{a^2}{4\eta} + \eta b^2.$$

Applying this inequality to the right-hand sides of Eqs. (E15)–(E17) with $\eta = 0.5 c_{\underline{m}} C_{\delta_M}^{-1/2}$, $\eta = 0.5 c_{\underline{k}} C_{\delta_K}^{-1/2}$, and $\eta = 0.5 c_{\underline{d}} C_{\delta_D}^{-1/2}$, respectively, yields

$$|||\dot{\mathbf{U}}_n(t)|||_{\mathbb{H}}^2 \leqslant \int_0^t |||\dot{\mathbf{U}}_n(\tau)|||_{\mathbb{V}}^2\} d\tau + \frac{C_{\delta_M}}{c_{\underline{m}}^2} \int_0^T \|\mathbf{f}(\tau)\|_{\mathbb{V}'}^2 d\tau, \tag{E18}$$

$$|||\dot{\mathbf{U}}_n(t)|||_{\mathbb{V}}^2 \leqslant \int_0^t |||\dot{\mathbf{U}}_n(\tau)|||_{\mathbb{V}}^2\} d\tau + \frac{C_{\delta_K}}{c_{\underline{k}}^2} \int_0^T \|\mathbf{f}(\tau)\|_{\mathbb{V}'}^2 d\tau, \tag{E19}$$

$$\int_0^t |||\dot{\mathbf{U}}_n(\tau)|||_{\mathbb{V}}^2 d\tau \leqslant \frac{C_{\delta_D}}{c_{\underline{d}}^2} \int_0^T \|\mathbf{f}(\tau)\|_{\mathbb{V}'}^2 d\tau. \tag{E20}$$

Adding Eq. (E18) to Eq. (E20) yields

$$|||\dot{\mathbf{U}}_n(t)|||_{\mathbb{H}}^2 \leqslant \left(\frac{C_{\delta_M}}{c_{\underline{m}}^2} + \frac{C_{\delta_D}}{c_{\underline{d}}^2} \right) \int_0^T \|\mathbf{f}(\tau)\|_{\mathbb{V}'}^2 d\tau, \tag{E21}$$

and adding Eq. (E19) to Eq. (E20) yields

$$|||\mathbf{U}_n(t)|||_{\mathbb{V}}^2 \leqslant \left(\frac{C_{\delta_K}}{c_{\underline{k}}^2} + \frac{C_{\delta_D}}{c_{\underline{d}}^2} \right) \int_0^T \|\mathbf{f}(\tau)\|_{\mathbb{V}'}^2 d\tau. \tag{E22}$$

[1] R. W. Clough and J. Penzien, *Dynamics of Structures* (McGraw-Hill, New York, 1975).
[2] K. J. Bathe and E. L. Wilson, *Numerical Methods in Finite Element Analysis* (Prentice-Hall, New York, 1976).
[3] L. Meirovitch, *Computational Methods in Structural Dynamics* (Sijthoff and Noordhoff, The Netherlands, 1980).
[4] O. C. Zienkiewicz and R. L. Taylor, *The Finite Element Method*, 4th ed. (McGraw-Hill, New York, 1989) (Vol. 1, 1989 and Vol. 2, 1991).
[5] J. Argyris and H. P. Mlejnek, *Dynamics of Structures* (North-Holland, Amsterdam, 1991).
[6] Y. K. Cheung and A. Y. T. Leung, *Finite Element Methods in Dynamics* (Kluwer Academic, Dordrecht, 1992).
[7] M. Géradin and D. Rixen, *Mechanical Vibrations* (Wiley, Chichester, UK, 1994).
[8] R. Ohayon and C. Soize, *Structural Acoustics and Vibration* (Academic, San Diego, London, 1998).

[9] T. T. Soong, *Random Differential Equations in Science and Engineering* (Academic, New York, 1973).
[10] E. J. Haug, K. K. Choi, and V. Komkov, *Design Sensitivity Analysis of Structural Systems* (Academic, New York, 1986).
[11] R. A. Ibrahim, ''Structural dynamics with parameter uncertainties,'' Appl. Mech. Rev. **40**(3), 309–328 (1987).
[12] P. D. Spanos and B. A. Zeldin, ''Galerkin sampling method for stochastic mechanics problems,'' J. Eng. Mech. **120**(5), 1091–1106 (1994).
[13] W. D. Iwan and H. Jensen, ''On the dynamical response of continuous systems including model uncertainty,'' Trans. ASME, J. Appl. Mech. **60**, 484–490 (1993).
[14] C. Lee and R. Singh, ''Analysis of discrete vibratory systems with parameter uncertainties. II. Impulse response,'' J. Sound Vib. **174**(3), 395–412 (1994).
[15] Y. K. Lin and G. Q. Cai, *Probabilistic Structural Dynamics* (McGraw-Hill, New York, 1995).
[16] E. Vanmarcke and M. Grigoriu, ''Stochastic finite element analysis of simple beams,'' J. Eng. Mech. **109**(5), 1203–1214 (1983).
[17] W. K. Liu, T. Belytschko, and A. Mani, ''Random field finite elements,'' Int. J. Numer. Methods Eng. **23**, 1832–1845 (1986).
[18] M. Shinozuka and G. Deodatis, ''Response variability of stochastic finite element systems,'' J. Eng. Mech. **114**(3), 499–519 (1988).
[19] P. D. Spanos and R. G. Ghanem, ''Stochastic finite element expansion for random media,'' ASCE J. Eng. Mech. **115**(5), 1035–1053 (1989).
[20] R. G. Ghanem and P. D. Spanos, *Stochastic Finite Elements: A Spectral Approach* (Springer, New York, 1991).
[21] M. Kleiber, D. H. Tran, and T. D. Hien, *The Stochastic Finite Element Method* (Wiley, New York, 1992).
[22] Y. K. Lin, *Probabilistic Theory of Structural Dynamics* (McGraw-Hill, New York, 1967) (Reprint: Krieger, Melbourne, FL, 1976).
[23] R. A. Ibrahim, *Parametric Random Vibration* (Wiley, New York, 1985).
[24] P. Kree and C. Soize, *Mathematics of Random Phenomena* (Reidel, Dordrecht, 1986).
[25] J. B. Roberts and P. D. Spanos, *Random Vibration and Statistical Linearization* (Wiley, New York, 1990).
[26] T. T. Soong and M. Grigoriu, *Random Vibration of Mechanical and Structural Systems* (Prentice-Hall, Englewood Cliffs, NJ, 1993).
[27] C. Soize, *The Fokker–Planck Equation for Stochastic Dynamical Systems and its Explicit Steady State Solutions* (World Scientific, Singapore, 1994).
[28] C. Soize, ''A nonparametric model of random uncertainties in linear structural dynamics,'' Publication du LMA-CNRS, ISBN 2-909669-16-5, **152**, 109–138 (1999).
[29] C. Soize, ''A nonparametric model of random uncertainties for reduced matrix models in structural dynamics,'' Probab. Eng. Mech. **15**(3), 277–294 (2000).
[30] Y. C. Fung, *Foundations of Solid Mechanics* (Prentice-Hall, Englewood Cliffs, NJ, 1968).
[31] C. Truesdell, *Mechanics of Solids, Vol. III, Theory of Viscoelasticity, Plasticity, Elastic Waves and Elastic Stability* (Springer, Berlin, 1984).
[32] P. G. Ciarlet, *Mathematical Elasticity, Vol. I: Three-Dimensional Elasticity* (North-Holland, Amsterdam, 1988).
[33] R. Dautray and J.-L. Lions, *Mathematical Analysis and Numerical Methods for Science and Technology* (Springer, Berlin, 1992).
[34] M. Reed and B. Simon, *Methods of Modern Mathematical Physics. I. Functional Analysis* (Academic, New York, 1980).
[35] F. Treves, *Basic Linear Partial Differential Equations* (Academic, New York, 1975).
[36] J. E. Marsden and T. J. R. Hughes, *Mathematical Foundations of Elasticity* (Prentice-Hall, Englewood Cliffs, NJ, 1983).
[37] J. T. Oden and J. N. Reddy, *Variational Methods in Theoretical Mechanics* (Springer, New York, 1983).
[38] M. L. Mehta, *Random Matrices*, Revised and enlarged 2nd ed. (Academic, New York, 1991).
[39] P. A. Deift, *Orthogonal Polynomials and Random Matrices: A Riemann-Hilbert Approach* (Courant Institute, New York University, 1999).
[40] N. M. Katz and P. Sarnak, *Random Matrices, Frobenius Eigenvalues and Monodromy* (American Mathematical Society, Providence, RI, 1999).
[41] D. Fox and P. B. Kahn, ''Higher order spacing distributions for a class of unitary ensembles,'' Phys. Rev. **134**(5B), 1151–1155 (1964).
[42] M. V. Moorthy, ''Inversion of the multi-dimensional Laplace transform-expansion by Laguerre series,'' Z. Angew. Math. Phys. **46**, 792–806 (1995).
[43] T. Nagao and K. Slevin, ''Laguerre ensembles of random matrices: Nonuniversal correlation functions,'' J. Math. Phys. **34**(6), 2317–2330 (1993).
[44] C. E. Shannon, ''A mathematical theory of communication,'' Bell Syst. Tech. J. **27**, 379–423 (1948); **27**, 623–659 (1948).
[45] E. T. Jaynes, ''Information theory and statistical mechanics,'' Phys. Rev. **106**(4), 620–630 (1957); **108**(2), 171–190 (1957).
[46] J. N. Kapur and H. K. Kesavan, *Entropy Optimization Principles with Applications* (Academic, San Diego, 1992).
[47] C. Fougeaud and A. Fuchs, *Statistique* (Dunod, Paris, 1967) (2nd ed. 1972).
[48] T. W. Anderson, *Introduction to Multivariate Statistical Analysis* (Wiley, New York, 1958).
[49] J. M. Hammersley and D. C. Handscomb, *Monte Carlo Methods* (Chapman and Hall, London, 1964).
[50] W. G. Cochran, *Sampling Techniques* (Wiley, New York, 1977).
[51] R. Y. Rubinstein, *Simulation and the Monte Carlo Method* (Wiley, New York, 1981).
[52] L. Devroye, *Non-Uniform Random Variate Generation* (Springer, New York, 1986).
[53] M. H. Kalos and P. A. Whitlock, *Monte Carlo Methods, Volume 1: Basics* (Wiley, New York, 1986).
[54] P. Bratley, B. L. Fox, and E. L. Schrage, *A Guide to Simulation*, 2nd ed. (Springer, New York, 1987).
[55] P. K. MacKeown, *Stochastic Simulation in Physics* (Springer, Singapore, 1997).
[56] K. Johansson, ''On fluctuations of eigenvalues of random Hermitian matrices,'' Duke Math. J. **91**(1), 151–204 (1998).

The matched-lag filter: Detecting broadband multipath signals with auto- and cross-correlation functions

John L. Spiesberger
Department of Earth and Environmental Science, 158 Hayden Hall, University of Pennsylvania, 240 South 33rd Street, Philadelphia, Pennsylvania 19104-6316

(Received 8 September 1999; accepted for publication 12 December 2000)

Signal detection is considered for uncertain noise variance and a broadband source of unknown waveform and emission time. The signal travels to the receivers along paths with unknown delays. Using a new ''matched-lag filter,'' the presence or absence of the signal is estimated from the auto- and cross-correlation functions of the receptions. Like a matched filter, correlation functions provide the first stage of gain in signal-to-noise ratio because the paths are assumed to be partially coherent. The second stage achieves additional gain by searching only over physically possible arrangements of signals in the auto- and cross-correlation functions while excluding forbidden arrangements. These stages enable the matched-lag filter to behave like a matched filter within a matched filter. In an ideal case, simulations of the matched-lag filter yield probabilities of detection that are, with one and two receivers, 4.1 and 366 times, respectively, that obtained from the conventional energy detector at a false-alarm probability of 0.001. The matched-lag filter has applications to wireless communications and the detection of acoustic signals from animals, vehicles, ships, and nuclear blasts. The matched-lag filter more completely describes signal structure than stochastic detection and communication theories whose specified auto-correlation function does not prohibit forbidden arrangements. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1350451]

PACS numbers: 43.60.Cg, 43.60.Gk, 43.60.Dh [JCB]

I. INTRODUCTION

A theory is developed for detecting broadband signals that travel to one or more receivers along an unknown number of paths. Some paths are assumed to be temporally resolved, so that their inverse bandwidths exceed some of their relative arrival times. The signals have unknown emission times, amplitudes, and waveforms. A reliable model for the propagation of energy need not exist or be used, though a model can be incorporated into the detector. It is assumed that some temporally resolved signals are partially coherent. Many broadband acoustic emissions from ships, vehicles, nuclear blasts, animals, underwater sources, and acoustic and electromagnetic communication transmitters appear to be of the type for which the theory is applicable. Under ideal circumstances, simulations of the theory yield significantly larger probabilities of detection than obtained from the conventional energy detector. In this paper, the word ''detection'' is defined to be a process of deciding if data consist of signal and noise or just noise.

From 1954 to present, some detection and communication theories treated multipath receptions stochastically by requiring the received auto-correlation functions of the signal and noise to have specified shapes (e.g., Refs. 1–6). With temporally unresolved paths, these approaches appear to offer a reasonable approach for modeling the data. However, when many paths are temporally *resolved*, these stochastic approaches are incomplete because signal-lags cannot occur at arbitrary separations, as shown below. Indeed, stochastic theories contain much less information about the structure of signals in correlation functions than the deterministic theory developed here.

The deterministic theory provides an organized means for dealing with data from all correlation functions. An auto-correlation function is used for one receiver, and all the auto- and cross- correlation functions are used when one has two or more receivers. Correlation functions offer a natural starting point because they provide the gain obtained with a matched filter.[7] With more than one receiver, it is interesting that stochastic detection and communication theories[1–6] do not address the use of auto- *and* cross-correlation functions, but instead utilize the auto-correlation functions from each receiver.

The deterministic structure of signals in correlation functions can be motivated by describing forbidden arrangements. Consider three paths arriving at a receiver. There are at most $3(3-1)/2=3$ signal-related peaks at positive lags in the auto-correlation function.[8] The lags of these peaks must satisfy ''lag-equations,[8]''

$$\tau_{ii}(m,n)=t_i(m)-t_i(n); \quad m>n\geq 1, \tag{1}$$

where the travel time for path m at receiver i is $t_i(m)$. It is impossible to have lags 2, 4, and 5 simultaneously occupied by signals where sample times are counted by their sample number. To see this, the difference in sample number between the third and first paths must be 5, to yield a signal at lag 5. This leaves one additional path which could arrive at sample numbers 1, 2, 3, or 4. None of these possibilities yields signals at lags 2, 4, and 5 (Table I). Allowed signal arrangements between auto- and cross-correlation functions have been described.[9] A ''matched-lag'' filter is defined to be a detector that utilizes only the allowed signal arrangements in correlation functions.[10] The number of allowed to forbidden arrangements turns out to be very small, and this intuitively explains why the matched-lag filter does well in

TABLE I. These are the lags at which signals occur in the auto-correlation function assuming there are three paths and the sample numbers at which paths one and three arrive are zero and five, respectively. The sample number at which the second path arrives, $t(2)$, determines the four allowed arrangements of signal lags. Note that the arrangement, 2, 4, and 5 never occurs. Thus, it is a forbidden arrangement.

$t(2)$	Occupied lags		
1	1	4	5
2	2	3	5
3	2	3	5
4	1	4	5

simulations. Stochastic detection and communication approaches, such as in Refs. 1–6, are incomplete because the specification of an auto-correlation function does not exclude forbidden signal arrangements (Sec. XI).

II. THE MATCHED-LAG FILTER: GENERAL FORMULATION

There are $\mathcal{R}(\mathcal{R}+1)/2$ auto- and cross-correlation functions from $\mathcal{R}$ receivers.[8] The time series of K data at receiver j is

$$\mathbf{r}_j, \quad j=1,2,3,\cdots,\mathcal{R},$$

where $\mathbf{r}_j$ is a $K\times 1$ vector. The correlation functions at nonnegative lags are

$$R_{ij}(p)\equiv\frac{1}{K}\sum_{k=1+p}^{K} r_i(k)r_j(k-p); \quad 1\leqslant i\leqslant j\leqslant\mathcal{R}; \quad p\geqslant 0, \tag{2}$$

where, for cross-correlation functions, k goes from 1 to $K+p$ for p less than zero and $i\neq j$. The kth row of $\mathbf{r}_j$ is $r_j(k)$. The correlation data are put in a vector,

$$\mathbf{R}\equiv[\mathbf{R}_{ij}], \quad 1\leqslant i\leqslant j\leqslant\mathcal{R}, \tag{3}$$

where only nonnegative lags are used from auto-correlation functions. Under hypothesis H_0, the data for the matched-lag filter are due to noise only and have joint density function $f(\mathbf{R}|H_0)$. The conditional probability of X given Y is denoted by $f(X|Y)$.

Joint density function for the correlation functions under hypothesis H_1

Under hypothesis H_1, signal and noise are present. The signal bearing lags in the auto- and cross-correlation functions have a deterministic structure given by Eq. (1) and Eq. (9) in Ref. 9. The equations are exact when there is (1) no interference between paths leading to a correlation peak being shifted in lag-space, and (2) no lack of coherence between paths in the auto- and cross-correlation functions. In practice, the lag equations are useful when there is interference and degradation of coherence as long as the correlations between paths lead to peaks near the lags specified in the lag-equations.[9,11]

Let $Q(i)$ be the upper bound for the difference in travel times between the last and first paths at receiver i. Further assume that the travel times coincide with a sample time at the receiver. Then one generates all the allowed arrangements of signals in auto-correlation functions by synthesizing all ways of arranging N_i-1 relative travel times,

$$t_i(m)-t_i(1), \quad m=2,3,4,\cdots,N_i \tag{4}$$

without replacement, among $Q(i)$ travel time bins where N_i is the number of paths at receiver i. There are,

$$B_i(N_i)\equiv\begin{pmatrix} Q(i) \\ N_i-1 \end{pmatrix}, \tag{5}$$

such arrangements. The b_ith arrangement of N_i-1 relative travel times determines the $P_i'(b_i,N_i)$ positive lags,

$$\psi_i(p|b_i,N_i); \quad p=1,2,3,\cdots,P_i'(b_i,N_i), \tag{6}$$

where signals occur in the auto-correlation function. The number of signal-related lags, $P_i'(b_i,N_i)$, satisfies

$$N_i-1\leqslant P_i'(b_i,N_i)\leqslant P_i(N_i), \tag{7}$$

where the maximum number of resolved signals at positive lag is $P_i(N_i)$ [Eq. (27) in Ref. 8]. One obtains less than $P_i(N_i)$ resolved signal-lags when two or more pairs of travel time differences are the same.

Arrangements b_i and b_j do not uniquely specify the difference in travel times between the first paths, $\tau_{ij}(1,1)=t_i(1)-t_j(1)$. The maximum time required for energy to propagate between the receivers is $D_{ij}=d_{ij}/c_{ij}$ where c_{ij} is the minimum speed at which energy propagates, and d_{ij} is the distance between the receivers. Since we know that $-D_{ij}\leqslant\tau_{ij}(1,1)\leqslant+D_{ij}$, there are $2D_{ij}+1$ possible lags in which $\tau_{ij}(1,1)$ can occur in the cross- correlation function. We only need to specify $\tau_{1j}(1,1)$ for $j=2,3,4,\cdots,\mathcal{R}$ to determine the lags, $\tau_{ij}(1,1)$, $2\leqslant i<j\leqslant\mathcal{R}$, because $\tau_{ij}(1,1)=\tau_{1j}(1,1)-\tau_{1i}(1,1)$.

The joint probability density function of the data is

$$f(\mathbf{R}|H_1)=\sum_{\mathbf{b},\boldsymbol{\tau}_{1j}(1,1),\mathbf{N}} f(\mathbf{R}|\mathbf{b},\boldsymbol{\tau}_{1j},\mathbf{N})f(\mathbf{b},\boldsymbol{\tau}_{1j}(1,1),\mathbf{N}), \tag{8}$$

where $\mathbf{b}=(b_1,b_2,b_3,\cdots,b_{\mathcal{R}})$, $\boldsymbol{\tau}_{1j}(1,1)=(\tau_{12}(1,1),\tau_{13}(1,1),\cdots,\tau_{1\mathcal{R}}(1,1))$ and $\mathbf{N}=(N_1,N_2,N_3,\cdots,N_{\mathcal{R}})$. The sum is over all possible combinations of the vectors where the signal arrangements are specified by Eq. (1) above and Eq. (9) in Ref. 9. The number of arrangements is finite because there are upper and lower bounds for the number of significant paths at each receiver. The filter could be written for more than one source by including sums on the number of different sources with their conditional probabilities. This is too cumbersome for an introductory paper.

III. THE MATCHED-LAG FILTER WITH AN ANALYTICAL SOLUTION: AN IDEAL CASE AND STARTING ASSUMPTIONS

An average likelihood ratio,[5] $\bar{\Lambda}=\overline{f(\mathbf{R}|H_1)}\times(\overline{f(\mathbf{R}|H_0)})^{-1}$, is used to account for uncertainties in the probability density function of the noise. An overline denotes an average over the noise's uncertainty. The ratio has an analytical form for an ideal case. This form is useful for getting a feel for the potential of using this receiver in other situations. The ratio will be evaluated for one and two re-

ceivers. When signals are present, it is assumed that the number of paths at each receiver, N, is a known constant. We assume that $Q(i)=Q$, $\forall i$ and $D_{12}=D$.

Under hypothesis H_0, the $k=1,2,3,\cdots,K$ measurements, $r_j(k)=e_j(k)$, at receivers $j=1,2$ are assumed to be mutually uncorrelated Gaussian random variables, $e_j(k)$, with mean zero and imperfectly known variance, ρ^2. The noise between receivers is assumed to be uncorrelated.

Under hypothesis H_1, the likelihood ratio is simple when the measurements contain N delayed and attenuated replicas of an emitted signal, $s(k)$, plus additive noise,

$$r_j(k)=\sum_{n=1}^{N} as(k-t_j(n))+e_j(k), \quad j=1,2. \tag{9}$$

We assume, for simplicity, that the amplitude, a, of each path is the same. Travel time is measured in units of the sample number at the receiver. To maintain simplicity, the likelihood ratio is developed for correlation lags p obeying $|p|\ll K$ and $K\gg 1$. Since the largest possible difference in travel time between the first and last paths is Q (Sec. II), $|p|\leqslant Q$.

To preserve mathematical simplicity, the average likelihood ratio is evaluated in an ideal situation where the emitted signal's auto-correlation function is a delta-function,

$$\sum_{k=1}^{K} s(k)s(k-p)\equiv \mathcal{E}\delta(p), \tag{10}$$

where $\delta(p)$ is one when p is zero and is zero otherwise. The energy in the emitted signal is

$$\mathcal{E}\equiv\sum_{k=1}^{K} s^2(k). \tag{11}$$

This approximation is most suitable for situations where the auto-correlation function has a known shape with small sidelobes. In this approximation, interference between paths does not change the lags at which the signal peaks occur in the correlation functions. The first paragraph of Sec. XI describes an approach for implementing the matched-lag filter for some real signal shapes.

IV. JOINT DENSITY FUNCTIONS UNDER HYPOTHESIS H_0: THE IDEAL CASE

For the noise-only case and for $|p|\ll K$, the variance of the positive lags in the auto-correlation function, and the variance of the lags in the cross-correlation function is (Appendix A)

$$\sigma_0^2\equiv\overline{(R_{ij}(p)-\overline{R_{ij}(p)})^2}\cong\frac{(\rho^2)^2}{K}. \tag{12}$$

The joint density function of the first Q positive lags in the auto-correlation function is (Appendix A)

$$f(\mathbf{R}_{ii}|H_0)=f(\mathbf{R}_{ii}|\sigma_0^2,H_0)f(\sigma_0^2); \quad i=1,2, \tag{13}$$

where the density,

$$f(\mathbf{R}_{ii}|\sigma_0^2,H_0)=(2\pi\sigma_0^2)^{-Q/2}\exp\left[-\frac{1}{2\sigma_0^2}\sum_{q=1}^{Q}R_{ii}^2(q)\right];$$
$$i=1,2, \quad Q\ll K, \tag{14}$$

describes a mutually uncorrelated zero mean Gaussian random variable. The probability for the variance of the noise is $f(\sigma_0^2)$.

Similarly, the joint density of the cross-correlation function is $f(\mathbf{R}_{12}|H_0)=f(\mathbf{R}_{12}|\sigma_0^2,H_0)f(\sigma_0^2)$ where the density is (Appendix A)

$$f(\mathbf{R}_{12}|\sigma_0^2,H_0)=(2\pi\sigma_0^2)^{-L/2}\exp\left[-\frac{1}{2\sigma_0^2}\sum_{l=-L_1}^{L_1}R_{12}^2(l)\right]. \tag{15}$$

The number of lags used from the cross-correlation function is $L=2(Q+D)+1=2L_1+1$ where the most negative and positive lags are $-L_1$ and L_1, respectively, and $L_1\ll K$.

The filter uses all the data from the two auto-correlation functions (except lag zero for now) and the cross-correlation function. The joint density of these three correlation functions is (Appendix A)

$$f(\mathbf{R}_{11},\mathbf{R}_{22},\mathbf{R}_{12}|H_0)=f(\mathbf{R}_{11}|\sigma_0^2,H_0)f(\mathbf{R}_{22}|\sigma_0^2,H_0)$$
$$\times f(\mathbf{R}_{12}|\sigma_0^2,H_0)f(\sigma_0^2). \tag{16}$$

Average density functions

For simplicity, a variable, $x\equiv\sigma_0^{-2}$, is introduced which is assumed to be uniformly distributed as $f(x)=(\hat{x}-\check{x})^{-1}$ on the interval $\check{x}\leqslant x\leqslant\hat{x}$. This parameterization is not made to necessarily coincide with a realistic case, but rather to explore the behavior of the filter when the noise's variance is uncertain.

Substituting x into Eq. (13) yields $f(\mathbf{R}_{ii}|H_0)=f(\mathbf{R}_{ii}|x,H_0)f(x)$. Its average is

$$\overline{f(\mathbf{R}_{ii}|H_0)}=\int_{\check{x}}^{\hat{x}}f(\mathbf{R}_{ii}|x,H_0)f(x)dx; \quad i=1,2,$$
$$=\frac{1}{(\hat{x}-\check{x})}\frac{1}{(2\pi)^{Q/2}}\mu_{\mathbf{R}_{ii}}^{-(Q/2+1)}$$
$$\times\left[\gamma\left(\frac{Q}{2}+1,\mu_{\mathbf{R}_{ii}}\hat{x}\right)-\gamma\left(\frac{Q}{2}+1,\mu_{\mathbf{R}_{ii}}\check{x}\right)\right], \tag{17}$$

where $\mu_{\mathbf{R}_{ii}}\equiv\frac{1}{2}\Sigma_{l=1}^{Q}R_{ii}^2(l)$ and the incomplete gamma function[12] is $\gamma(a,b)\equiv\int_0^b e^{-t}t^{a-1}\,dt$.

The average value of the density from one auto- and two cross-correlation functions is [Eq. (16)],

$$\overline{f(\mathbf{R}_{11},\mathbf{R}_{22},\mathbf{R}_{12}|H_0)}$$
$$\equiv\int_{\check{x}}^{\hat{x}}f(\mathbf{R}_{11}|x,H_0)f(\mathbf{R}_{22}|x,H_0)f(\mathbf{R}_{12}|x,H_0)f(x)dx$$
$$=\frac{1}{(\hat{x}-\check{x})}\frac{1}{(2\pi)^{Q+\frac{L}{2}}}\mu_1^{-(Q+L/2+1)}$$
$$\times\left[\gamma\left(Q+\frac{L}{2}+1,\mu_1\hat{x}\right)-\gamma\left(Q+\frac{L}{2}+1,\mu_1\check{x}\right)\right], \tag{18}$$

where $\mu_1 \equiv \frac{1}{2}(\{\Sigma_{p=1}^{Q} R_{11}^2(p) + R_{22}^2(p)\} + \Sigma_{n=-L_1}^{L_1} R_{12}^2(n))$.

V. JOINT DENSITY FUNCTIONS FOR AUTO-CORRELATION FUNCTIONS UNDER HYPOTHESIS H_1: THE IDEAL CASE

A. Allowed arrangements of signals

Since N paths arrive at each receiver, the number of allowed signal arrangements, $B_i(N)$, in each auto-correlation function is the same [Eq. (5)]. The b_ith arrangement is characterized by the lags, $\psi_i(p|b_i,N)$; $p=1,2,3,\cdots,P_i'(b_i,N)$, at which signals occur and their redundancy at each lag, $\eta_i(p|b_i,N)$, $p=1,2,3,\cdots,P_i'(b_i,N)$ where $\eta_i(p|b_i,N) \geqslant 1$. The b_ith allowed lag-redundancy set contains $2P_i'(b_i,N)$ elements, $\{\psi_i(p|b_i,N), \eta_i(p|b_i,N)\}$, $p=1,2,3,\cdots,P_i'(b_i,N)$.

Some arrangements are the same, because there are two sets of different relative travel times that yield the same arrangement. For example, the two travel time sets, $t_i \in \{0,1,3\}$ and $t_i \in \{0,2,3\}$, both yield the same positive auto-correlation function lags 1, 2, and 3, and redundancy functions, $\eta(p)=1, 2$, and 1 for $p=1, 2$, and 3, respectively. Travel time sets that are time reversals of one another generate the same lag-redundancy set. This can be proven by computing the auto-correlation function for the travel times $t_i(n)$, $n=1,2,3,\cdots,N$, which have signals at positive lags given by Eq. (1). The time reversals, $-t_i(n)$, $n=1,2,3,\cdots,N$, yield signals at lags, $\tau_{ii}(m,n) = -t_i(m) - (-t_i(n)) = -(t_i(m) - t_i(n))$; $N \geqslant m > n \geqslant 1$, which are opposites of those in Eq. (1). One then notes that auto-correlation function lags are symmetric about zero. As Q increases, the ratio of the number symmetrical to unsymmetrical sets gets small (Appendix B).

B. Probability density functions for auto-correlation functions

A resolved signal has a value of

$$A = \frac{a^2 \mathcal{E}}{K} \tag{19}$$

in the auto- and cross-correlation functions (Appendix A). When signal and noise are present, the variance of the noise and reverberation in an auto-correlation function is

$$\sigma_1^2 = \sigma_0^2 \left[1 + 2\left(1 + \frac{N-1}{4Q} \right) l^2 \right], \tag{20}$$

where the time-averaged signal-to-noise intensity ratio over K samples at the receiver is

$$l^2 = \frac{AN}{\rho^2} \tag{21}$$

(Appendix A). The covariance of the nonsignal component is negligible compared with the variance, σ_1^2, when, as we assume, the time-averaged intensity signal-to-noise ratio is small at the receiver, so $l^2 \ll 1$ (Appendix A). The nonsignal components in the auto-correlation function may then be approximated as mutually uncorrelated Gaussian random variables with zero means and variances σ_1^2.

The joint density of the positive lags in auto-correlation function i is

$$f(\mathbf{R}_{ii}|H_1) = \sum_{b_i=1}^{B_i(N)} f(\mathbf{R}_{ii}|b_i,\sigma_1^2) f(b_i) f(\sigma_1^2), \tag{22}$$

where the probability is

$$\begin{aligned} f(\mathbf{R}_{ii}|b_i,\sigma_1^2) &= (2\pi\sigma_1^2)^{-Q/2} \\ &\times \exp\left[-\frac{1}{2\sigma_1^2} \sum_{p=1}^{P_i'(b_i,N)} (R_{ii}(\psi_i(p|b_i,N)) \right. \\ &- \eta_i(p|b_i,N)A)^2 - \frac{1}{2\sigma_1^2} \\ &\times \left. \sum_{p=P_i'(b_i,N)+1}^{Q} R_{ii}^2(\Psi_i(p|b_i,N)) \right]; \quad i=1,2, \end{aligned} \tag{23}$$

and where $\Psi_i(p|b_i,N)$ are the lags where *no* signals occur at positive lag. The probability of obtaining the b_ith allowed arrangement of signals will be taken to be the one of most ignorance, so it is uniformly distributed as $f(b_i) = (B_i(N))^{-1}$. Thus, the $B_i(N)$ travel time sets (Eq. (4)) are equally likely.

Average density functions for auto-correlation functions

In analogy to Sec. IV, we define $y \equiv \sigma_1^{-2}$ to be uniformly distributed on the interval $\check{y} \leqslant y \leqslant \hat{y}$ with density $f(y) = (\hat{y} - \check{y})^{-1}$. The average of Eq. (22) is

$$\begin{aligned} \overline{f(\mathbf{R}_{ii}|H_1)} &\equiv \int_{\check{y}}^{\hat{y}} f(y) f(\mathbf{R}_{ii}|H_1)\, dy \\ &= \frac{f(y)}{(2\pi)^{Q/2}} \frac{1}{B_i(N)} \sum_{b_i=1}^{B_i(N)} \mu_{b_i}^{-(Q/2+1)} \\ &\times \left[\gamma\left(\frac{Q}{2}+1, \mu_{b_i}\hat{y} \right) - \gamma\left(\frac{Q}{2}+1, \mu_{b_i}\check{y} \right) \right], \end{aligned} \tag{24}$$

where

$$\begin{aligned} \mu_{b_i} \equiv \frac{1}{2} \sum_{l=1}^{Q} R_{ii}^2(l) + \frac{A^2}{2} \sum_{p=1}^{P_i'(b_i,N)} \eta_i(p|b_i,N)^2 \\ - A \sum_{p=1}^{P_i'(b_i,N)} \eta_i(p|b_i,N) R_{ii}(\psi_i(p|b_i,N)). \end{aligned} \tag{25}$$

VI. JOINT DENSITY FUNCTIONS FOR TWO AUTO- AND ONE CROSS-CORRELATION FUNCTION UNDER HYPOTHESIS H_1: THE IDEAL CASE

A. Modeling the noise and reverberation

The variance of the noise plus reverberation in the cross-correlation function is (Appendix A)

$$\alpha_1^2 = \sigma_0^2(1 + 2l^2). \tag{26}$$

We see that the variances of the noise plus reverberation in the auto- and cross-correlation functions are approximately the same when $N-1 \ll 4Q$ [Eqs. (20), (26)]. A simple approximation is made so that these variances are approximated by their weighted average, $\beta^2 \equiv (2Q\sigma_1^2 + L\alpha_1^2)(2Q+L)^{-1}$. This weighting is chosen because there are $2Q$ positive auto-correlation lags and L cross-correlation lags.

The nonsignal components in the correlation functions are approximated as $2Q+L$ mutually uncorrelated Gaussian random variables with mean zero and variance β^2. This is valid for our assumed case of l^2 small [Eqs. (21), (A7), (A8)].

B. Allowed arrangements

The number of lags that contain a signal in the the cross-correlation function is denoted by $M(b_i,b_j) \leqslant N^2$, which depends on the arrangements of relative travel times at the two receivers. The maximum value of N^2 occurs when all the signal-related lags are resolved. The lags at which they occur are $\tau_{12}(1,1)+\phi_{ij}(p|b_i,b_j)$, $p=1,2,3,\cdots,M(b_i,b_j)$ where $-D \leqslant \tau_{12}(1,1) \leqslant D$. The relative travel times of the paths are assumed to be statistically independent at the two receivers, so all the allowed arrangements are equally likely.

C. Density function

The joint density of the data in all the correlation functions is [Eq. (8)]

$$f(\mathbf{R}|H_1) = \sum_{i=1}^{B_i(N)} f(\mathbf{R}_{11}|b_i,\beta^2,H_1) \sum_{j=1}^{B_j(N)} f(\mathbf{R}_{22}|b_j,\beta^2,H_1) \times \sum_{\tau_{12}(1,1)=-D}^{D} f(\mathbf{R}_{12}|b_i,b_j,\tau_{12}(1,1),\beta^2,H_1) \times f(b_i)f(b_j)f(\tau_{12}(1,1))f(\beta^2). \quad (27)$$

Note that the density functions for the two auto-correlation functions are independent except for their condition on β^2. We define the density function of $\tau_{12}(1,1)$ to be the one of least information, so it is uniformly distributed as $f(\tau_{12}(1,1))=(2D+1)^{-1}$.

The density functions for the auto-correlation functions are given by Eq. (22), except σ_1^2 is replaced with β^2. Finally, the conditional density of the cross-correlation function is

$$f(\mathbf{R}_{12}|b_i,b_j,\tau_{12}(1,1),\beta^2,H_1) = (2\pi\beta^2)^{-L/2}\exp\left[\frac{-1}{2\beta^2}\sum_{n=1}^{M(b_i,b_j)}(R_{12}(\tau_{12}(1,1)+\phi_{ij}(n|b_i,b_j)) - \zeta_{ij}(n|b_i,b_j)A)^2 - \frac{1}{2\beta^2}\sum_{n=M(b_i,b_j)+1}^{L} R_{12}^2(\Phi_{ij}(n|b_i,b_j,\tau_{12}(1,1)))\right], \quad (28)$$

where the number of signal pairs at relative lag $\phi_{ij}(n|b_i,b_j)$ in the cross-correlation function is $\zeta_{ij}(n|b_i,b_j) \geqslant 1$ and the lags *without* signals in the cross- correlation function are $\Phi_{ij}(n|b_i,b_j,\tau_{12}(1,1))$; $n=M(b_i,b_j)+1,\cdots,L$.

D. Average density function

As before, the variance of the noise is parameterized as $w \equiv \beta^{-2}$ with uniform density, $f(w)=(\hat{w}-\check{w})^{-1}$, on the interval $\check{w} \leqslant w \leqslant \hat{w}$. Then the average density of the positive lags in the two auto-correlation functions and the L lags in the cross-correlation function is

$$\overline{f(\mathbf{R}|H_1)} = \int_{\check{w}}^{\hat{w}} f(\mathbf{R}|H_1)f(w)dw = \frac{f(w)}{(2\pi)^{Q+(L/2)}}\frac{1}{B_i(N)B_j(N)(2D+1)} \times \sum_{b_i=1}^{B_i(N)}\sum_{b_j=1}^{B_j(N)}\sum_{\tau_{12}(1,1)=-D}^{D}\mu_{ij}(b_i,b_j,\tau_{12}(1,1))^{-(Q+(L/2)+1)} \times\left[\gamma\left(Q+\frac{L}{2}+1,\mu_{ij}(b_i,b_j,\tau_{12}(1,1))\hat{w}\right) - \gamma\left(Q+\frac{L}{2}+1,\mu_{ij}(b_i,b_j,\tau_{12}(1,1))\check{w}\right)\right], \quad (29)$$

where

$$\begin{aligned}\mu_{ij}(b_i,b_j,\tau_{12}(1,1)) &\equiv \mu_1 - A\sum_{p=1}^{P_i'(b_i,N)}\eta_i(p|b_i,N)R_{11}(\psi_i(p|b_i,N)) \\ &\quad - A\sum_{p=1}^{P_j'(b_j,N)}\eta_j(p|b_j,N)R_{22}(\psi_j(p|b_j,N)) \\ &\quad - A\sum_{p=1}^{M_{ij}(b_i,b_j)}(\zeta_{ij}|b_i,b_j)R_{12}(\tau_{12}(1,1)+\phi_{ij}(p|b_i,b_j)) \\ &\quad + \frac{A^2}{2}\left(\sum_{p=1}^{P_i'(b_i,N)}\eta_i^2(p|b_i,N) + \sum_{p=1}^{P_j'(b_j,N)}\eta_j^2(p|b_j,N) + \sum_{p=1}^{M_{ij}(b_i,b_j)}\zeta_{ij}^2(p|b_i,b_j)\right). \end{aligned} \quad (30)$$

VII. AVERAGE LIKELIHOOD RATIOS FOR THE MATCHED-LAG FILTER: THE IDEAL CASE

With one receiver, the average likelihood ratio for the filter is

$$\overline{\Lambda_1} = \overline{f(\mathbf{R}_{ii}|H_1)}\,(\overline{f(\mathbf{R}_{ii}|H_0)})^{-1}, \quad (31)$$

where the numerator and denominator come from Eqs. (24) and (17), respectively. Similarly, the average likelihood ratio with two receivers is

$$\overline{\Lambda_2} = \overline{f(\mathbf{R}|H_1)}\,(\overline{f(\mathbf{R}_{11},\mathbf{R}_{22},\mathbf{R}_{12}|H_0)})^{-1}, \quad (32)$$

where the numerator and denominator come from Eqs. (29) and (18), respectively.

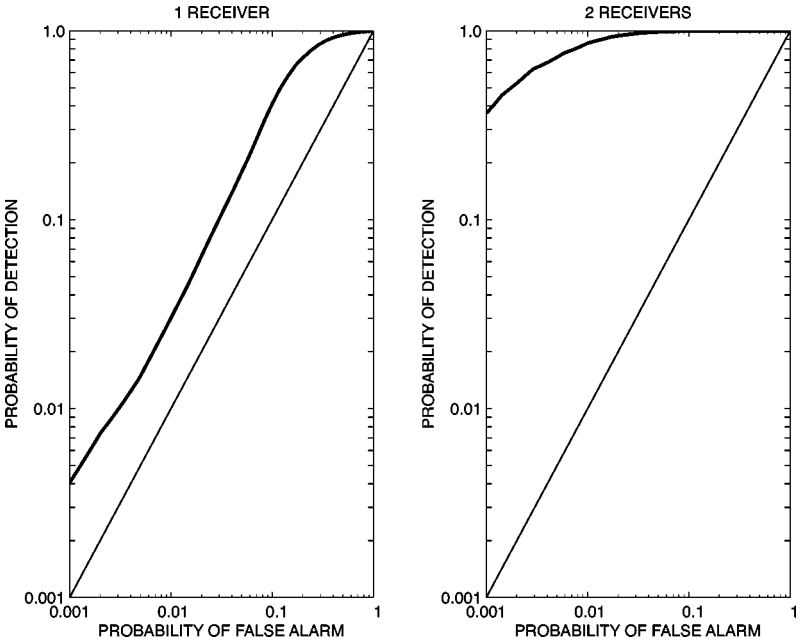


FIG. 1. **Left**: Receiver operating curves (Refs. 5, 13–15) for the energy detector (thin line) and the matched-lag filter (thick line) using the values in Eq. (38) for one receiver. **Right**: Same except this is for two receivers.

VIII. AVERAGE LIKELIHOOD RATIOS FOR A CONVENTIONAL RECEIVER: THE ENERGY DETECTOR

The conventional alternative to the matched-lag filter is the energy detector.[13] Suppose Z mutually uncorrelated Gaussian random variables, $\mathbf{z}$, with mean zero and variance ϵ_0^2 are measured at a receiver under hypothesis H_0. Under hypothesis H_1, the Z samples consist of the same kind of random variable except the variance is ϵ_1^2. Suppose both variances are uncertain. Under hypothesis H_0, the average density function of the data is

$$\overline{f(\mathbf{z}|H_0)}=\int_{\check{\epsilon}_0^2}^{\hat{\epsilon}_0^2}\frac{1}{(2\pi\epsilon_0^2)^{Z/2}}\exp\left[-\frac{1}{2\epsilon_0^2}\sum_{i=1}^{Z}z^2(i)\right]f(\epsilon_0^2)d\epsilon^2, \tag{33}$$

and under hypothesis H_1 the average density is

$$\overline{f(\mathbf{z}|H_1)}=\int_{\check{\epsilon}_1^2}^{\hat{\epsilon}_1^2}\frac{1}{(2\pi\epsilon_1^2)^{Z/2}}\exp\left[-\frac{1}{2\epsilon_1^2}\sum_{i=1}^{Z}z^2(i)\right]f(\epsilon_1^2)d\epsilon^2, \tag{34}$$

where the lower and upper bounds for the variance are denoted by $\check{\epsilon}^2$ and $\hat{\epsilon}^2$, respectively. The energy $\Sigma_{i=1}^{Z}z^2(i)$ is the only datum used for this receiver.

As before, we change variables to $u_z\equiv\epsilon_0^{-2}$ and $v_z\equiv\epsilon_1^{-2}$ and assume their density functions are uniform as $f(u_z)\equiv(\hat{u}_z-\check{u}_z)^{-1}$ and $f(v_z)\equiv(\hat{v}_z-\check{v}_z)^{-1}$. Then,

$$\overline{f(\mathbf{z}|H_0)}=\frac{1}{\mu_z^{(Z/2)+1}(\hat{u}_z-\check{u}_z)(2\pi)^{Z/2}}\left[\gamma\left(\frac{Z}{2}+1,\mu_z\hat{u}_z\right)-\gamma\left(\frac{Z}{2}+1,\mu_z\check{u}_z\right)\right], \tag{35}$$

$$\overline{f(\mathbf{z}|H_1)}=\frac{1}{\mu_z^{(Z/2)+1}(\hat{v}_z-\check{v}_z)(2\pi)^{(Z/2)}}\left[\gamma\left(\frac{Z}{2}+1,\mu_z\hat{v}_z\right)-\gamma\left(\frac{Z}{2}+1,\mu_z\check{v}_z\right)\right], \tag{36}$$

where $\mu_z\equiv\frac{1}{2}\Sigma_{i=1}^{Z}z^2(i)$ with bounds $\hat{u}_z\equiv\check{\epsilon}_0^{-2}$, $\check{u}_z\equiv\hat{\epsilon}_0^{-2}$, $\hat{v}_z\equiv\check{\epsilon}_1^{-2}$ and $\check{v}_z\equiv\hat{\epsilon}_1^{-2}$. The average likelihood ratio for the energy detector is

$$\overline{\Lambda_3}=\overline{f(\mathbf{z}|H_1)}\left(\overline{f(\mathbf{z}|H_0)}\right)^{-1}, \tag{37}$$

where the numerator and denominator come from Eqs. (36) and (35), respectively.

IX. EXAMPLES

A. Nonstationary or poorly bounded noise

Using the Neyman–Pearson criterion,[5] receiver operating curves[5,14,15] are plotted from 10 000 simulations of the the average likelihood ratios in Eqs. (31) and (37) for the matched-lag filter and the energy detector for

$N=3$ (# paths)
$Q=7$ (maximum difference in travel time from first to last path)
$10\log_{10}(A^2/\sigma_1^2)=0.9$ (dB for resolved signal in the autocorrelation function)
$K=1000$ (# of samples digitized at the receiver) (38)
$l^2=0.1$ signal-to-noise intensity ratio of the digitized samples
$\pm 50\%$ variation in the sample standard deviation
$D=0$ (# of time samples for field to go between the receivers)

(Fig. 1). The 50% variation in the standard deviation of the noise at the receiver(s) is due to nonstationary data, a lack of prior measurements of the noise-only field, or a lack of confidence in knowing if the signal is present or not from prior measurements from which the noise is estimated. This varia-

tion is thus bounded by $\check{\rho}=\rho/2$ and $\hat{\rho}=1.5\rho$ where ρ is the true standard deviation. This determines the limits in (17) as $\check{x}=\hat{\sigma}_0^{-2}$ and $\hat{x}=\check{\sigma}_0^{-2}$ where from (12), $\check{\sigma}_0^2=(\check{\rho}^2)^2/K$ and $\hat{\sigma}_0^2=(\hat{\rho}^2)^2/K$. The limits in (24) are $\check{y}=\hat{\sigma}_1^{-2}$ and $\hat{y}=\check{\sigma}_1^{-2}$ where (20) uses $\check{\sigma}_1^2$ and $\hat{\sigma}_1^2$ for $\check{\sigma}_0^2$ and $\hat{\sigma}_0^2$ respectively. In (29), $\check{w}=\hat{\beta}^{-2}$ and $\hat{w}=\check{\beta}^{-2}$ where β and $\hat{\beta}$ use $\{\check{\sigma}_1^2,\check{\alpha}_1^2\}$ and $\{\hat{\sigma}_1^2,\hat{\alpha}_1^2\}$ respectively and $\check{\alpha}_1^2$ and $\hat{\alpha}_1^2$ in (26) use $\check{\sigma}_0^2$ and $\hat{\sigma}_0^2$ respectively. For the energy detector in (35) and (36), $\check{\epsilon}_0^2=\check{\rho}^2$, $\hat{\epsilon}_0^2=\hat{\rho}^2$, $\check{\epsilon}_1^2=\check{\rho}^2+AN$, $\hat{\epsilon}_1^2=\hat{\rho}^2+AN$. The signal-to-noise intensity ratio is obtained from Eq. (21). Note that $l^2\ll 1$ as required for neglecting the off-diagonal terms in the correlation matrix for the noise and reverberation (Appendix A). Also note that $Q\ll K$, as required for assuming stationary values for the noise in the auto-correlation function.

For the values in Eq. (38), the probability of detection and false alarm are equal for the energy detector (Fig. 1). The addition of the small amount of signal energy does not increase the sample standard deviation significantly, and the energy detector yields no information regarding the presence or absence of the signal. The matched-lag filter yields probabilities of detection that are 4.1 and 366 times that of the energy detector at a 0.001 probability of false alarm for one and two receivers respectively.

B. Stationary noise

When the uncertainty in the noise is due to sampling K independent Gaussian random variables with mean zero and standard deviation ρ, the 95% confidence limits for the sample standard deviation are

$$\rho\left(1\pm\frac{2^{3/2}}{K^{1/2}}\right), \tag{39}$$

(Ref. 16). For one receiver, there are $K=1000$ samples, yielding an $\pm 8.9\%$ variation in the sample standard deviation. At a false-alarm probability of 0.001, the probabilities of detection for the matched-lag filter and energy detector are 0.099 and 0.95 respectively,

For two receivers, there are $2K=2000$ samples for estimating the sample standard deviation, with 95% confidence limits of $\pm 6.3\%$ variation in the sample standard deviation. At a false-alarm probability of 0.001, the probabilities of detection for the matched-lag filter and energy detector are 0.82 and 0.9997 respectively.

X. INTERPRETATION

The ideal matched-lag filter greatly outperforms the energy detector when the *a priori* variance of the noise is poorly estimated because the matched-lag filter incorporates robust features of any reverberant signal, i.e., that the signal-related lags must occur in physically allowed arrangements (Fig. 1). The improvement increases with the number of receivers. Evidently, it becomes less likely that noise can masquerade as a physically possible signal arrangement as the number of receivers increases. Indeed, the ratio of allowed to forbidden arrangements gets very small as the number of receivers or paths increases (Appendix C, Table II). When the noise uncertainty comes from finite sample sizes of stationary variance, the energy detector outperforms the ideal matched-lag filter, but by decreasing margins as the number of receivers increases (Sec. IX). Adding lag zero from the auto-correlation functions to the matched-lag filter incorporates the information from the energy detector. The decision threshold for the matched-lag filter would then be based on the likelihood ratio of the joint density function of lag zero with the other lags in the correlation functions.

TABLE II. The entries are $\log_{10} Y$ where Y is the ratio of the number of allowed to forbidden signal arrangements in the correlation functions as a function of the number of receivers and paths [Eq. (C1)]. Values are computed for $D=0$ and $Q=7$, as in Sec. IX. The ratio gets small very quickly with the number of receivers. The number of allowed and forbidden arrangements is the same for one receiver and two paths.

	# Paths		
# Receivers	2	3	4
1		−0.48	−1.4
2	−4.8	−10	−15
3	−9.6	−19	−29
4	−14	−28	−43
5	−19	−38	−57

The matched-lag filter is a matched filter within a matched filter.[7] The first stage of gain comes from using correlation functions. The second stage comes from searching for signals only among the physically allowed signal arrangements, which are often sparse compared with the number of forbidden arrangements.

XI. CONCLUSIONS

It may be computationally impractical to implement the likelihood ratio for a matched-lag filter by searching over all signal-arrangements, though massively parallel or quantum computers[17] may someday help. For example, suppose 16 paths of 0.016 s width are resolved in a 10 s interval after propagating 3000 km in the Pacific.[18] One would search an auto-correlation function in bins of about 0.016 ms width, allowing for misfit. There would be $Q=10/0.016=625$ bins, and $B_i(16)\sim 6\times 10^{29}$ arrangements [Eq. (5)]. It is practical to implement a form of the matched-lag filter by picking few enough peaks from the correlation functions so that one can test whether their lags and amplitudes are a good fit to an allowed arrangement. The probability that noise can mimic a good fit drops quickly with the number of receivers because the number of allowed to forbidden arrangements may be extremely small (Table II).

In some applications, signal-peaks may have large sidelobes in correlation functions. Lack of information about sidelobe shape and size will make it complicated to implement a matched-lag filter. In other applications, sidelobes are small, and the matched-lag filter would be similar to that described here. For example, there are countless situations in which animal calls and wireless communication systems yield small sidelobes.

The matched-lag filter should work when there are regions within correlation functions where gaps exist between signal-related lags.[11,19] There can be unresolved signals sharing similar lags, but there must be gaps between the signal regions. Then one can search for signal arrangements that are similar to observed, with some appropriate allowance for

misfits. This approach works when location, not detection, is an issue.[9,11] It would seem appropriate to use a stochastic detection method on regions of correlation functions containing no gaps between signal energy. Some hybrid approach then ought to be optimum.

The matched-lag filter was evaluated by assuming that all the travel time arrangements were equally likely. This evaluation is synonymous with the fact that the matched-lag filter can detect signals without any model for the propagation of acoustic or electromagnetic waves. In cases where it is possible to place bounds on the relative travel times of the paths, the *a priori* probability density function of the lag-redundancy arrangement, $f(b_i)$ in Eqs. (22), (27), can be modified to fit expectations to improve the probabilities of detection.[20] It appears that the matched-lag filter provides a theory that links the detection of broadband signals with location techniques such as matched-field processing,[21] when reliable models are available for the propagation of acoustic or electromagnetic waves.

A proof is provided via a counter example to show that some stochastic detection and communication theories (e.g., Refs. 1–6) provide incomplete descriptions of the structure of signals in auto-correlation functions because they permit forbidden signal arrangements. Consider three paths, each of amplitude eight, that always arrive at times 1, 2, 3, or 4 s. The four possible sets of arrival times are {1,2,3}, {1,2,4}, {1,3,4}, and {2,3,4} s, respectively. The corresponding auto-correlation functions for each at lags 0, 1, 2, and 3 s are {6,4,2,0}, {6,2,2,2}, {6,2,2,2}, and {6,4,2,0}. Assume each travel time set is equally likely. Then the expected value of these four auto-correlation functions is {6,3,2,1}. For the detection or communication problem, we now insist the auto-correlation function is $E[R_{11}(p)]=\{6,3,2,1\}$. Next, we note that the expected value of two auto-correlation functions, {6,4,2,0} and {3,0,1,2}, is $E[R_{11}(p)]$ where {6,4,2,0} is an allowed arrangement but {3,0,1,2} is forbidden because it is impossible to have signals simultaneously at lags 0, 2, and 3. Therefore a specified auto-correlation function does not exclude forbidden arrangements. Conversely, specifications of the allowed lag-amplitude arrangements and their probabilities of occurrences uniquely determine the auto- and cross-correlation functions of the signal. Thus the matched-lag filter contains more information about signal structure than the stochastic approach, and the matched-lag filter will yield higher probabilities of detection.

A comparison between a matched-lag filter and a stochastic detection theory was not made because the author is unaware of a published receiver operating curve for a stochastic approach that utilizes statistical assumptions about the signal and noise that are similar to those used here. Some stochastic theories do not allow the variance of the noise to be uncertain.[1–5] Almost all situations of practical interest are associated with uncertain variance. Some adaptive stochastic approaches do not allow for a single snapshot of the data[22] as used here. This paper's comparison between the matched-lag filter and the commonly used energy detector was made because identical assumptions could be used regarding the statistics of the noise and the signal.

A matched-lag filter ought to allow communication in multipath conditions with lower energy or signal-to-noise ratios than previously possible. For a discreet memoryless binary channel,[23] and the case in the right panel of Fig. 1, the capacity of the matched-lag filter is 0.68 bits at a false-alarm probability of 0.1. In contrast, the capacity of the energy detector is 0 bits at all false-alarm probabilities because the probability of detection and false-alarm are equal.

ACKNOWLEDGMENTS

Dr. John Zittle suggested to the author in 1996 that the lag equations in Ref. 8 might be applied to a detection problem. I thank Mary E. Putt, Ted G. Birdsall, Roy Greenfield, and John C. Burgess for their stimulating conversations and suggestions. I thank the reviewers for their suggestions. This work was supported in part by the Office of Naval Research Grant No. N00014-97-1-0613.

APPENDIX A: NOISE AND REVERBERATION IN CORRELATION FUNCTIONS

1. Noise only

Under hypothesis H_0, the joint probability density of the correlation function is obtained by substituting Eq. (9) in Eq. (2),

$$R_{ij}(p)=\frac{1}{K}\sum_{k=1+p}^{K} e_i(k)e_j(k-p); \quad i=1,2; \quad p\geqslant 0. \tag{A1}$$

Except for lag zero of the auto-correlation function, the mean of $R_{ij}(p)$ is zero because $e_i(k)$ has zero mean.

It is straightforward to show that the covariance,

$$\overline{R_{ij}(p)R_{ij}(q)} = \frac{1}{K^2}\sum_{k=1+p}^{K}\sum_{l=1+q}^{K}\overline{e_i(k)e_j(k-p)e_i(l)e_j(l-q)}; \quad p,q\geqslant 0, \tag{A2}$$

vanishes when $p\neq q$ (not shown). The covariance also vanishes when p and q can be either sign, as long as $p\neq q$, and $p\neq -q$ when $i=j$ for auto-correlation functions.

When $|p|\ll K$ and $K\gg 1$, $R_{ij}(p)$ is a sum of $K-|p|\gg 1$ uncorrelated [Eq. (A1)] random variables. When the variance of $e_i(k)$ is known, $R_{ij}(p)$ approaches a Gaussian random variable because of the central limit theorem.[24] When $|p|\ll K$, the reader may verify that the variance of the correlates in auto- and cross-correlation functions is given by Eq. (12), and is approximately independent of lag.

We estimate how many products of uncorrelated Gaussian random variables there need to be in Eq. (A1) so that $R_{ij}(p)$ is Gaussian with variance given by Eq. (12). Using a standard method,[25] it can be shown that the density function of u, which is the product of two uncorrelated Gaussian random variables with mean zero and variance ρ^2 is $f(u)=(\pi\rho^2)^{-1}\mathrm{BesselK}(0,u\rho^{-2})$, $u\geqslant 0$ and $f(u)=f(-u)$. The modified Bessel function of the second kind of order zero is $\mathrm{BesselK}(0,x)$ which approaches infinity as $x\to 0$. Different products in Eq. (A1) are uncorrelated but are not statistically

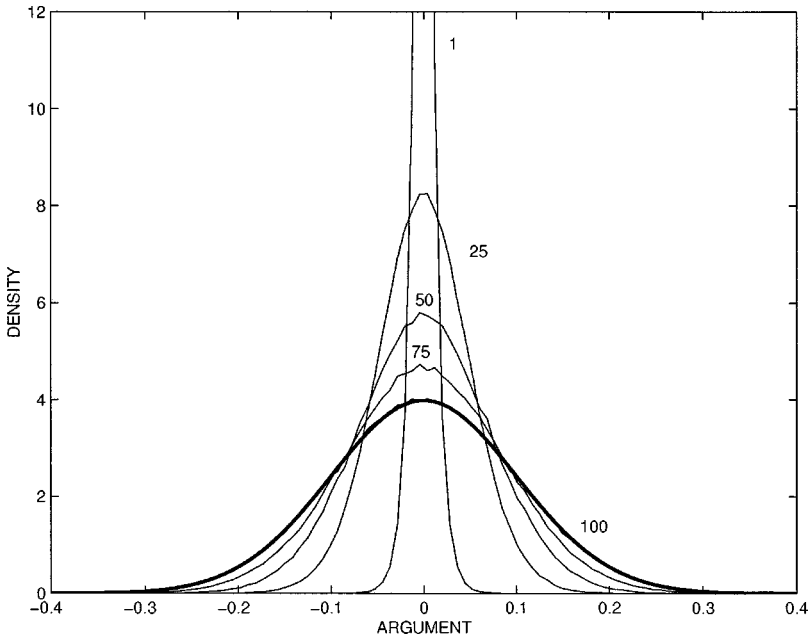


FIG. A1. Five empirical density functions (thin curves) compared with a Gaussian random variable (thick curve) having mean zero and variance given by Eq. (12) for $\rho^2=1$. The numbers equal $K-(1+p)+1$ in Eq. (A1), which is the number of summed products of uncorrelated Gaussian random variables which go into the correlation value at lag p. Note convergence to the Gaussian form of the correct variance when the number of summed products is 100.

independent. When there are 100 terms in the summation in Eq. (A1), $R_{ij}(p)$ approaches normality with the theoretical variance given by Eq. (12) (Fig. A1).

The covariance, $\overline{R_{ii}(p)R_{ij}(q)}=(1/K^2)\times\Sigma_{k=1+p}^{K}\Sigma_{l=1+q}^{K}e_i(k)e_i(k-p)e_i(l)e_j(l-q)$; $p,q\geqslant 0$, vanishes because the noise is uncorrelated between receivers i and j (Sec. III). The conditional density functions of the data at one or two receivers are then mutually uncorrelated Gaussian random variables with mean zero and variance σ_0^2 as in Eqs. (14) and (15).

2. Signal plus noise

With the presence of signal and noise, substituting Eq. (9) into (2), and using Eqs. (10), (11), yields

$$R_{ij}(p)=K^{-1}\delta(t_i(m)-t_j(n))\times\sum_{m=1}^{N}\sum_{n=1}^{N}a_i(m)a_j(n)\mathcal{E}+\Delta_{ij}(p), \tag{A3}$$

for nonnegative lags. The component due to noise and reverberation is

$$\Delta_{ij}(p)\equiv K^{-1}\sum_{k=1+p}^{K}\sum_{m=1}^{N_i}e_j(k-p)a_i(m)s(k-t_i(m))+K^{-1}\sum_{k=1+p}^{K}\sum_{n=1}^{N_j}e_i(k)a_j(n)s(k-p-t_j(n))+K^{-1}\sum_{k=1+p}^{K}e_i(k)e_j(k-p). \tag{A4}$$

For $p<0$, k goes from 1 to $K+p$. The expected value of $\Delta_{ij}(p)$ over the noise is zero. Also,

$$\overline{\Delta_{ij}^2(p)}=\frac{\rho_j^2}{K^2}\mathcal{E}\sum_{m=1}^{N_i}a_i^2(m)+\frac{\rho_i^2}{K^2}\mathcal{E}\sum_{n=1}^{N_j}a_j^2(n)+\rho_i^2\rho_j^2(K-|p|)+\frac{2\delta(i-j)\rho_j^2}{K^2}\mathcal{E}\sum_{m=1}^{N_j}\sum_{n=1}^{m-1}a_j(m)a_j(n)\times\delta(2p-t_j(m)+t_j(n)), \tag{A5}$$

where the variance of the noise at receiver i is ρ_i^2, which, so far, is allowed to vary from receiver to receiver.

For simplicity, assume $a_i(n)=a$, $N_i=N$, $\rho_i^2=\rho^2$, and $|p|\ll K\ \forall i$. Using Eq. (A3), one shows that Eq. (19) gives the amplitude, A, of a resolved signal peak in the correlation function. Eq. (A5) becomes

$$\overline{\Delta_{ij}^2(p)}=\frac{\rho^4}{K}+\frac{2\rho^2A}{K}[N+\delta(i-j)\delta(2p-t_j(m)+t_j(n))];\quad 1\leqslant n<m\leqslant N. \tag{A6}$$

For auto-correlation functions, the variance is slightly nonstationary because it depends on the lag. The variance becomes more stationary as N increases.

It is convenient to introduce the time averaged signal-to-noise intensity ratio at each receiver, $l^2\equiv\overline{K^{-1}\Sigma_{k=1}^{K}Na^2s^2(k)}\,\overline{(K^{-1}\Sigma_{k=1}^{K}e^2(k))^{-1}}$, which yields Eq. (21) from Eq. (19).

For cross-correlation functions, $i\neq j$ in Eq. (A6) and using Eq. (21) yields Eq. (26). For auto-correlation functions, $i=j$ in Eq. (A6), and there is an additional term which contributes for $2p=t_j(n)-t_j(m)$. Since $2p$ is even, and since there are about $N(N-1)/2$ positive signal lags in the auto-correlation function, about one-half of these are even on average. So the average variance of the noise in the auto-correlation functions is a weighted average of $(N(N-1)/2)/2=N(N-1)/4$ terms which include the extra variance, and $Q-N(N-1)/4$ terms which do not have the extra variance. (Q is the number of positive lags used from

the auto-correlation function.) The weighted average is, $\sigma_1^2 \equiv Q^{-1}((N(N-1)/4)(2\rho^2 A(N+1)/K+(\rho^2)^2/K)+(Q-N(N-1)/4)(2\rho^2 AN/K+(\rho^2)^2/K))$ which yields Eq. (20) using Eq. (21).

The covariance of the nonsignal components in the cross-correlation function is computed in analogy with Eq. (A4). For $a_i(m)=a$, $N_i=N$, and $\rho_i^2=\rho^2$, and $\overline{\Delta_{ij}(p)\Delta_{ij}(q)}=\sigma_0^2 l^2 N^{-1}(\Sigma_{m=1}^{N}\Sigma_{n=1}^{N}\delta[p-q-(t_i(m)-t_i(n))]+\Sigma_{m=1}^{N}\Sigma_{n=1}^{N}\delta[p-q-(t_j(m)-t_j(n))]+2\delta(i-j)\times\delta[p+q-(t_i(m)-t_i(n))])$. The ratio of the covariance to the variance is

$$\frac{\overline{\Delta_{ij}(p)\Delta_{ij}(q)}}{\sigma_1^2}\sim \mathrm{O}\left(\frac{l^2}{N}\right), \tag{A7}$$

which is small when $l^2 \ll 1$. Similarly, the covariance between the nonsignal components in the auto- and cross-correlation function is about $\overline{\Delta_{ii}(p)\Delta_{ij}(q)} \sim \sigma_0^2 l^2 N^{-1}(\Sigma_{n=1}^{N}\Sigma_{m=1}^{N}\delta(p+q-(t_i(n)-t_j(m)))+\delta(p-q-(t_j(m)-t_i(n))))$ when $a_i(m)=a$, $N_i=N$, and $\rho_i^2=\rho^2$. This covariance divided by the variance in the auto-correlation function [Eq. (20)], is,

$$\frac{\overline{\Delta_{ii}(p)\Delta_{ij}(q)}}{\sigma_1^2}\sim \mathrm{O}\left(\frac{2l^2}{N}\right), \tag{A8}$$

which is small when $l^2 \ll 1$.

APPENDIX B: NUMBER OF SYMMETRICAL TRAVEL TIME SETS

First consider symmetrical travel time sets with $t(N)-t(1)=\tau$ where $N-1 \leqslant \tau \leqslant Q$. Travel times are measured from zero to τ, where τ is a sample number. It is convenient to talk about a signal in a ''bin'' where bins 0 and τ are filled with signals, leaving $N-2$ signals to symmetrically put in the intervening bins. All symmetrical sets must have signals in bins,

$$n \text{ and } \tau-n;\quad n=0,1,2,\cdots,\tau. \tag{B1}$$

There are four cases to consider.

In case 1, τ and N are even. There cannot be a signal at bin $\tau/2$ otherwise N would be odd. Work with bins $\tau/2+1$ through $\tau-1$, which are referred to as the upper half. Half of the remaining $N-2$ remaining signals must go in the upper half with the remaining going in the lower half to maintain symmetry according to Eq. (B1). The upper half contains $(\tau/2)-1$ bins, and there are $\Upsilon_1(\tau,N)$ ways of arranging the $(N-2/2)$ samples among them (Eq. B2).

$$\Upsilon_1(\tau,N)=\begin{pmatrix}\frac{\tau}{2}-1\\ \frac{N-2}{2}\end{pmatrix};\quad \Upsilon_2(\tau,N)=\begin{pmatrix}\frac{\tau}{2}-1\\ \frac{N-3}{2}\end{pmatrix};$$
$$\Upsilon_3(\tau,N)=\begin{pmatrix}\frac{\tau-1}{2}\\ \frac{N-2}{2}\end{pmatrix}. \tag{B2}$$

In case 2, τ is even and N is odd. There must be a signal in bin $\tau/2$. The are $(\tau/2)-1$ upper half bins, $(\tau/2)+1$ through $\tau-1$. They are filled with $(N-3)/2$ signals which can be done in $\Upsilon_2(\tau,N)$ ways to make a symmetrical set. In case 3, τ is odd and N is even. There cannot be a signal in bin $\tau/2$. There are $(\tau-1)/2$ upper half bins, $(\tau-1)/2+1$ through $\tau-1$. These are filled with $(N-2)/2$ signals which can be done in $\Upsilon_3(\tau,N)$ ways to make a symmetrical set. In case 4, τ and N are odd. It is impossible to construct a symmetrical set of travel times for this case. The number of symmetrical sets is computed by summing over all values of τ, i.e., from $N-1$ to Q, $B_s=\Sigma_{n=1}^{Q-N/2+1}\Upsilon_1(N+2(n-1),N)+\Sigma_{n=1}^{Q-N/2+1}\Upsilon_3(N-1+2(n-1),N)$; N even, Q even. This simplifies to

$$B_s=\frac{2}{N}(Q-N+2)\begin{pmatrix}\frac{Q}{2}\\ \frac{N-2}{2}\end{pmatrix};\quad N \text{ even, } Q \text{ even.} \tag{B3}$$

Results for other odd and even combinations of N and Q are similar (not shown).

APPENDIX C: SPARSENESS OF ALLOWED ARRANGEMENTS OF SIGNALS

Let Q and L be the numbers of lags used from each auto- and cross-correlation function, respectively. Suppose there are N paths and B signal-arrangements at each receiver. Let these be independent of i [Eq. (5)] and assume these arrangements are independent between receivers. Recall that specifying $\tau_{1j}(1,1)$, $j=1,2,3,\cdots,\mathcal{R}$ specifies the values of $\tau_{ij}(1,1)$ in all the cross-correlation functions. For $\mathcal{R}$ receivers there are $\mathcal{A}(\mathcal{R})=B^{\mathcal{R}}(2D+1)^{\mathcal{R}-1}$ signal arrangements assuming the maximum travel time difference between each pair of receivers is $\pm D$.

Note that the number of ways of arranging a total of b balls in u urns is $Z(b,u)=((u-1)!)^{-1}\Pi_{i=1}^{u-1}(b+i)$ where each urn can hold zero to b balls. The number of forbidden and allowed signal arrangements in each auto-correlation function is $Z(P,Q)$ where $P \equiv N(N-1)/2$. The number of forbidden and allowed signal arrangements in each cross-correlation function is $Z(N^2,L)$ because there are N^2 signal lags to put among the L cross-correlation lags. Then for $\mathcal{R}$ receivers, the number of forbidden and allowed signal arrangements is $\mathcal{T}(\mathcal{R})=Z^{\mathcal{R}}(P,Q)Z^{\mathcal{R}-1}(N^2,L)$. The ratio of allowed to forbidden signal arrangements is then,

$$Y \equiv \frac{\mathcal{A}(\mathcal{R})}{\mathcal{T}(\mathcal{R})-\mathcal{A}(\mathcal{R})}. \tag{C1}$$

[1] R. C. Davis, ''The detectability of random signals in the presence of noise,'' Trans. I.R.E. **PGIT-3**, 52–62 (1954).

[2] D. C. Youla, ''The use of the method of maximum-likelihood in estimating continuous-modulated intelligence which has been corrupted by noise,'' Trans. I.R.E. **PGIT-3**, 90–105 (1954).

[3] D. Middleton, ''On the detection of stochastic signals in additive normal noise-Part I,'' IRE Trans. Inf. Theory **IT-3**, 86–121 (1957).

[4] R. Price and P. E. Green, Jr., ''Signal processing in radar astronomy—Communication via fluctuating multipath media,'' Lincoln Laboratory, M.I.T., Lexington, Ma., Tech. Rept. 234, Oct. (1960).

[5] C. W. Helstrom, *Statistical Theory of Signal Detection* (Pergammon, New York, 1968).
[6] D. B. Kilfoyle and A. B. Baggeroer, ''The state of the art in underwater acoustic telemetry,'' IEEE J. Ocean Eng. **25**, 4–27 (2000).
[7] D. O. North, ''An analysis of the factors which determine signal-noise discrimination in pulsed carrier systems,'' RCA Laboratory Report PTR-6C; Proc. IEEE **51**, 1016–1027 (1963) (1943).
[8] J. L. Spiesberger, ''Identifying cross-correlation peaks in the presence of echoes with application to optimal passive localization of calling animals and tomographic mapping of the environment,'' J. Acoust. Soc. Am. **100**, 910–917 (1996).
[9] J. L. Spiesberger, ''Linking auto and cross-correlation functions with correlation equations: Application to estimating the relative travel times and amplitudes of multipath,'' J. Acoust. Soc. Am. **104**, 300–312 (1998).
[10] J. L. Spiesberger, ''Detecting multipath signals with the matched-lag filter,'' Proc. 1999 IEEE International Conference on Acoustics, Speech, and Signal Processing, March 15–19, **III**, 1189–1192 (1999).
[11] J. L. Spiesberger, ''Locating animals from their sounds and tomography of the atmosphere: experimental demonstration,'' J. Acoust. Soc. Am. **106**, 837–846 (1999).
[12] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals Series and Products* (Academic, New York, 1965).
[13] Whalen, *Detection of Signals in Noise* (Academic, New York, 1971).
[14] T. G. Birdsall, ''The receiver operating curve,'' Ph.D. thesis, U. of Michigan, 1973.
[15] W. W. Peterson, T. G. Birdsall, and W. C. Fox, ''The theory of signal detectability,'' Trans. I.R.E. **PGIT-4**, 171–212 (1954).
[16] S. F. Arnold, *Mathematical Statistics* (Prentice-Hall, New Jersey, 1990).
[17] J. Preskill, ''Battling decoherence: the fault-tolerant quantum computer,'' Phys. Today **52**, 24–32 (1999).
[18] J. L. Spiesberger, E. Terray, and K. Prada, ''Successful ray modeling of acoustic multipaths over a 3000 km section in the Pacific,'' J. Acoust. Soc. Am. **95**, 3654–3657 (1994).
[19] J. A. Colosi, S. M. Flatte, 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).
[20] I. M. G. Lourtie and G. C. Carter, ''Signal detectors for random ocean media,'' J. Acoust. Soc. Am. **92**, 1420–1424 (1992).
[21] H. P. Bucker, ''Use of calculated sound fields and matched-field detection to locate sound sources in shallow water,'' J. Acoust. Soc. Am. **59**, 368–373 (1976).
[22] A. Steinhardt, ''Adaptive multisensor detection and estimation,'' in *Adaptive Radar Detection and Estimation*, edited by S. Haykin and A. Steinhardt (Wiley, New York, 1992), Chap. 3, pp. 91–158.
[23] R. G. Gallager, *Information Theory and Reliable Communication* (Wiley, New York, 1968).
[24] R. J. Serfling, *Approximation Theorems of Mathematical Statistics* (Wiley, New York, 1980).
[25] G. Casella and R. Berger, *Statistical Inference* (Wadsworth and Brooks/Cole, Pacific Grove, 1990), pp. 148–151.

A compressive gammachirp auditory filter for both physiological and psychophysical data

Toshion Irino[a)]
NTT Communication Science Laboratories, 2-4, Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0237, Japan

Roy D. Patterson[b)]
Centre for Neural Basis of Hearing, Department of Physiology, University of Cambridge, Downing Street, Cambridge CB2 3EG, United Kingdom

(Received 13 October 2000; revised 17 January 2001; accepted 5 March 2001)

A gammachirp auditory filter was developed by Irino and Patterson [J. Acoust. Soc. Am. **101**, 412–419 (1997)] to provide a level-dependent version of the linear, gammatone auditory filter, with which to explain the level-dependent changes in cochlear filtering observed in psychophysical masking experiments. In this 'analytical' gammachirp filter, the chirp varied with level and there was no explicit representation of the change in filter gain or compression with level. Subsequently, Carney *et al.* [J. Acoust. Soc. Am. **105**, 2384–2391 (1999)] reviewed Carney and Yin's [J. Neurophysiol. **60**, 1653–1677 (1988)] reverse-correlation (revcor) data and showed that the frequency glide of the chirp does not vary with level in their data. In this article, the architecture of the analytical gammachirp is reviewed with respect to cochlear physiology and a new form of gammachirp filter is described in which the magnitude response, the gain, and the compression vary with level but the chirp does not. This new 'compressive' gammachirp filter is used to fit the level-dependent revcor data reported by Carney *et al.* (1999) and the level-dependent masking data reported by Rosen and Baker [Hear. Res. **73**, 231–243 (1994)]. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1367253]

PACS numbers: 43.64.Bt, 43.66.Dc [LHC]

I. INTRODUCTION

The gammatone auditory filter was developed by Boer (1975) and de Boer and de Jongh (1978) to describe cochlear impulse responses measured in cats with the reverse-correlation, or revcor, technique. The envelope of the impulse response was approximated with the gamma distribution from statistics; the carrier, or fine structure, of the impulse response was a sinusoid, or tone, at the center frequency of the filter (Johannesma, 1972). Carney and Yin (1988) fitted the gammatone to the data from more than 150 primary auditory fibers over a wide range of stimulus levels, and showed that it provided a good approximation to these impulse response data. At about the same time, Schofield (1985) reported that the magnitude response of the gammatone filter could explain the masking data gathered by Patterson (1976) to derive the shape of the auditory filter psychophysically. Patterson *et al.* (1992) then reported that when the order of the gammatone was four, there was good agreement between the magnitude response of the gammatone filter and the rounded exponential, or roex, auditory filter used to fit a wide range of human masking data by a number of modelers (for a review, see Patterson and Moore, 1986). In essence, this meant that if one made the reasonable assumption that the roex auditory filter had a minimum phase characteristic, then it would be equivalent to a gammatone auditory filter. Slaney (1993), Cooke (1993) and Patterson *et al.* (1995) extended the mathematical framework of the gammatone filter family, and described methods for constructing a range of gammatone filterbanks for use in auditory modeling.

A. Level dependence

Psychophysical measurements of the auditory filter shape (e.g., Lutfi and Patterson, 1984; Rosen and Baker, 1994) indicate that the filter is approximately symmetric in frequency at low stimulus levels, but it becomes asymmetric as stimulus level increases with the low-frequency skirt becoming much shallower than the high-frequency skirt at high stimulus levels. These findings are consistent with physiological observations of basilar membrane motion (e.g., Pickles, 1988; Ruggero, 1992). The original versions of the gammatone filter were linear and so they could not account for level-dependent changes in filter shape. Carney (1993) introduced level-dependent gain and bandwidth into an extended version of the gammatone with feedback, but it had a symmetric frequency response. Lyon (1996, 1997) developed a one-zero version of the gammatone filter to introduce asymmetry in frequency; he reduced the zeros in an IIR version of the gammatone filter (Slaney, 1993) leaving just one at zero to raise the low-frequency tail of the filter. Irino and Patterson (1997) developed a level-dependent gamma*chirp* auditory filter to accommodate the level-dependent asymmetry observed in human masking data (Lutfi and Patterson, 1984; Moore and Glasberg, 1987; Moore *et al.*, 1990; Rosen and Baker, 1994). In this analytic gammachirp, the rate of chirp, or frequency modulation, at the start of the impulse response varies with level as the asymmetry varies with level.

About the same time, de Boer and Nuttal (1997) and

a)Electronic mail: irino@cslab.kecl.ntt.co.jp
b)Electronic mail: roy.patterson@mrc-cbu.cam.ac.uk

Recio *et al.* (1998) extended measurements of the glide, or chirp, in the instantaneous frequency (IF) of the impulse response of the basilar membrane, originally discovered by Møller and Nilsson (1979). The trajectories of the IF glides resemble the glide in the gammachirp filter which was initially introduced to explain the psychophysical masking data. Subsequently, Carney *et al.* (1999) reanalyzed the revcor functions of Carney and Yin (1988) and showed that the instantaneous frequency did, indeed, chirp. The results were somewhat surprising, however, because the trajectory of the instantaneous frequency was not level dependent (see also de Boer and Nuttal, 2000). This was inconsistent with the level-dependent chirp described by Irino and Patterson (1997) when fitting human masking data with the analytical gammachirp filter. The problem arose because Irino and Patterson chose the simplest form of gammachirp that would enable them to make the gammatone level dependent, without regard for the constraints implied by cochlear physiology. Moreover, the simplest form of gammachirp cannot account for the level-dependent gain and compression observed physiologically around the peak frequency (Pickles, 1988; Recio *et al.*, 1998). To overcome these problems, Irino and Patterson (2000) have developed a modified version of the gammachirp filter in which there are explicit changes in gain and compression with level, but no change in the chirp with level as required by the revcor data of Carney *et al.* (1999).

The purpose of this article is to extend the domain of this compressive gammachirp auditory filter, and demonstrate that it can explain the level dependent effects observed in both psychophysical and physiological data within a unified filtering framework. In Sec. II we describe the analytic and compressive gammachirp filters; in Sec. III we describe the fit to the physiological revcor data (Carney *et al.*, 1999); and in Sec. IV we describe the fit to the human masking data (Rosen and Baker, 1994).

II. GAMMACHIRP AUDITORY FILTERS

This section presents (a) the mathematical background of the gammachirp filter, (b) the modification of the analytic gammachirp into the compressive gammachirp filter, and (c) how these filters are used to fit the data in Secs. III and IV.

A. Analytic gammachirp

The analytic, complex form of the gammachirp auditory filter (Irino and Patterson, 1997) is

$$g_{ca}(t) = a t^{n_1-1} \exp(-2\pi b_1 \mathrm{ERB}(f_{r1}) t) \times \exp(j 2\pi f_{r1} t + j c_1 \ln t + j\phi_1), \quad (1)$$

where time $t>0$; a is the amplitude; n_1 and b_1 are parameters defining the envelope of the gamma distribution; c_1 is the chirp factor; f_{r1} is the asymptotic frequency; $\mathrm{ERB}(f_{r1})$ is the equivalent rectangular bandwidth (Glasberg and Moore, 1990); ϕ_1 is the initial phase; and $\ln t$ is the natural logarithm of time. When $c_1=0$, Eq. (1) reduces to the complex impulse response of the gamma*tone* filter (Patterson *et al.*, 1995). The gammachirp filter in Eq. (1) was derived analytically, as the function satisfying minimum uncertainty constraints in the space of the Mellin transform (Irino and Patterson, 1997), and so it is referred to as an analytic gammachirp.

The instantaneous frequency $f_i(t)$ is the derivative of the phase function in Eq. (1); that is

$$f_i(t) = f_{r1} + c_1/t. \quad (2)$$

It is a function of time and it asymptotes to the frequency, f_{r1}, as t becomes large. The trajectory of the instantaneous frequency towards f_{r1} is an upchirp when $c_1<0$ and a downchirp when $c_1>0$. When $c_1=0$, Eq. (1) is the complex form of the gammatone filter.

The Fourier magnitude spectrum of Eq. (1) is

$$|G_{CA}(f)| = a_\Gamma \cdot |G_T(f)| \cdot \exp(c_1 \theta_1), \quad (3)$$

$$\theta_1 = \arctan\left(\frac{f-f_{r1}}{b_1 \mathrm{ERB}(f_{r1})}\right). \quad (4)$$

The first factor of the right hand-side of Eq. (3), a_Γ, is the amplitude and the second factor, $|G_T(f)|$, is the Fourier magnitude spectrum of the gammatone filter. The detailed forms of these factors are described in Appendix A. The third factor in Eq. (3), $\exp(c_1\theta_1)$, is an asymmetric function since θ_1 is an antisymmetric function centered at the asymptotic frequency, f_{r1} as shown in Eq. (4). This factor applies asymmetry to the gammatone and converts the composite frequency response into that of a gammachirp auditory filter. The asymmetric function is a low-pass filter when $c_1<0$, a high-pass filter when $c_1>0$, and unity when $c_1=0$. The peak, or best, frequency can be obtained from the derivative of Eq. (3) and it is

$$f_{p1} = f_{r1} + c_1 b_1 \mathrm{ERB}(f_{r1})/n_1. \quad (5)$$

When fitting the gammachirp filter to human masking data, Irino and Patterson (1997) varied the asymmetry parameter linearly with the probe level of the signal, P_s, measured in dB. In a typical case $c_1 = 3.38 - 0.107 P_s$; c_1 is negative when P_s is greater than about 30 (dB SPL).

B. Filter shape of the analytic gammachirp

In this subsection, we investigate the frequency response of the analytic gammachirp filter. In most cases, the amplitude factor, a_Γ, has been normalized and considered separately from the discussion of the filter shape. We discuss the background and the effect of the amplitude factor in Appendix A and ignore it by setting $a_\Gamma=1$ in Eq. (3).

Figure 1 illustrates how Irino and Patterson (1997) produced a set of gammachirp filters, $|G_C(f)|$, with varying asymmetry, by multiplying a fixed gammatone filter, $|G_T(f)|$, together with a set of asymmetric functions, $\exp(c_1\theta_1)$. The fixed, gammatone auditory filter (GT) is shown by the solid line in the lower part of the figure. The low-pass asymmetric functions (LP-AF) are shown by the fan of dashed lines that pass through the same origin as the gammatone filter; c_1 was varied from 0 to –2 as level increased. The ordinate for the GT and LP-AF filters is on the right-hand side of Fig. 1. Each of the composite gammachirp filters (GC) is produced by multiplying the fixed GT by the appropriate LP-AF (or by adding the decibel values in the figure). The ordinate for the GC filters is on the left-hand side of Fig. 1. The vertical dotted line shows the fixed center

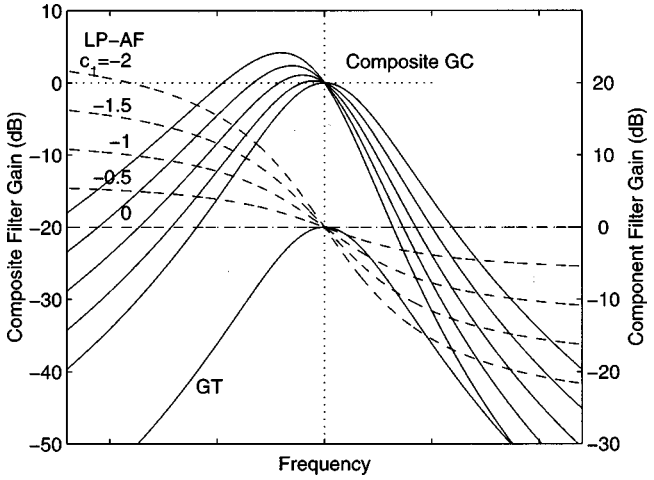


FIG. 1. A set of analytic gammachirp filters, GC (solid lines; left ordinate), produced by cascading a fixed gammatone filter, GT, with a low-pass asymmetric function, LP-AF, whose range varies with level (dashed lines; right ordinate).

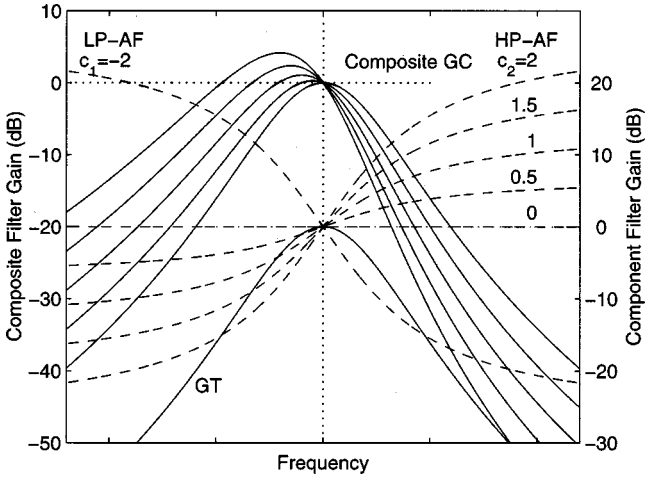


FIG. 2. A set of analytic gammachirp filters, GC (solid lines; left ordinate), produced by cascading a fixed gammatone filter, GT, first with a low-pass asymmetric function, LP-AF (dashed line; right ordinate), and then with a high-pass asymmetric function, HP-AF, whose range increases as level decreases (dashed lines; right ordinate).

frequency of the gammatone auditory filter in the lower part of the figure and the gammachirp with $c_1=0$ in the upper part of the figure; these two filters are, of course, the same. The gain at the peak of this analytic gammachirp filter *increases* with level, as c_1 becomes more negative, and at the same time, the peak frequency of the analytic gammachirp decreases.

There are several problems with the analytic gammachirp filter as a representation of cochlear filtering. The gain at the peak of the filter *increases* with level and the tails of the filters diverge below the center frequency rather than converging. Observations of basilar membrane motion show that as level increases peak frequency decreases as in the analytic gammachirp, but gain *decreases* (Pickles, 1988). Moreover, since the chirp parameter, c_1, is level dependent as described in Irino and Patterson (1997), the trajectory of the chirp changes with level as suggested by Eq. (2). Carney *et al.* (1999) have shown that, although there is a chirp in the impulse response of the cat's cochlear filter, *the form of the chirp does not vary with level* as it does in the analytic gammachirp.

In the next subsections, we describe how the analytic gammachirp filter can be modified to produce a gammachirp filter which is more representative of cochlear observations, and which can explain both the human masking data of Rosen and Baker (1994) and the revcor data of Carney *et al.* (1999).

C. Adapting the architecture of the gammachirp filter to accommodate cochlear observations

Since the chirp does not vary with level it is probably a property of the passive basilar membrane observed at high stimulus levels and post mortem (Recio *et al.*, 1998). The highest-level gammachirp in Fig. 1 has a c_1 of -2, and a brief study revealed that the frequency glide of this gammachirp filter was similar to that observed by Carney *et al.* (1999). This led Irino and Unoki (1999) to reverse the logic of the gammachirp when applying it to the physiology of the cochlea. Instead of simulating the auditory filter with a symmetric gamma*tone* filter at low levels that becomes an asymmetric gammachirp as level increases (for unspecified reasons), they suggested simulating the high-level, passive basilar membrane filter with a fixed, asymmetric, gammachirp filter, which then becomes progressively more symmetric as level decreases by virtue of the active mechanism in the cochlea. They noted that the product of any pair of asymmetric functions is another asymmetric function of the same general form. When an arbitrary value, c_a, is used to define an analytic gammachirp filter, it can be rewritten as follows:

$$\begin{aligned}|G_{CAa}(f)| &= |G_T(f)|\cdot\exp(c_a\,\theta_1)\\ &= |G_T(f)|\cdot\exp(c_1\,\theta_1)\cdot\exp\{(c_a-c_1)\,\theta_1\}\\ &= |G_T(f)|\cdot\exp(c_1\,\theta_1)\cdot\exp(c_2\,\theta_1)\\ &= |G_{CA}(f)|\cdot\exp(c_2\,\theta_1),\end{aligned}\tag{6}$$

where $c_2=c_a-c_1$. $|G_{CA}(f)|$ is an analytic gammachirp filter tied specifically to c_1 as in Eq. (3). Equation (6) shows that a gammachirp filter with an arbitrary shape can be derived from the combination of an initial gammachirp filter with parameter c_1 and an asymmetric function with a carefully chosen c_2. This led them to generate the level-dependent set of gammachirp filters shown in Fig. 2. Each of the composite responses shown by solid lines in the upper part of the figure is the product of (1) a fixed gammachirp filter produced by cascading the gammatone filter (GT) in the lower part of the figure with the fixed, low-pass asymmetric function (LP-AF) having $c_1=-2$, and (2) a second asymmetric function whose c_2 varies between 0 and $+2$, becoming more *positive* as level *decreases*. The result, of course, is the same set of composite gammachirp filters as in Fig. 1. This form of analytic gammachirp filter does not solve the problem of representing cochlear filtering with a family of gammachirp functions; the chirp in the impulse response still varies with level and gain still increases with level. But now the change with level is associated with a high-pass asym-

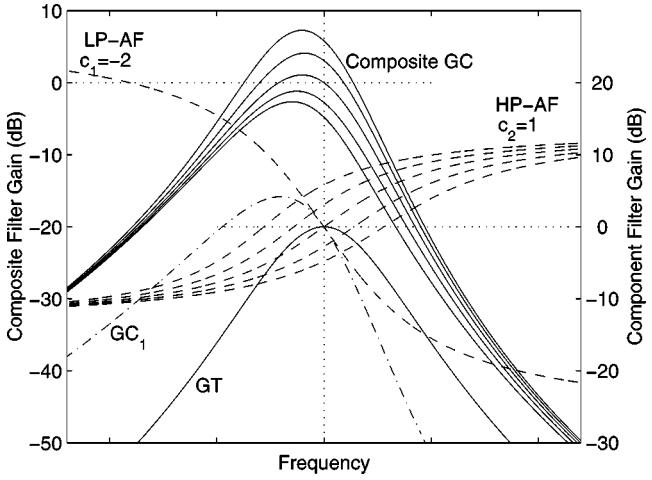


FIG. 3. A set of gammachirp filters, GC (solid lines; left ordinate), produced by cascading a fixed gammatone filter, GT, first with a low-pass asymmetric function, LP-AF (dashed line; right ordinate), and then with a high-pass asymmetric function, HP-AF, whose center frequency, f_{r2}, increases with level (dashed lines; right ordinate). A fixed gammachirp filter, GC_1 (dashed and dotted line; right ordinate) is produced by GT and LP-AF.

metric function (HP-AF) whose effect increases as level decreases, and this form of gammachirp can be developed into a form that is more representative of cochlear physiology.

D. A compressive gammachirp auditory filter

The next step in the process is to rewrite the angular variables of the asymmetric functions in terms of physiological variables and parameters that vary with level, which will then make it possible to fix the values of the asymmetry parameters, c_1 and c_2, so that the chirp does not vary with level. Starting with the form of Eq. (6) in which the representation of both asymmetric functions is explicit,

$$|G_{CC}(f)| = |G_T(f)| \cdot \exp(c_1\,\theta_1) \cdot \exp(c_2\,\theta_2) = |G_{CA}(f)| \cdot \exp(c_2\,\theta_2), \tag{7}$$

the angular variables are rewritten in terms of the center frequencies and bandwidths of the passive gammachirp filter and the level-dependent asymmetric functions. If the filter center frequencies are f_{r1} and f_{r2}, respectively, then from Eq. (4)

$$\theta_1 = \arctan\left(\frac{f - f_{r1}}{b_1 \mathrm{ERB}(f_{r1})}\right) \quad \text{and} \quad \theta_2 = \arctan\left(\frac{f - f_{r2}}{b_2 \mathrm{ERB}(f_{r2})}\right). \tag{8}$$

In this form, the chirp parameters, c_1 and c_2, can be fixed with, for example, $c_1 = -2$, and $c_2 = 1$, and the level dependency can be associated with the center frequency of the *high-pass* asymmetric function, f_{r2}. The center frequency of the low-pass asymmetric function, f_{r1}, is fixed, so that the best frequency of the gammachirp that represents the passive basilar membrane is fixed. If then, for simplicity, $b_2 = b_1$, a new gammachirp filter system is defined with component asymmetric functions and composite gammachirp filters as shown in Fig. 3 by the dashed and solid lines, respectively.[1]

The fixed analytic gammachirp, $|G_{CA}(f)|$, is intended to represent the passive basilar membrane response which is not level dependent; it is presented by the dashed and dotted line in the lower part of the figure (GC_1). This filter largely determines the form of the chirp, and, so, the chirp no longer varies significantly with level. The asymmetric function, $\exp(c_2\,\theta_2)$, of the compressive gammachirp is similar to that of the analytic gammachirp inasmuch as it has a high-pass characteristic that changes with level and renders the composite gammachirp *more symmetric at lower levels*. However, in order to simulate the linearity, or lack of compression, observed in the cochlea at frequencies remote from the best frequency, the asymmetric functions are made to be flat spectrum at frequencies outside the passband. The change with level in the composite gammachirp is achieved primarily by varying the center frequency of the asymmetric function relative to the best frequency of the fixed gammachirp that represents the basilar membrane response. The high-pass asymmetric functions (HP-AF) are shown by the fan of dashed lines in Fig. 3; the function moves up in frequency as level increases. Each of the composite, compressive gammachirp filters is produced by multiplying the fixed GC_1 by the appropriate HP-AF (or by adding the decibel values in the figure). The frequency ratio between the peak frequency f_{p1} in Eq. (5) and the center frequency f_{r2} of the high-pass asymmetric function is

$$f_{\mathrm{rat}} = f_{r2}/f_{p1}. \tag{9}$$

Now, the gain of the composite gammachirp increases and its bandwidth decreases as the center frequency of the high-pass asymmetric function decreases, which agrees, at least qualitatively, with the physiological data. Moreover, the best frequency of the composite filter still shifts down in frequency as level increases, albeit by a relatively small amount in this example. Perhaps most importantly, the chirp no longer varies with level as required by the revcor data of Carney *et al.* (1999). Since this form of the gammachirp filter is compressive around the peak frequency, it is referred to as a compressive gammachirp auditory filter.

The high-pass asymmetric function, $\exp(c_2\,\theta_2)$, does not have an analytic impulse response. An asymmetric compensation filter is developed in Appendix B to enable the simulation of the compressive gammachirp impulse response, so that

$$g_{cc}(t) = a_c \cdot g_{ca}(t) * h_c(t), \tag{10}$$

where a_c is a constant, $g_{ca}(t)$ is the gammachirp impulse response from Eq. (1), and $h_c(t)$ is the impulse response of the asymmetric compensation filter [Appendix B, Eqs. (B4)–(B8)].

III. FITTING THE COMPRESSIVE GAMMACHIRP FILTER TO REVCOR DATA

We characterize the level dependency of the revcor data (Carney *et al.*, 1999) by the level dependency of the parameters of the real function used to represent the auditory filters; that is, the gammatone, the analytic gammachirp, and the compressive gammachirp. We then compare these three filters in terms of the rms fit to physiological data, the number of coefficients in the filter, and the characteristics of the chirp.

A. Parameters

For the analytic gammachirp, the parameters in the fit are, b_1, c_1, f_{r1}, and ϕ_1 in Eq. (1) and the delay of the zero point, τ, which represents the delay of the traveling wave in the cochlea. In the case of the gammatone, the chirp parameter c_1 is set to zero. The compressive gammachirp has eight parameters: b_1, c_1, f_{r1}, ϕ_1, τ, b_2, c_2, and f_{rat}. The amplitude factors a and a_c were not used since both the revcor data and the filter response are normalized. In addition, the order of the gamma distribution, n_1, was fixed to 4 as results of a preliminary simulation and previous studies of the gammatone (de Boer and de Jongh, 1978; Carney and Yin, 1988) and gammachirp (Irino and Patterson, 1997) indicated that this was the appropriate value. The level dependency was introduced as a linear function of noise level, N_r (dB), in the experiment. For example, if the level dependency is applied to the parameter b_1, the formula is

$$b_1 = b_1^{(0)} + b_1^{(1)}(N_r - 80) = b_1^{(0)} + b_1^{(1)} N_{rd}. \tag{11}$$

For convenience in the fitting procedure, we introduced a noise-level parameter, N_{rd}, and we set the coefficient $b_1^{(0)}$ to the value when $N_{rd} = 80$ (dB), which is the common signal level measured for all units. The selection of the level-dependent parameters is described in the following sections.

B. Fitting procedure

The root-mean-squared difference between the revcor data and the filter response was minimized using the Levenberg–Marquardt method (Press *et al.*, 1988). Both the data and the filter were normalized in advance since the gain of the revcor functions varied little with signal level (Carney *et al.*, 1999; Patuzzi and Robertson, 1988). Carney and Yin (1988) measured revcor functions in auditory nerve fibers (or units) over a range of noise levels that depended on how long they held the unit. With one unit they managed ten noise levels (Carney *et al.*, 1999), but with most units it was far fewer. The number of coefficients for the gammachirp filter varies from as few as four up to about ten, and as a result, the coefficients for many units would be theoretically underdetermined. It is also difficult to specify good starting values for the parameters in the sense of producing rapid convergence on a stable solution. Consequently, several techniques were introduced to overcome these problems. First, the fits were performed in two stages; initially, the more stable envelope was fitted to get starting values for the envelope parameters, and then the full impulse response with the fine structure was fitted. Second, for each of these stages, the data at the highest level, 80 dB SPL, was fitted first to determine the zeroth-order coefficients (e.g., $b_1^{(0)}$). Then, with these as starting values, all of the data were fitted to determine all of the coefficients, simultaneously. Finally, we divided the time range of the rms calculation for a given response into two parts—before and after the envelope peak. The number of rms differences is thereby doubled and this is known to assist convergence in least-squared-error methods.

1. Envelope fit

The envelope of the revcor data was calculated over a 20-ms duration as the absolute value of the complex response derived with the Hilbert transform. At the extremes of the interval, where the basilar membrane response was virtually nonexistent, the root-mean-squared (rms) value of the noise was calculated from the revcor data, between 0 and 1 ms and between 18 and 20 ms. An envelope threshold was set at 1.5 times this rms noise level. Envelope values above this threshold were used for calculation of the rms difference. For the analytic gammachirp and the gammatone, the parameters b_1 and τ were included in the least-mean-square process because it is they that mainly determine the envelope shape and the temporal location. For the compressive gammachirp, all parameters except for ϕ_1 were included because they all affect the envelope. Initially, the revcor data for 80 dB SPL were fitted to determine the zeroth-order coefficients $b_1^{(0)}$ and $\tau^{(0)}$. They were then used as the initial values in the next step. The revcor data for all signal levels including 80 dB were used to determine the first-order coefficients $b_1^{(1)}$ and $\tau^{(1)}$ as well as the zeroth-order coefficients $b_1^{(0)}$ and $\tau^{(0)}$. This stage determines the shape of the envelope and its location on the time axis.

2. Waveform fit

Once the envelope was determined with the above procedure, the rest of parameters, f_{r1}, c_1, and ϕ_1 were used to minimize the error between the raw revcor data and the real part of the impulse response for the analytic gammachirp. For the compressive gammachirp, all parameters other than the delay τ are used at this stage. The rms difference between the data and the filter response was calculated within the range where the envelope of the gammachirp is greater than 2% of the peak value. This threshold was established in a preliminary simulation as a tradeoff between maximization of the response length used in the fit and minimization of the background noise in low-power parts of the revcor data. The rms error was defined as the difference between the revcor waveform and the real part of the filter response. The final value was normalized by the rms amplitude of the revcor waveform and specified in dB. As before, the fitting was performed first at 80 dB SPL and then for all signal levels simultaneously.

C. Results: Tradeoffs between accuracy and number of coefficients

This section describes the relationship between the number of coefficients assigned to each parameter and the degree of the fit. The revcor data for Unit 86100u25 with a best frequency of about 2000 Hz was selected from the examples in Fig. 1 of Carney *et al.* (1999) because this unit was measured at ten signal levels which was more than for any other unit. We excluded the data at the lowest level, 20-dB data, because of the relatively high noise level.

1. Fitting with the gammatone filter

First, the revcor data of unit86100u25 was fitted with the real impulse response of the gammatone filter [i.e., $c_1 = 0$ in

TABLE I. Relationship between the number of coefficients and root-mean-square (rms) error for gammatone filters fitted to revcor functions from a cat auditory nerve as reported in Fig. 1(E) (Unit 86100u25, BF=2000 Hz) of Carney *et al.* (1999). The first column shows the total number of coefficients. The second to fifth columns show the number of coefficients used for the gammatone parameter at the top of that column: 1 indicates a filter parameter that is constant across noise level and 2 indicates a linear dependence of the parameter on signal level. The last column shows rms error in dB. The row with the asterisk* shows the parameter set used for Fig. 4.

No. of coefficients	f_{r1}	b_1	ϕ_1	τ	Error
8	2	2	2	2	−10.1
7*	2	2	2	1	−10.1
7	1	2	2	2	−9.8
7	2	2	1	2	−10.0
7	2	1	2	2	−7.3
6	1	2	2	1	−9.9
6	1	2	1	2	−9.8
5	1	2	1	1	−9.8
4	1	1	1	1	−7.6

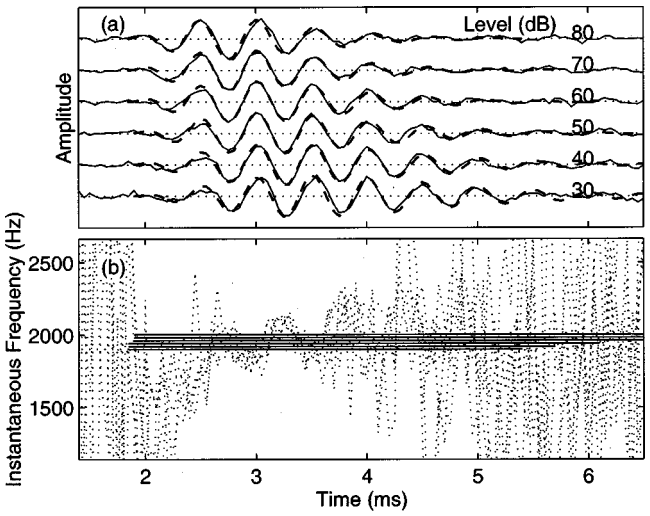


FIG. 4. (a) Normalized revcor functions from unit 86100u25 (solid lines) and the filter responses (dashed lines) of the best-fitting gammatone filter with the parameter set indicated by the asterisk (*) in Table I. The amplitude of the response was normalized by the input signal level indicated by the number on the right-hand side. (b) Instantaneous frequency estimated from the best-fitting gammatone filter (solid lines) and from the revcor functions (dotted lines) using a Hilbert transform after smoothing with a three-point rectangular window.

Eq. (1)]. The parameters for the gammatone are f_{r1}, b_1, ϕ_1, and τ. Table I shows the total number of coefficients (first column), the assignment of coefficients to the parameter (second to fifth columns), and the final rms error in dB between the revcor data and the gammatone filter response (sixth column). The table shows that the gammatone filter can fit the revcor data with an rms error of as little as −10.1 dB when all of the parameters are level dependent. This is a reasonable limit given the noise in the revcor data; the signal-to-noise ratio is less than about 20 dB because the rms noise level is about 10% of the maximum value in the revcor data. At the same time, it is clear that an rms error close to the limit can be achieved with far fewer parameters, so long as the envelope parameter b_1 is level dependent. When b_1 is not level dependent as shown in rows with a 1 in column two, the error increases to between −7.3 and −7.6 dB, which is more than 2 dB worse than the best fit. The reason is illustrated in Fig. 4(a) which presents the revcor impulse responses with solid lines; as signal level increases, the duration of the response decreases and the envelope becomes shorter. The zero-crossings of the carrier are, however, virtually invariant as Carney *et al.* (1999) pointed out. So, it is basically the envelope that is level dependent in these revcor data and b_1 is the main envelope parameter in the gammatone filter.

The impulse responses of the gammatone filters are also plotted in Fig. 4(a) with dashed lines; the case where f_{r1}, b_1, and τ are level dependent is indicated by the asterisk in Table I. This version was used to illustrate the best fit, or the upper limit available with the gammatone filter, rather than the most efficient in terms of the number of parameters per data point. These solutions are presented primarily for comparison with the gammachirp filters in the next subsections. There is good agreement between the data and these gammatone filter responses except in the initial, raising portion of the response between 2 and 2.5 ms.

The instantaneous frequency (IF) of the response was calculated using the Hilbert transform of the average revcor data, smoothed with a three-point rectangular window to reduce abrupt changes in the estimated value. Figure 4(b) shows the instantaneous frequencies estimated for the six levels shown in Fig. 4(a) (dotted lines). The accuracy of the instantaneous frequency data improves rapidly over the first 2 ms and the functions for different levels converge between about 2 and 3.5 ms provided the amplitude is not too low. These IF trajectories show that the chirp does not vary much with level as Carney *et al.* (1999) pointed out. The IF trajectories from the fitted gammatone filters are presented by the solid *straight* lines, and they make it clear that it is impossible to simulate the chirp in the data with the gammatone because the IF of the gammatone is constant, by definition. The main disparities in Fig. 4(b) are observed over the initial portion of the response, and they illustrate that there is a serious limitation in the ability of the gammatone filter to explain revcor data.

2. Fitting with the analytic gammachirp filter

The analytic gammachirp filter of Eq. (1) was used for the fit to the same revcor data of unit86100u25 with five parameters, f_{r1}, b_1, c_1, ϕ_1, and τ. The strategy was the same as described above for the gammatone filter. Table II shows the results. The rms error is reduced to about −11 dB provided the envelope parameter, b_1, is level dependent as indicated by a ''2'' in column three. The fit is improved over that of the gammatone filter by about 1 dB even when the filter has the same number of variable coefficients (e.g., 5 or 6). Making the other parameters level dependent had little effect on the rms error. For example, a comparison between the twelfth and sixteenth rows shows that the delay parameter, τ, does not need to be level dependent. Moreover, the phase parameter, ϕ_1, can be set to zero with little increment in the error term because the initial phase also occurs in the chirp parameter, c_1.[2] In contrast, when the envelope param-

TABLE II. Relationship between the number of coefficients and the rms error for the best-fitting analytic gammachirp filter to the same revcor functions as in Table I. 0 indicates ϕ_1 value of 0. See the caption of Table I for details.

No. of coefficients	f_{r1}	b_1	c_1	ϕ_1	τ	Error
10	2	2	2	2	2	−11.1
9	1	2	2	2	2	−11.0
9	2	2	2	2	1	−11.2
9	2	2	2	1	2	−11.1
9	2	2	1	2	2	−11.2
9	2	1	2	2	2	−7.7
8*	1	2	2	2	1	−11.0
8	1	2	2	1	2	−11.0
8	1	2	1	2	2	−11.0
8	1	1	2	2	2	−7.6
7	1	2	2	1	1	−11.0
7	1	2	1	1	2	−11.0
7	1	2	1	2	1	−11.0
7	1	1	2	2	1	−8.1
7	1	1	2	1	2	−7.7
6	1	2	1	1	1	−11.0
6	1	1	2	1	1	−8.3
5	1	2	1	0	1	−10.7
5	1	1	1	1	1	−8.1

eter, b_1, is not level dependent (rows with a 1 in column three), the rms error increases by around 3 dB to about −8 dB. The reason is the same as for the gammatone filter; the duration of the impulse response decreases with increasing level.

The real parts of the impulse responses of the analytic gammachirp filters are shown with dashed lines in Fig. 5(a); in this version there are eight coefficients as indicated in the row with the asterisk in Table II. This version was selected to fit this unit and other units described in Sec. III D. There is an improvement in the fit, primarily in the first few milliseconds, which reduces the rms error by about 1 dB. The effect is more obvious in Fig. 5(b) which shows the IF trajectories

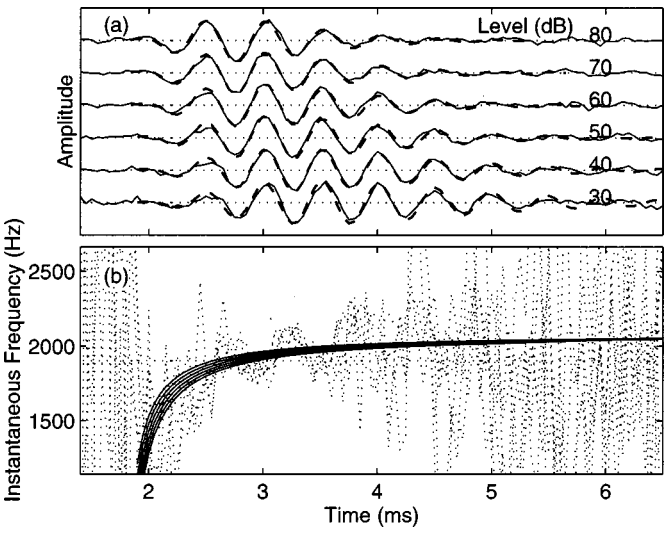


FIG. 5. (a) The same revcor functions as Fig. 4 (solid lines) with responses of the analytic gammachirp filter (dashed lines) using the parameter set indicated by the asterisk (*) in Table II. (b) Instantaneous frequency estimated from the best-fitting analytic gammachirp filter (solid lines) and from the same revcor functions (dotted lines).

TABLE III. Relationship between the number of coefficients and the rms error for the best-fitting compressive gammachirp filter to the same revcor functions as in Table I. The last column shows the average compression between the input and output signal levels in dB/dB. See the captions of Tables I and II for details.

No.	f_{r1}	b_1	c_1	ϕ_1	τ	f_{rat}	b_2	c_2	Error	cmp
10	2	1	1	1	1	2	1	1	−12.0	0.65
10	2	1	1	0	1	2	2	1	−10.0	0.92
9*	2	1	1	0	1	2	1	1	−12.0	0.65
9	1	1	1	1	1	2	1	1	−11.8	0.67
9	1	1	1	0	1	2	2	1	−11.5	0.74
9	2	1	1	1	1	1	1	1	−7.7	1.00
8	1	1	1	0	1	2	1	1	−12.0	0.63
8	2	1	1	0	1	1	1	1	−7.7	1.00

for the analytic gammachirp filters with solid lines [see also Eq. (2)]; they rise smoothly with the average of the revcor data over the duration of the response which is clear of background noise. Thus, the analytic gammachirp provides a good explanation of the revcor data. This analytic gammachirp fit is, however, somewhat different from the one applied by Irino and Patterson (1997) to human masking data; in that case, the chirp parameter, c_1, was level dependent and the envelope parameter, b_1, was not level dependent. This disparity in the level-dependent parameters is one shortcoming of the analytic gammachirp as a model for both physiological and psychophysical data.[3]

3. Fitting with the compressive gammachirp filter

The compressive gammachirp filter of Eq. (10) was used to fit the same revcor data of unit86100u25 with eight parameters, f_{r1}, b_1, c_1, ϕ_1, τ, f_{rat}, b_2, and c_2. In this case, we restricted the maximum number of parameters to ten because the fit would be mathematically underdetermined with nine sets of revcor data. Moreover, ten parameters is sufficient to survey the upper limit of the fitting process. Table III shows the results in the same format as previously: the total number of coefficients (first column), the assignment to parameter (second to ninth columns), and the rms error in dB (tenth column). There is also an extra column showing the average compression in dB/dB (11th column) which will be described shortly. The minimum rms error was −12.0 dB which is better than with the analytic gammachirp by about 1 dB and better than with the gammatone by about 2 dB. When the frequency ratio, f_{rat}, in Eq. (9) is not level dependent (the sixth and eighth rows), the rms error increases about 4 dB to −7.7 dB. So, the relative position in frequency of the initial gammachirp and the high-pass compensation filter needs to be level dependent. Making the other parameters level dependent had little effect on rms error. It also proved possible to set the phase parameter, ϕ_1, to zero without increasing the rms error. The value of the delay parameter, τ, is about 1.8 ms, which is the same as for the analytic gammachirp.

The real parts of the impulse responses of the compressive gammachirp filters are shown with dashed lines in Fig. 6(a); in this version there are nine coefficients as indicated by the asterisk in Table III. Disparities between the revcor data and the filter responses are now limited to the initial and final

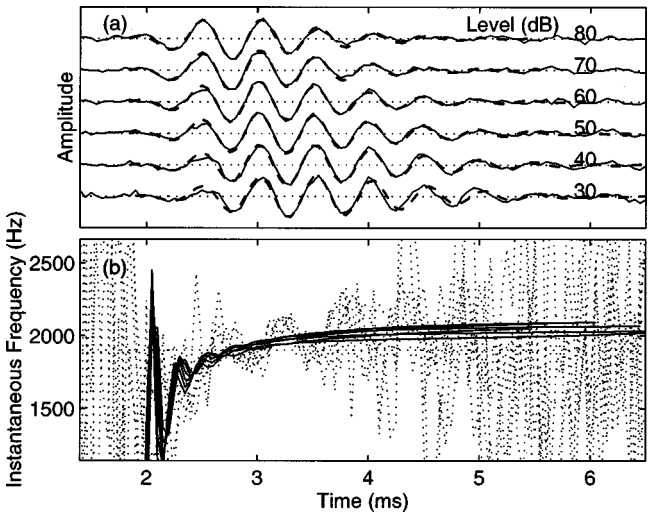


FIG. 6. (a) The same revcor functions as in Fig. 4 (solid lines) with responses of the best-fitting compressive gammachirp filter (dashed lines) using the parameter set indicated by the asterisk (*) in Table III. (b) Instantaneous frequency estimated from the compressive gammachirp filter (solid lines) and from the same revcor functions (dotted lines).

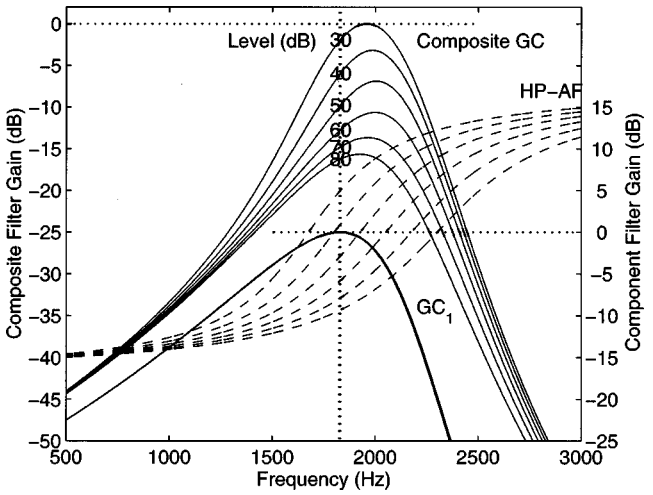


FIG. 7. The set of compressive gammachirp filters, GC (solid lines; left ordinate) derived from the normalized revcor data of Unit 86100u25, corresponding to the impulse responses in Fig. 6(a). The filters were produced by cascading a passive gammachirp filter, GC_1 (lower solid line; right ordinate) with a high-pass asymmetric function, HP-AF (dashed lines; right ordinate) whose center frequency varies with level.

parts of the responses at 30 and 40 dB; the data and filters are almost coincident when the relative amplitude of the waveform is sufficiently large. The level dependence of the frequency ratio, f_{rat}, introduces a level-dependent envelope as described shortly. The IF trajectories of the compressive gammachirp filters are plotted with solid lines in Fig. 6(b). They start with an oscillation and rise with the average of the revcor data. They contrast with the smooth IF trajectories for the analytic gammachirp filters in Fig. 5(b) and Eq. (2). The oscillation is produced by the interaction between the initial, analytic gammachirp filter and the high-pass asymmetric compensation filter. The rapid oscillation seems to contribute to the improvement in the fit, although it is difficult to analyze the effect in detail because the onsets of the revcor functions are at, or below, the noise level.

Figure 7 shows the frequency responses (Composite GC) of the compressive gammachirp filters shown in Fig. 6(a). Each is the composite response of the initial, analytic gammachirp filter (GC_1) and a high-pass asymmetric function (HP-AF). The initial, analytic gammachirp is essentially independent of noise level. The frequency ratio, f_{rat}, is the dominant, level-dependent parameter and it is positively correlated with the noise level. The gain of the auditory filter varies within the passband and it is roughly constant outside the passband. As signal level increases, the peak gain decreases and the bandwidth generally increases. As a result, the envelope in the waveform becomes shorter as the frequency ratio, f_{rat}, becomes larger, since the envelope duration is negatively correlated with bandwidth, provided the carrier frequency does not fluctuate too much.

The peak frequency increases with noise level at first and then it decreases a little. This differs from the typical physiological observation that peak frequency decreases monotonically as level increases. The discrepancy is probably due to a lack of sensitivity to the frequency parameter, f_{r1}, in the current fitting procedure. The fits were restricted to time domain. It is possible that estimation of the change in peak frequency could be improved by introducing constraints from the frequency responses of the revcor data.

The filter gain of the compressive gammachirp varies with noise level. The gain was normalized to 0 dB for a noise level of 30 dB SPL. The gain decreases as noise level increases and is about −16 dB when noise level is 80 dB SPL. Thus, the input–output function of this filter is compressive and the average growth ratio between 30 and 70 dB is about 0.65 dB/dB (see the last column in Table III, labeled ''cmp,'' for an alternative set of coefficients). The compressive gain only occurs in the region of the peak frequency of the filter; the degree of compression decreases and the tails of the filters converge as frequency diverges from the peak frequency. Such compressive gain has been repeatedly reported both physiologically (e.g., Pickles, 1988; Ruggero, 1992) and psychophysically (Oxenham and Plack, 1997). It is not possible to obtain this kind of compression either by varying the amplitude parameter, a, of the analytic gammachirp [Eq. (1)] or the amplitude parameter, a_c, of the compressive gammachirp [Eq. (10)]. It is necessary to invoke a mechanism which affects the amplitude only around the peak frequency. This is accomplished by the interaction of the two component filters in the compressive gammachirp filter.

In summary, the compressive gammachirp filter provides the best account of the revcor data of the three models in terms of rms error, IF trajectory, and compressive gain.

D. Results: Individual units

The frequency glides derived from several units in the same cat (86100) were reported in Fig. 4 of Carney *et al.* (1999). The instantaneous frequencies were determined directly using the zero crossings of the revcor waveforms without any parametric model. First-order regression lines were plotted on the same figure to show the general trends. We have characterized the IF trajectory with parametric models, namely, the analytic gammachirp and the compressive gam-

TABLE IV. Parameter values and rms errors for the best-fitting analytic gammachirp filters to the revcor functions from various units of cat 86100 (see Fig. 4 of Carney *et al.*, 1999). The first column shows the unit number. The second column shows the number of sound pressure levels measured for the unit. The last column shows rms error in dB. The remaining columns show the best-fitting coefficients for f_{r1}, b_1, c_1, ϕ_1 τ when $n_1=4$, $N_{rd}=N_r-80$.

Unit	Level no.	f_{r1}	b_1	c_1	ϕ_1	τ	Error
20	2	303	$2.39+0.010N_{rd}$	$1.18-0.001N_{rd}$	$0.96-0.000N_{rd}$	2.10	−10.0
2	2	463	$2.57+0.015N_{rd}$	$1.29+0.017N_{rd}$	$1.92+0.113N_{rd}$	2.10	−10.7
22	4	1150	$2.66+0.022N_{rd}$	$0.11+0.012N_{rd}$	$1.18+0.102N_{rd}$	2.46	−10.3
7	5	1660	$1.82+0.013N_{rd}$	$-0.60+0.002N_{rd}$	$-1.17+0.022N_{rd}$	1.78	−12.7
25	9	2085	$1.89+0.016N_{rd}$	$-1.39-0.009N_{rd}$	$-0.41-0.054N_{rd}$	1.73	−11.0
26	6	2676	$1.54+0.013N_{rd}$	$-2.54+0.022N_{rd}$	$2.05+0.123N_{rd}$	1.50	−10.2
18	2	3100	$1.38+0.021N_{rd}$	$-2.86+0.017N_{rd}$	$-2.59+0.389N_{rd}$	1.42	−8.5

machirp. In this section we describe the fitting of these IF models to the revcor data presented by Carney *et al.* (1999) in their Fig. 4 (units 2, 7, 18, 20, 22, and 26).

1. Parameter values

We used the analytic gammachirp with eight coefficients and the compressive gammachirp with nine coefficients as indicated by asterisks in Tables II and III, respectively. They both have more than the minimum number of coefficients required to fit one of the units (25), but it seemed desirable to use the same parameter set for all units. It is also the case that the revcor functions for several units (units 2, 18, and 20) were only measured at two signal levels. Fairly substantial preliminary simulation was used to select initial coefficients for each unit to get reasonable convergence. The simulations were then repeated with one additional coefficient (nine for the analytic gammachirp; ten for the compressive gammachirp) to confirm that the results are stable.

The values of coefficients associated with the fits that produced the minimum rms error are listed in Table IV for the analytic gammachirp and in Table V for the compressive gammachirp. The first column shows the unit number; the second column shows the number of noise levels used in the fit; and the column labeled error shows the rms error in dB. The extra column labeled ''cmp'' in Table V shows the compression factor in dB/dB. The remaining columns show the parameter values as before. The average rms error for the analytic and compressive gammachirp are −10.5 dB and −10.4 dB, respectively. So, the quality of the fits is about the same. The problem is that there are not really sufficient noise levels for units other than 25.

For the analytic gammachirp (Table IV), the parameters b_1, c_1, and ϕ_1 are level dependent. Both the envelope parameter, b_1, and chirp parameter, c_1, are negatively correlated with the frequency, f_{r1}, when noise level N_{rd} is constant; the phase parameter, ϕ_1, does not show a clear tendency. It is difficult to imagine a physical realization of this phase term in a cochlear simulation where the parameter is gradually changing with location. For the compressive gammachirp (Table V), only the frequency parameters, f_{r1} and f_{rat}, are level dependent. The phase parameter, ϕ_1, is zero. All of the other parameters are level independent. The delay parameter, τ, is level independent in both filter models and the values are virtually the same for corresponding units in Tables IV and V. The values are about 2 ms when the frequency parameter, f_{r1}, is below 1000 Hz; they are less than 2 ms and negatively correlated to best frequency when f_{r1} is above 1000 Hz.

The average compression of the input-output function is about 0.65 dB/dB when the revcor data span six to nine noise levels and about 0.8 dB/dB when the revcor data span four to five noise levels. No consistent value is obtained for units with only two noise levels. The best estimate of the compression value is obtained from revcor data measured for a wide dynamic range, and a relatively large number, of noise levels. So, a value of about 0.6 dB/dB in the frequency region of 2 kHz seems a reasonable overall summary of the current results with regard to compression. This value is in good agreement with values derived from physiological observations of input-output functions in auditory nerve fibers of guinea-pig with similar CFs (Cooper and Yates, 1994). They are also in good agreement with recent psychophysical mea-

TABLE V. Parameter values, rms errors, and degree of compression for the best-fitting compressive gammachirp filters to the same revcor functions as in Table IV. The last column shows the compression ratio between the input and output of the filter in dB/dB. Columns 3–10 show the best-fitting coefficients when $n_1=4$. See the caption of Table IV for further details.

Unit	No.	f_{r1}	b_1	c_1	ϕ_1	τ	f_{rat}	b_2	c_2	Error	cmp
20	2	$290+0.094N_{rd}$	2.67	−1.15	0	2.10	$1.00+0.0015N_{rd}$	1.96	2.84	−9.5	0.94
2	2	$420+0.004N_{rd}$	1.94	−0.38	0	2.08	$1.21+0.0043N_{rd}$	1.72	2.18	−10.9	0.75
22	4	$1110+2.017N_{rd}$	2.02	−0.63	0	2.33	$1.16+0.0042N_{rd}$	1.51	1.03	−9.8	0.82
7	5	$1600-0.085N_{rd}$	1.62	−1.70	0	1.52	$1.09+0.0030N_{rd}$	0.94	1.30	−10.4	0.81
25	9	$2150-0.144N_{rd}$	1.67	−3.03	0	1.77	$1.26+0.0069N_{rd}$	1.57	1.29	−12.0	0.65
26	6	$2500-0.034N_{rd}$	1.65	−2.92	0	1.66	$1.21+0.0050N_{rd}$	1.38	1.48	−11.2	0.67
18	2	$3100+0.000N_{rd}$	1.36	−3.37	0	1.46	$1.24+0.0168N_{rd}$	1.08	0.99	−8.8	0.25

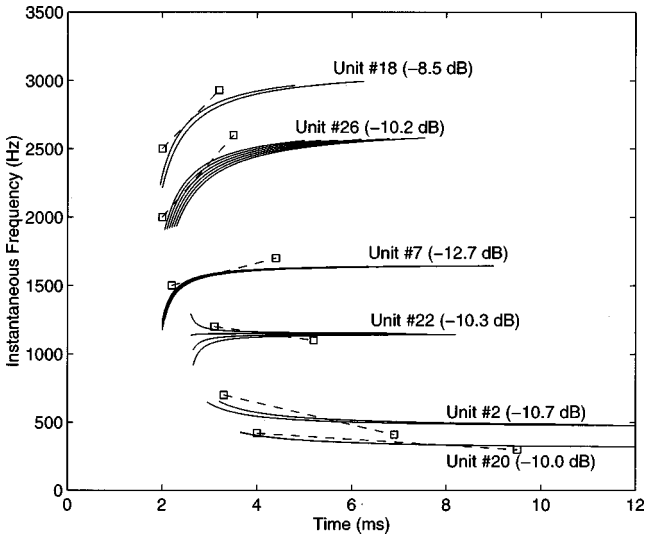


FIG. 8. Instantaneous frequency trajectories for six units from cat 86100. Dashed lines terminated with squares are adapted from the first-order regression lines for 80 dB SPL in Fig. 4 of Carney *et al.* (1999). Solid lines show the instantaneous frequency estimated from the analytic gammachirp filter, for all sound pressure levels. The filter coefficients are shown in Table IV. The unit number and rms error (in parentheses) are shown beside the curves.

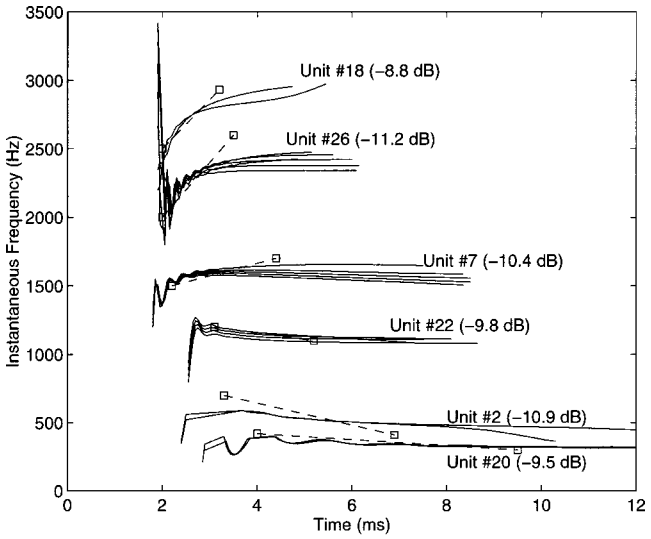


FIG. 9. Instantaneous frequency trajectories for six fibers of cat 86100. Dashed lines show first-order regression lines for 80 dB SPL by Carney *et al.* (1999, Fig. 4). Solid lines show instantaneous frequencies from the best-fitting compressive gammachirp filters. The filter coefficients are shown in Table V. The rms errors are shown in parentheses.

surements of compression derived from forward masking experiments in the frequency region 1500–3000 Hz (Hicks and Bacon, 1999).

2. Instantaneous frequency

The IF trajectories were calculated from the complex responses of the analytic and compressive gammachirp filters using the parameter values listed in Tables IV and V. The dashed lines with squares in Figs. 7 and 8 show the first-order regression lines fitted to the IF values derived at the 80-dB level by Carney *et al.* (1999, Fig. 4) for units, 2, 7, 18, 20, 22, and 26. The solid lines show the IF trajectories within the range where the envelope magnitude is greater than 2% of the envelope peak (the same criterion as was used when calculating the rms error in Sec. II B).

Figure 8 shows that the IF trajectories for the analytic gammachirp flow through the bounds delimited by the regression lines. The IFs for units 7, 26, and 18 glide upward from a very low IF to asymptote in the region of the best frequency of the unit. This is consistent with the results for unit 25 described previously. The IFs for units 2 and 20 glide downward from a very high IF to asymptote in the region of the best frequency of the unit. Theoretically, the downward frequency glide starts at infinity. The IFs for unit 22 glide in to the best frequency either from well below or well above depending on the sign of the chirp parameter, c. For example, the c values are 0.11 and -0.37 for noise levels of 80 and 40 dB, respectively. It is difficult to imagine the mechanism associated with an IF starting from infinite frequency, or one whose starting point flips from infinity to 0 as noise level increases. In summary, the analytic gammachirp provides a reasonable fit to the IF data in terms of rms error, but it does not provide useful insights into the cochlear mechanisms underlying the initial portion of the IF trajectory.

Figure 9 shows that the IF trajectories for the compressive gammachirp also flow through the bounds delimited by the regression lines. The IF trajectories for units 2, 20, and 22 rise rapidly from zero frequency to a maximum value and, then, *decrease* gradually. Since the IFs do not start from infinite frequency, the responses are more realistic in terms of the underlying cochlear mechanics. The trajectories for units 26 and 18 have initial frequencies higher than their asymptotic frequencies and they oscillate on the way to the asymptote. The maxima in these functions arise from low-amplitude parts of the filter response; no corresponding maxima are observable in the revcor data because of the low signal-to-noise ratio. It is, nevertheless, interesting to note that there are oscillations in the original IF data of units 26 and 18 (Fig. 4 of Carney *et al.*, 1999), although they do not have the same period partly due to methodological differences in the IF calculation. In summary, the compressive gammachirp provides a reasonable fit to the IF data in terms of rms error.

IV. FITTING THE COMPRESSIVE GAMMACHIRP TO HUMAN MASKING DATA

Irino and Patterson (1997) demonstrated that the analytic gammachirp filter can explain a wide range of notched-noise masking data. The most demanding data were those of Rosen and Baker (1994) who measured masking over a large dynamic range. In this section, we show how the compressive gammachirp can be fitted to these data, and thereby raise the prospect that the compressive gammachirp filter can provide the basis for a unified model of frequency selectivity in the auditory system.

A. Method and parameters

In the psychophysical experiment, the observer is required to detect a brief probe tone in the presence of a low-pass or high-pass noise masker (Patterson, 1974). As the dis-

placement between the tone frequency and the edge of the noise increases, threshold decreases over a range that can be as much as 60 dB. It is assumed that the observer listens for the tone through an auditory filter centered in the region of the tone to maximize the signal level and minimize the noise level in the decision variable. The slope of the noise cutoff is very steep and the spectrum level of the noise is flat, and in this case, the function relating probe threshold to the frequency of the noise cutoff is closely related to the integral of the auditory filter. This relationship provides the basis for the fitting procedure (Patterson *et al.*, 1982) and the large dynamic range of the threshold measurements means that the procedure can be highly sensitive to filter shape. Over the years, the assumptions of the model have been tested and the procedure has been improved in conjunction with the fitting technique, for example, to ensure that the auditory filter is centered close to the tone frequency and to accommodate filter asymmetry (Patterson and Nimmo-Smith, 1980). Reviews of the development and extension of the technique are provided in Patterson and Moore (1986) and Rosen and Baker (1994). Basically, however, the technique remains the same; probe level at threshold is used in conjunction with carefully designed noise maskers to estimate the integral of the auditory filter, and functions like the gammachirp are then fitted to sets of these thresholds to determine the optimal parameter values to represent the auditory filter.

The compressive gammachirp was fitted to the notched-noise data of Rosen and Baker (1994) using basically the same procedure as in Irino and Patterson (1997).[4] The Levenberg–Marquardt method (Press *et al.*, 1988) was employed to minimize the rms difference between the masked threshold in dB and the predicted value of masked threshold, also in dB. Since the data are now thresholds values, in dB, the value of the rms error is somewhat different from that used to fit to revcor data. It is also worth noting that the rms errors are based on values with different units near the probe frequency, unlike the rms error in the fit to a single unit of revcor data. These factors do not affect the consistency of filter derivation with the revcor data because a single compressive gammachirp filter is used for each parameter set. The probe frequency in the psychophysical experiment was 2000 Hz and the data were gathered for a very wide range of both probe and masker levels.

The compressive gammachirp has essentially nine parameters (n_1, b_1, c_1, f_{r1}, ϕ_1, b_2, c_2, f_{rat}, and τ) as described in Eqs. (1)–(8) of Sec. II. However, the phase parameter, ϕ_1, and the delay parameter, τ, cannot be specified in this case because we use the magnitude spectrum, $|G_{\mathrm{CC}}(f)|$, to fit the masking data. Instead, the procedure requires two other parameters which are the efficiency of the detector, K, that determines the listeners response, and a limit on the dynamic range, r, of the filter. In this case, r was set to -100 dB which is virtually zero. The amplitude parameter a_c falls out of the fitting equations with masking data since both the data and the filter response are normalized. Once again, the order of the gamma distribution, n_1, was fixed at 4.

To control for off-frequency listening (Patterson and Nimmo-Smith, 1980), the asymptotic frequency, f_{r1}, of the analytic gammachirp, $g_{ca}(t)$, and thus its peak frequency, f_{p1}, were varied to locate the filter position that produces the maximum signal-to-noise ratio in each listening condition. The fits to the revcor data (Tables III and V) show that the frequency ratio, f_{rat}, needs to vary with level to produce a good fit, but that the other parameters do not. Accordingly, the frequency ratio, f_{rat}, was set to be a linear function of probe level, P_s, in dB; the remaining parameters (b_1, c_1, b_2, c_2, and K) were taken to be level-independent variables.

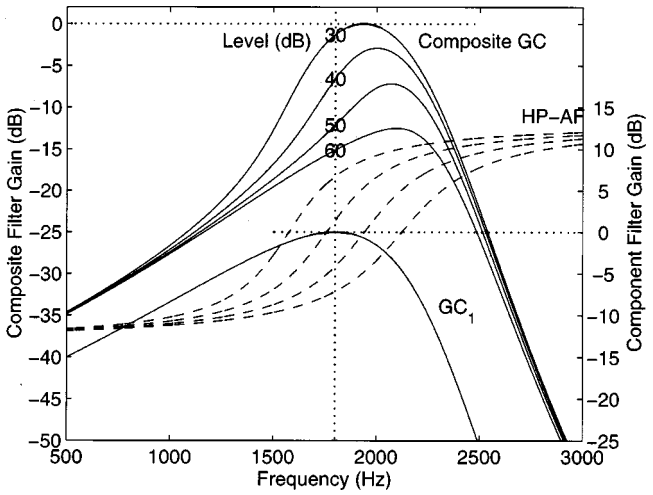


FIG. 10. The set of compressive gammachirp filters, GC (solid lines; left ordinate) derived from Rosen and Baker's data (1994). The filters were produced by cascading a passive gammachirp filter, GC_1 (lower solid line; right ordinate) with a high-pass asymmetric function, HP-AF (dashed lines; right ordinate) whose center frequency varies with level.

B. Results

The best fit resulted in a rms error of 1.27 (dB), which is slightly better than the 1.36 dB produced with by Irino and Patterson (1997) with the original, analytic gammachirp.[5] The parameter values in this case are $b_1=2.02$; $b_2=1.14$; $c_1=-3.70$; $c_2=0.979$; $f_{\mathrm{rat}}=0.573+0.0101P_s=1.381+0.0101\cdot(P_s-80)$; $K=-5.03$. So, it is sufficient to make the ratio of frequencies level dependent to explain the masking data; none of the other parameters need vary with level in this specific case. It is also interesting to note that, although the masking data and the revcor data are completely different, the parameter values for the masking data fall within the range of parameter values listed in Table V for the fit of the compressive gammachirp to revcor data.

The magnitude spectra for the best-fitting filters are shown in Fig. 10 for the case where the peak frequency, f_{p1}, of the fixed, analytic gammachirp (lower solid line) is 1800 Hz (the asymptotic frequency, f_{r1}, is about 2300 Hz). Since the frequency ratio, f_{rat}, is level dependent, the high-pass asymmetric function is level dependent as indicated by the set of dashed curves. Consequently, the composite, compressive gammachirp filter is level dependent as shown by the set of solid lines in the upper portion of the figure. As signal level increases, the peak gain decreases and the bandwidth generally increases (although it is slightly non-monotonic). Here, then, is an auditory filter with level-dependent gain in

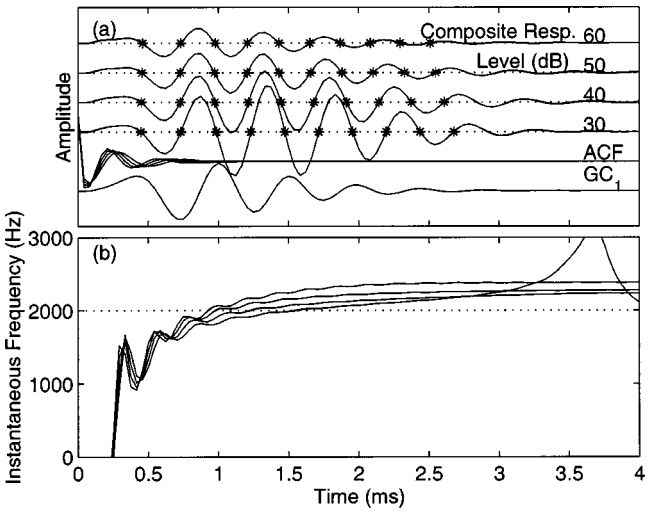


FIG. 11. The impulse response (a) and instantaneous frequency trajectory (b) of the compressive gammachirp filter shown in Fig. 10. In panel (a), the bottom line shows the response of the initial gammachirp filter, GC_1, and just above it the responses of the high-pass, asymmetric compensation filters, ACF. The upper four lines show the responses of the composite compressive gammachirp filters as a function of stimulus level (dB). Since the responses are normalized by the input signal level, the maximum amplitude corresponds to the filter gain. The asterisks show the times of the seven zero-crossings in the responses. In panel (b), the four lines present the instantaneous frequency trajectory for the four composite gammachirp filters in panel (a).

the passband and linear tails outside the passband, as required physiologically, that can accommodate human masking data over a wide dynamic range. In contrast to the physiological observations, the peak frequency increases slightly with level. This is because the peak frequency, f_{p1}, of the analytic gammachirp is assumed to be fixed. The discrepancy with physiological data can be resolved if f_{p1} is made level dependent and negatively correlated with level, as is the case for units 25 and 26 in Table V. It is not possible, however, to determine the relationship between f_{p1} and level with notched-noise data because the fitting procedure includes the assumption that the listeners use ''off-frequency listening'' to optimize signal detection (Patterson and Nimmo-Smith, 1980). That is, the listeners use the filter that gives the maximum signal-to-noise ratio in any given condition and the optimum filter varies with masking condition.

The impulse responses of the filters shown in Fig. 10 are plotted in Fig 11(a). In this case, the unspecified parameters were set to zero (the phase, ϕ_1, and the latency, τ). The filter gain (relative amplitude) decreases as level increases, but the locations of the zero crossings (indicated by asterisks) are almost level independent at least for the first few cycles where the amplitude of the filter is sufficiently high. The filter responses are similar to those from the revcor data shown in Fig. 6(a) except that the gain of the revcor data was normalized. The IF trajectories calculated from the impulse responses in Fig. 11(a) are plotted in Fig. 11(b). The characteristics of the frequency-glide are quite similar to those of the IF trajectories shown in Fig. 6(b). So, the results are consistent with the physiological observations of Carney *et al.* (1999).

The gain of the compressive gammachirp filter decreases as level increases. So, there is compression between the input and output of the filter in the passband. The average compression is about 0.6 dB/dB which is similar to the values obtained with the revcor data of units 25 and 26 in Table V.

V. SUMMARY

The analytic, gammachirp auditory filter of Irino and Patterson (1997) has been extended to produce a new gammachirp auditory filter that can provide a unified framework for both physiological and psychophysical modeling of frequency selectivity and compression. This compressive gammachirp filter consists of an analytic gammachirp with a fixed, negative chirp parameter and a high-pass asymmetric function with a positive chirp parameter. The compressive gammachirp filter provides a better fit to the revcor data of Carney *et al.* (1999), in terms of rms error, than either the gammatone auditory filter or the analytic gammachirp filter. It is also more compatible with observations of basilar membrane motion and primary auditory nerve data which show that the chirp in the impulse response is not level dependent. The compressive gammachirp filter was also fitted to the human masking data of Rosen and Baker (1994) without the level-dependent chirp previously required to fit such data. It appears that both cat revcor data and human masking data can be explained with a single compressive gammachirp filter that has only one level-dependent parameter, which is the ratio of the frequencies that defined the positions of the filter's components—the fixed analytic gammachirp and the high-pass asymmetric function. Thus, the compressive gammachirp filter provides a relatively simple model for characterizing the cochlear frequency glide and predicting human masking data within a unified framework.

ACKNOWLEDGMENTS

We would like to thank L. Carney for providing us with the revcor data and Richard Baker for helpful comments on an earlier draft of the paper. This research was mainly performed when the first author was seconded to, and the second author was a visiting researcher at ATR-HIP, Kyoto. The research was partially supported by CREST (Core Research for Evolutional Science and Technology) of the Japan Science and Technology Corporation (JST) and a grant from ATR-HIP to the second author at the Center for the Neural Basis of Hearing in Cambridge.

APPENDIX A: AMPLITUDE FACTOR OF THE GAMMACHIRP

The Fourier magnitude spectrum of the gammachirp filter in Eq. (1) is

$$|G_C(f)| = \frac{a|\Gamma(n_1+jc_1)|}{|2\pi\sqrt{\{b_1\mathrm{ERB}(f_{r_1})\}^2+(f-f_{r_1})^2}|^n}\cdot \exp(c_1\,\theta_1)$$

$$= \frac{a\,|\Gamma(n_1+jc_1)|}{\Gamma(n_1)} \times \frac{\Gamma(n_1)}{|2\pi\sqrt{\{b_1\mathrm{ERB}(f_{r_1})\}^2+(f-f_{r_1})^2}|^n}\cdot \exp(c_1\,\theta_1)$$

$$= a_\Gamma \cdot |G_T(f)| \cdot \exp(c_1\,\theta_1), \tag{A1}$$

$$\theta_1 = \arctan\left(\frac{f-f_{r_1}}{b_1\mathrm{ERB}(f_{r_1})}\right). \tag{A2}$$

The scalar values a, b_1, c_1, f_{r_1}, and n_1 are the amplitude, bandwidth, chirp factor, asymptotic frequency, and envelope factor of the filter, and $\mathrm{ERB}(f_{r_1})$ is the equivalent rectangular bandwidth. The first term, a_Γ, in Eq. (A1) is the power normalization factor from the definition of the Fourier transform. When the envelope factor, n_1, is an integer,

$$a_\Gamma = \frac{a|\Gamma(n_1+jc_1)|}{\Gamma(n_1)} = \frac{a}{\Gamma(n_1)}\cdot\sqrt{\frac{\pi}{c_1\sinh(\pi c_1)}}\cdot\prod_{k=0}^{n_1-1}|k+jc_1|$$

$$= a\sqrt{\frac{\pi c_1}{\sinh(\pi c_1)}}\cdot\prod_{k=1}^{n_1-1}\left|1+j\frac{c_1}{k}\right|. \tag{A3}$$

The second factor in Eq. (A1), $|G_T(f)|$, is the Fourier magnitude spectrum of the gamma*tone* auditory filter. The third factor, $\exp(c_1\,\theta_1)$, is the asymmetric function which introduces asymmetry into the frequency response and converts it into a gamma*chirp* auditory filter. This is a lowpass filter when $c_1<0$, a highpass filter when $c_1>0$, and unity when $c_1=0$.

The peak gain of the analytic gammachirp filter in Fig. 1 is

$$\frac{|G_C(f_{p1})|}{|G_C(f_{r_1})|} = \frac{\exp\{c_1\cdot\arctan(c_1/n_1)\}}{\{1+(c_1/n_1)^2\}^{n_1/2}}, \tag{A4}$$

where f_{p_1} is the peak frequency defined in Eq. (5).

The peak gain can be normalized to unity (0 dB) using Eq. (A4). It is also possible to normalize the power using a_Γ in Eq. (A3). In this case, the peak gain decreases slightly as the chirp factor, c, decreases (to about -0.5 dB when $c_1=-2$). There is, however, no clear physiological criterion for selecting a particular gain factor.

In the process of fitting the gammachirp to human masking data, Irino and Patterson (1997) attributed any shifts in peak frequency to 'off-frequency listening' (Patterson and Nimmo-Smith, 1980), and any changes in gain were aggregated into the efficiency constant, K, which is described with reference to the roex filter in Patterson *et al.* (1982). So, the peak gain was essentially normalized to unity when fitting the data. As a result, the effects of level on gain and peak frequency, and thus the relationship to physiological data, were not discussed in that paper.

The amplitude factor cannot be derived from the revcor data because the discharge rates of auditory nerve fibers saturate as the noise level increases (Patuzzi and Robertson, 1988; Carney *et al.*, 1999). As a result, both the amplitude of the filter response and the data were normalized in the process of fitting filters to the revcor data. The compressive gammachirp filter has the advantage that the compression characteristics can be determined without the need to specify the amplitude factor, as demonstrated in Fig. 3. It does, however, mean that further experiments will be necessary to determine whether the amplitude factor is level-dependent.

APPENDIX B: ASYMMETRIC COMPENSATION FILTER (ACF) AND ASYMMETRIC FUNCTION

Equation (7) shows that the compressive gammachirp, $G_{CC}(f)$, is defined as the combination of another gammachirp, $G_{CA}(f)$, and an asymmetric function, $\exp(c_2\,\theta_2)$. The impulse response of the gammachirp is described in Eq. (1). The impulse response of the asymmetric function, however, is not well-defined because the phase response is not specified in Eqs. (3) and (7). Recently, while implementing a gammachirp filterbank for auditory modeling, Irino and Unoki (1998, 1999) developed an IIR asymmetric compensation filter, $H_C(f)$, to simulate the asymmetric function such that

$$|H_C(f)| \cong \exp(c\cdot\theta), \tag{B1}$$

$$G_C(f) \cong G_T(f)\cdot H_C(f), \tag{B2}$$

and, for the impulse response,

$$g_c(t) \cong g_t(t) * h_c(t). \tag{B3}$$

The asymmetric compensation filter is defined in the z plane as

$$H_C(z) = \prod_{k=1}^{N} H_{Ck}(z), \tag{B4}$$

$$H_{Ck}(z) = \frac{(1-r_k e^{j\varphi_k} z^{-1})(1-r_k e^{-j\varphi_k} z^{-1})}{(1-r_k e^{j\phi_k} z^{-1})(1-r_k e^{-j\phi_k} z^{-1})}, \tag{B5}$$

$$r_k = \exp\{-k\cdot p_1\cdot 2\pi b\mathrm{ERB}(f_r)/f_s\}, \tag{B6}$$

$$\phi_k = 2\pi\{f_r + p_0^{k-1}\cdot p_2\cdot c\cdot b\mathrm{ERB}(f_r)\}/f_s, \tag{B7}$$

$$\varphi_k = 2\pi\{f_r - p_0^{k-1}\cdot p_2\cdot c\cdot b\mathrm{ERB}(f_r)\}/f_s, \tag{B8}$$

where p_0, p_1, and p_2 are positive coefficients, f_s is the sampling rate, and N is the number of filters in the cascade. When $N=4$, $p_0=2$, and p_1 and p_2 are properly set as a function of $|c|$, the rms error in the Fourier magnitude spectra between the original gammachirp filter, $|G_C(f)|$, and the composite filter, $|G_T(f)|\cdot|H_C(f)|$, is reduced to less than 1 dB within the range required by parameters b and c. The accuracy is greatest near the asymptotic frequency, f_r, and, as a result, the impulse responses are in excellent agreement. [For details, see Irino and Unoki (1998, 1999).]

As a result, the impulse response of the asymmetric compensation filter is used to fit the revcor data in the time domain, and the asymmetric function is used to fit the human masking data in the frequency domain.

[1]The idea of combining an analytic gammachirp with a high-pass asymmetric function was originally inspired by the NonLinear Resonant Tectorial Membrane (NL-RTM) model of Allen (1997, 1998). In this model, two filters interact as a function of level: the first filter has a bandpass characteristic and it represents the basilar membrane (BM) travelling wave; the second filter has more of a highpass characteristic and it represents the interaction between the basilar and tectorial membranes (BM-TM filter). The location of the peak of the BM motion moves relative to the BM-TM filter as a function of signal level. The magnitude of mechanical vibration is plotted as a function of cochlear location, as for an excitation pattern. These excitation patterns can be reinterpreted as the magnitude spectra of auditory filters using the conversion scheme proposed by Glasberg and Moore (1990), and in this case, the filter shape is found to be similar to those in Fig. 3. It still remains, however, to determine whether the conversion is applicable to physiological observations of basilar membrane motion.

[2]It is easy to show that the chirp parameter, c_1, affects the initial phase. When $c_1 \ln t$ in Eq. (1) is expanded to $c_1 \ln t/t_0 + c_1 \ln t_0$, t_0 is a constant, and so $c_1 \ln t_0$ is a phase term.

[3]We fitted gammachirp filters to revcor functions derived with sustained noise and based on the assumption of linearity. It has been reported that when the measurements are made with clicks, the envelope becomes longer as level increases (Lin and Guinan, 2000), which contradicts the linearity assumption. To accommodate such data, it would be essential to introduce a mechanism to estimate the internal level of a transient and a controller to set the appropriate values of the filter parameters for that level. Such a mechanism is beyond the scope of this article.

[4]There are two further differences between the fit of Irino and Patterson (1997) and that used here: a correction for equal loudness contour (ELC) has been included as suggested by Glasberg and Moore (1990), and the fixed absolute threshold value of 22.7 dB was omitted because it was initially introduced by Rosen and Baker (1994) for their specific data and is not applicable to other masking data.

[5]The fit with the original gammachirp (in Fig. 2 of Irino and Patterson, 1997) was recalculated with the conditions in footnote 4; the result was 1.36 dB, which is slightly higher than 1.33 dB originally reported.

Allen, J. B. (**1997**). ''OHCs shift the excitation pattern via BM tension,'' in *Diversity in Auditory Mechanics*, edited by Lewis, E. R., *et al.* (World Scientific, Singapore).

Allen, J. B. (**1998**). ''A bio-mechanical model of the ear to predict auditory masking,'' in Proc. Computational Hearing (NATO Advanced Study Institute), Il Ciocco, Italy.

Carney, L. H. (**1993**). ''A model for the response of low-frequency auditory-nerve fibers in cat,'' J. Acoust. Soc. Am. **93**, 401–417.

Carney, L. H., and Yin, T. C. T. (**1988**). ''Temporal coding of resonances by low-frequency auditory nerve fibers: single-fiber responses and a population model,'' J. Neurophysiol. **60**, 1653–1677.

Carney, L. H., Megean, J. M., and Shekhter, I. (**1999**). ''Frequency glides in the impulse responses of auditory-nerve fibers,'' J. Acoust. Soc. Am. **105**, 2384–2391.

Cooke, M. (**1993**). *Modelling Auditory Processing and Organisation* (Cambridge U. P., Cambridge).

Cooper, N. P., and Yates, G. K. (**1994**). ''Nonlinear input-output functions derived from the responses of guinea-pig cochlear nerve fibres: Variations with characteristic frequency,'' Hear. Res. **78**, 221–234.

de Boer, E. (**1975**). ''Synthetic whole-nerve action potentials for the cat,'' J. Acoust. Soc. Am. **58**, 1030–1045.

de Boer, E., and de Jongh, H. R. (**1978**). ''On cochlear encoding: Potentialities and limitations of the reverse-correlation technique,'' J. Acoust. Soc. Am. **63**, 115–135.

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.

Glasberg, B. R., and Moore, B. C. J. (**1990**). ''Derivation of auditory filter shapes from notched-noise data,'' Hear. Res. **47**, 103–138.

Hicks, L. M., and Bacon, S. P. (**1999**). ''Psychophysical measures of auditory nonlinearities as a function of frequency in individuals with normal hearing,'' J. Acoust. Soc. Am. **105**, 326–338.

Irino, T., and Patterson, R. D. (**1997**). ''A time-domain, level-dependent auditory filter: The gammachirp,'' J. Acoust. Soc. Am. **101**, 412–419.

Irino, T., and Patterson, R. D. (**2000**). ''A gammachirp framework for auditory filtering: Unification of cochlear frequency-glide data and psychoacoustical masking data,'' 12th International Symposium on Hearing, Mierlo, The Netherlands.

Irino, T., and Unoki, M. (**1998**). ''A time-varying, analysis/synthesis auditory filterbank using the gammachirp,'' IEEE Int. Conf. Acoust. Speech Signal Processing (ICASSP-98), Seattle, WA, pp. 3653–3656.

Irino, T., and Unoki, M. (**1999**). ''A analysis/synthesis auditory filterbank based on an IIR implementation of the gammachirp,'' J. Acoust. Soc. Jpn. **20**, 397–406.

Johannesma, P. I. M. (**1972**). ''The pre-response stimulus ensemble of neurons in the cochlear nucleus,'' in *Symposium on Hearing Theory* (IPO, Eindhoven, The Netherlands), pp. 58–69.

Lin, T., and Guinan, J. J., Jr. (**2000**). ''Auditory-nerve-fiber responses to high-level clicks: Interference patterns indicate that excitation is due to the combination of multiple drives,'' J. Acoust. Soc. Am. **107**, 2615–2630.

Lutfi, R. A., and Patterson, R. D. (**1984**). ''On the growth of masking asymmetry with stimulus intensity,'' J. Acoust. Soc. Am. **76**, 739–745.

Lyon, R. F. (**1996**). ''The all-pole gammatone filter and auditory models,'' Forum Acusticum '96, Antwerp, Belgium.

Lyon, R. F. (**1997**). ''All-pole models of auditory filtering,'' in *Diversity in Auditory Mechanics*, edited by Lewis, E. R., *et al.* (World Scientific, Singapore).

Moore, B. C. J., and Glasberg, B. R. (**1987**). ''Formulae describing frequency selectivity as a function of frequency and level, and their use in calculating excitation patterns,'' Hear. Res. **28**, 209–225.

Moore, B. C. J., Peters, R. W., and Glasberg, B. R. (**1990**). ''Auditory filter shapes at low center frequencies,'' J. Acoust. Soc. Am. **88**, 132–140.

Møller, A. R., and Nilsson, H. G. (**1979**). ''Inner ear impulse response and basilar membrane modelling,'' Acustica **41**, 258–262.

Oxenham, A. J., and Plack, C. J. (**1997**). ''A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing,'' J. Acoust. Soc. Am. **101**, 3666–3675.

Patterson, R. D. (**1974**). ''Auditory filter shape,'' J. Acoust. Soc. Am. **55**, 802–809.

Patterson, R. D. (**1976**). ''Auditory filter shapes derived with noise stimuli,'' J. Acoust. Soc. Am. **59**, 640–654.

Patterson, R. D., and Moore, B. C. J. (**1986**). ''Auditory filters and excitation patterns as representations of frequency resolution,'' in *Frequency Selectivity in Hearing*, edited by Moore, B. C. J. (Academic Press, London), pp. 123–177.

Patterson, R. D., and Nimmo-Smith, I. (**1980**). ''Off-frequency listening and auditory-filter asymmetry.'' J. Acoust. Soc. Am. **67**, 229–245.

Patterson, R. D., Allerhand, M., and Giguère, C. (**1995**). ''Time-domain modelling of peripheral auditory processing: a modular architecture and a software platform,'' J. Acoust. Soc. Am. **98**, 1890–1894.

Patterson, R. D., Nimmo-Smith, I., Weber, D. L., and Milroy, R. (**1982**). ''The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold,'' J. Acoust. Soc. Am. **72**, 1788–1803.

Patterson, R. D., Robinson, K., Holdsworth, J. W., McKeown, D., Zhang, C., and Allerhand, M. (**1992**). ''Complex sounds and auditory images,'' in *Auditory Physiology and Perception*, edited by Cazals, Y., Demany, L. and Horner, K. (Pergamon, Oxford), 429–446.

Patuzzi, R., and Robertson, D. (**1988**). ''Tuning in the mammalian cochlea,'' Physiol. Rev. **68**, 1009–1081.

Pickles, J. O. (**1988**). *An Introduction to the Physiology of Hearing* (Academic, London).

Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (**1988**). *Numerical Recipes in C* (Cambridge U. P., Cambridge).

Recio, A. R., Rich, N. C., Narayan, S. S., and Ruggero, M. A. (**1998**). ''Basilar-membrane response to clicks at the base of the chinchilla cochlea,'' J. Acoust. Soc. Am. **103**, 1972–1989.

Rose, S., and Baker, R. J. (**1994**). ''Characterising auditory filter nonlinearity,'' Hear. Res. **73**, 231–243.

Ruggero, M. A. (**1992**). ''Responses to sound of the basilar membrane of the mammalian cochlea,'' Curr. Opin. Neurobiol. **2**, 449–456.
Ruggero, M. A., Rich, N. C., Recio, A. R., Narayan, S. S., Robles, L. (**1997**). ''Basilar-membrane response to tones at the base of the chinchilla cochlea,'' J. Acoust. Soc. Am. **101**, 2151–2163.
Schofield, D. (**1985**). ''Visualizations of speech based on a model of the peripheral auditory system,'' NPL Report DITC 62/85.
Slaney, M. (**1993**). ''An efficient implementation of the Patterson-Holdsworth auditory filterbank,'' Apple Computer Technical Report #35.

Frequency glides in click responses of the basilar membrane and auditory nerve: Their scaling behavior and origin in traveling-wave dispersion

Christopher A. Shera[a)]
Eaton-Peabody Laboratory of Auditory Physiology, Massachusetts Eye and Ear Infirmary, 243 Charles Street, Boston, Massachusetts 02114 and Department of Otology and Laryngology, Harvard Medical School, Boston, Massachusetts 02115

(Received 17 November 2000; accepted for publication 27 February 2001)

Frequency modulations (or glides), reported in impulse responses of both the auditory nerve and the basilar membrane, represent a change over time in the instantaneous frequency of oscillation of the response waveform. Although the near invariance of glides with stimulus intensity indicates that they are not the consequence of nonlinear or active processes in the inner ear, their origin has remained otherwise obscure. This paper combines theory with experimental data to explore the basic phenomenology of glides. When expressed in natural dimensionless form, glides are shown to have a universal form nearly independent of cochlear location for characteristic frequencies (CFs) above approximately 1.5 kHz (the ''scaling region''). In the apex of the cochlea, by contrast, glides appear to depend strongly on CF. In the scaling region, instantaneous-frequency trajectories are shown to be approximately equal to the ''inverse group delays'' of basilar-membrane transfer functions measured at the same locations. The inverse group delay, obtained by functionally inverting the transfer-function group-delay-versus-frequency curve, specifies the frequency component of a broadband stimulus expected to be driving the cochlear partition at the measurement point as a function of time. The approximate empirical equality of the two functions indicates that glides are closely related to cochlear traveling-wave dispersion and suggests that they originate primarily through the time dependence of the effective driving pressure force at the measurement location. Calculations in a one-dimensional cochlear model based on solution to the inverse problem in squirrel monkey [Zweig, J. Acoust. Soc. Am. **89**, 1229–1254 (1991)] support this conclusion. In contrast to previous models for glides, which locate their origin in the differential build-up and decay of multiple micromechanical resonances local to each radial cross section of the organ of Corti, the model presented here identifies glides as the global consequence of the dispersive character of wave propagation in the cochlea.
[DOI: 10.1121/1.1366372]

PACS numbers: 43.64.Bt, 43.64.Kc, 43.66.Ba [LHC]

I. INTRODUCTION

Since their first description in derived impulse responses of auditory nerve fibers (Møller and Nilsson, 1979), frequency modulations (or ''glides'') have been reported in responses of both the basilar-membrane (e.g., de Boer and Nuttall, 1997; Recio *et al.*, 1998; Recio and Rhode, 2000) and the auditory nerve (e.g., Carney *et al.*, 1999; Lin and Guinan, 2000). Glides represent a change over time in the instantaneous frequency of oscillation of the response waveform. Frequency modulations in the response are remarkably robust to changes in the measurement paradigm. Glides are seen, for example, both in mechanical and neural impulse responses to acoustic clicks (e.g., Recio *et al.*, 1998; Lin and Guinan, 2000; Recio and Rhode, 2000) as well as in estimates of impulse responses obtained by cross- or reverse correlation using broadband noise stimuli (e.g., de Boer and Nuttall, 1997; Carney *et al.*, 1999). Furthermore, glides are nearly independent of stimulus intensity over the full dynamic range of hearing (e.g., de Boer and Nuttall, 1997; Recio *et al.*, 1998; Carney *et al.*, 1999) and even maintain their general form postmortem (e.g., Recio *et al.*, 1998). Although glides appear largely invariant to changes in the stimuli, auditory-nerve responses indicate that glides do depend on measurement location within the cochlea (Carney *et al.*, 1999): Whereas instantaneous-frequency trajectories recorded from the base ($CF \gtrsim 1500$ Hz) increase with time, approaching the local characteristic frequency (CF) from below, trajectories appear mirrored in the apex ($CF \lesssim 750$ Hz), where instantaneous frequencies generally decrease with time, approaching CF from above.[1]

Although the near invariance of glides with stimulus intensity and physiological condition indicates that glides are not the consequence of physiologically active processes in the inner ear, their origin has remained otherwise obscure. Glides have been derived as components of ''optimal auditory filters,'' (Irino and Patterson, 1997, 2000), simulated using bandpass filters constructed from an array of artfully placed poles in the complex plane (de Boer and Nuttall, 1996; Shekhter and Carney, 1997), and suggested to originate via the differential build-up or decay of multiple micromechanical resonances and/or vibrational modes within the

a)Electronic mail: shera@epl.meei.harvard.edu

organ of Corti (de Boer and Nuttall, 1995; Lin and Guinan, 2000; Mountain, 2000). Even before the additional complexity of strong variation in glides with CF was appreciated, de Boer and Nuttall (1997) proposed that consistency with glides become a standard for evaluating cochlear models, arguing that there ''is currently little or no theoretical basis for understanding the glide phenomenon.''

Motivated by the wide range of mechanisms previously proposed, we take in this paper a step toward providing a theoretical account of glides. Here, we attempt to understand how much of the phenomenology can be explained within the framework of a simple cochlear model. We focus primarily on glides as manifest in the basal turns of the cochlea, where mechanical responses are better characterized and where, as we show below, glides appear to obey a simple scaling relation.

II. SCALING BEHAVIOR OF GLIDES

A. Scaling and nonscaling regions

Carney *et al.* (1999) computed instantaneous-frequency trajectories $f_{\text{in}}(x,t)$ from auditory-nerve-fiber (ANF) impulse responses (Carney and Yin, 1988) derived by reverse-correlation (revcor functions).[2] Carney *et al.* quantified glides by estimating the average time rate of change of the instantaneous frequency (IF) from a linear fit to the trajectory in the region about the peak of the envelope of the revcor function. The resulting ''glide slopes'' (in kHz/ms), assembled from a population of fibers with characteristic frequencies (CFs) spanning the range 0.25–4.5 kHz, are reproduced in Fig. 1(a) together with a loess trend line to guide the eye (Cleveland, 1993). As reported by Carney *et al.*, the scatterplot reveals considerable systematic variation in glide slope. The trend line indicates that glide slopes generally increase with increasing CF. Slopes are typically negative at the lowest CFs, pass through zero near 750 Hz, and increase strongly above zero at higher CFs.

To understand how much of this variation in glide slope reflects actual changes in the form of the glides, as opposed to changes in the characteristic frequency, and thus the effective local time scale, it is helpful to remove the dependence on this time scale by expressing glide slopes in natural dimensionless units. We do this by expressing all frequencies relative to the fiber CF and all time intervals in periods of that CF. We thus transform a glide slope in kHz/ms according to (kHz/CF)/(ms·CF); the net effect is to divide the glide slopes in Fig. 1(a) by the square of the characteristic frequency.[3] The resulting dimensionless glide slopes are shown in Fig. 1(b). The behavior of the trend line suggests a natural division into two regions based on fiber CF: (1) an apical region below roughly 1.5 kHz where the dimensionless glide slope varies strongly with CF; and (2) a basal ''scaling region'' above roughly 1.5 kHz where the average slope appears approximately independent of CF. Although scatter in the data makes the precise boundary between regions difficult to determine—and no sharp border is likely to exist in any case—the existence of this scaling region suggests that the form of the glide remains approximately invariant in the basal turns of the cochlea.[4]

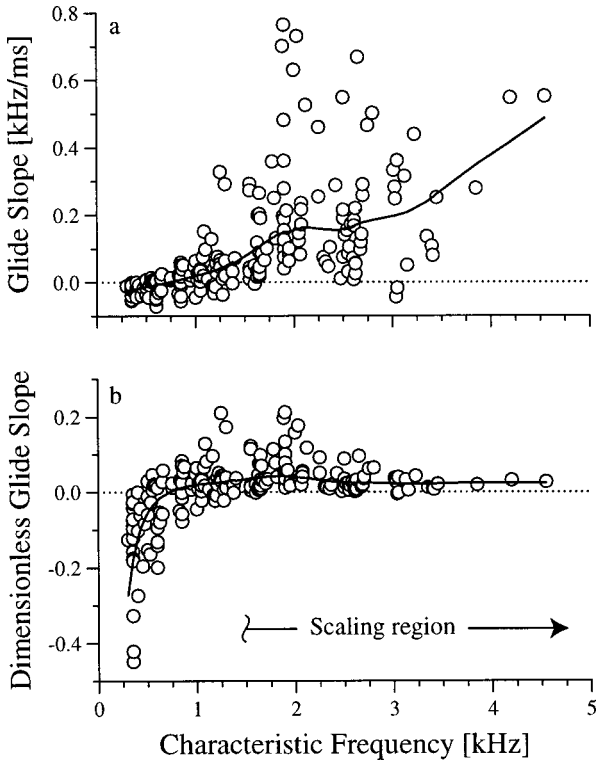


FIG. 1. Glide slopes and their variation with CF. The top graph [panel (a)] shows a scatterplot versus CF of glide slopes computed by Carney *et al.* (1999, Fig. 7) from the instantaneous-frequency trajectories of revcor functions obtained from 214 auditory-nerve fibers from eight cats. For each fiber, the value shown as CF represents the mean best frequency averaged over the stimulus intensities studied. A loess trend line (Cleveland, 1993) computed from the data is shown to guide the eye. (We discuss the variability among fibers of comparable CF in Sec. II A.) The bottom graph [panel (b)] shows the same data expressed in natural dimensionless form obtained by normalizing glide slopes by the square of the fiber CF (in kHz). Constancy of the dimensionless glide slope at CFs above roughly 1.5 kHz suggests that the form of the glide remains nearly invariant in the basal turns of the cochlea (the ''scaling region'').

B. Invariance of glides in the basal scaling region

Figure 2 demonstrates the approximate invariance of glides in the basal scaling region. The top panel shows IF trajectories $f_{\text{in}}(x,t)$ computed from basilar-membrane (BM) measurements at six different locations in the chinchilla cochlea (six different cochleae) by Recio and Rhode (2000, Fig. 4). To remove delays associated with the acoustic system and middle-ear transmission, a delay of 0.2 ms, representing the approximate latency to the onset of umbo vibration[5] (Recio and Rhode, 2000), was subtracted from the data.[6] Together, the trajectories represent CFs spanning about $1\frac{1}{2}$ octaves. Although differences among animals increase the scatter and preclude a definitive comparison, to a good first approximation plotting the trajectories in the dimensionless form suggested by Fig. 1 unifies the data across CF (bottom panel).[7] Although additional data may reveal systematic deviations from scaling—such as those evident in measurements of stimulus-frequency otoacoustic emission group delay (Shera and Guinan, 2000a, b)—Fig. 2(b) indicates that glides in the basal region of the chinchilla cochlea have a nearly universal form independent of CF.

Invariance of the normalized IF trajectory across CF

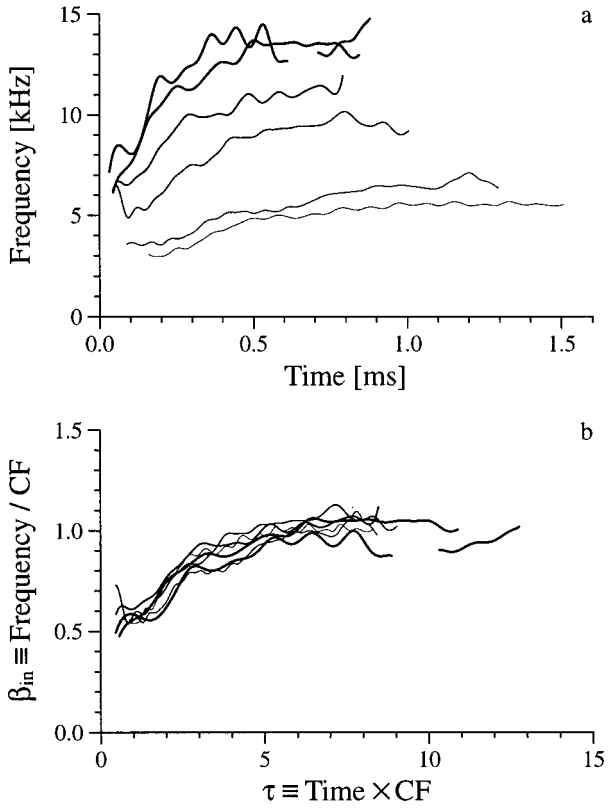


FIG. 2. Approximate invariance of glides across CF. The top graph [panel (a)] shows IF trajectories obtained from BM impulse responses measured at six locations in six chinchilla cochleae (Recio and Rhode, 2000, Fig. 4). To remove delays associated with the acoustic system and middle-ear transmission, a delay of 0.2 ms, representing the approximate latency to the onset of umbo vibration (Recio and Rhode, 2000), was subtracted from the data. To help distinguish the different curves, linewidths have been made proportional to CF. The bottom graph [panel (b)] illustrates the unification across CF achieved by expressing the trajectories in panel (a) in dimensionless form as a function of time measured in periods of CF.

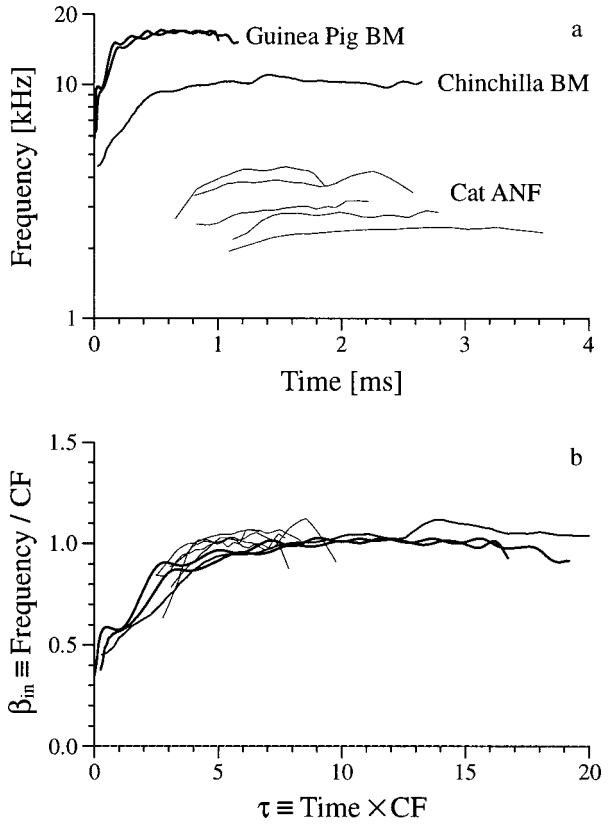


FIG. 3. Approximate invariance of glides across species. The figure shows IF trajectories obtained from ANF and BM impulse responses in three species (distinguished by linewidth). The top graph [panel (a)] shows IF trajectories $f_{in}(x,t)$ computed from (1) five cat ANF revcor functions with CFs spanning almost an octave (Carney and Yin, 1988; Carney *et al.*, 1999); (2) basilar-membrane (BM) responses to acoustic clicks in the chinchilla (Recio *et al.*, 1998, Fig. 6A, averaged across level); and (3) two derived BM impulse responses (input–output cross-correlation functions) measured with pseudorandom noise in the guinea pig (e.g., de Boer and Nuttall, 1997). A delay of 1 ms was subtracted from the cat ANF data to compensate for acoustic, synaptic, and neural propagation delays present in the revcor function (Carney, personal communication). A delay of 0.35 ms, representing the approximate latency to the onset of incus vibration (Recio *et al.*, 1998), was subtracted from the chinchilla data. Since the guinea-pig data were normalized by stapes vibration, no delay correction was necessary. The guinea-pig data are from experiments 9916 and 9924 at equivalent noise levels of 70 and 40 dB SPL, respectively. The bottom graph [panel (b)] illustrates the approximate unification achieved across species by re-expressing the trajectories in panel (a) in dimensionless form as a function of time measured in periods of CF.

(and the resulting constancy of the dimensionless glide slope) constitutes a time-domain manifestation of the approximate local scaling symmetry (Zweig, 1976; Siebert, 1968; Sondhi, 1978) manifest by basilar-membrane transfer functions (Rhode, 1971; Gummer *et al.*, 1987) and neural tuning curves (e.g., Kiang and Moxon, 1974; Liberman, 1978). Local scaling symmetry implies that rather than depending on position and frequency independently, mechanical transfer functions and tuning curves in fact depend on the two variables f and x primarily in the dimensionless combination $f/f_{CF}(x)$, where $f_{CF}(x)$ is the CF at location x (i.e., the cochlear position-frequency map). In the time domain, scaling implies that corresponding basilar-membrane and neural impulse responses, and thus their instantaneous-frequency trajectories, depend on t and x through the dimensionless combination $tf_{CF}(x)$ representing time measured in periods of the characteristic frequency. Figure 2(b) demonstrates this time-domain scaling in glides: To a good approximation, normalized IF trajectories are functions of the single dimensionless variable τ, where $\tau(x,t) \equiv tf_{CF}(x)$. We therefore define the normalized IF trajectory, $\beta_{in}(\tau)$, by the equation

$$\beta_{in}[\tau(x,t)] \equiv f_{in}(x,t)/f_{CF}(x). \quad (1)$$

Figure 2(b) emphasizes another important feature of glides in the scaling region: They are relatively *slow*, generally occurring over many cycles of the characteristic frequency. The period of the glide—arbitrarily defined, say, as the time needed for the instantaneous frequency to come to within 5% of the CF—is typically at least five cycles in duration. The time course of the glide thus represents a significant fraction of the energy-weighted average group delay of the response, even at low stimulus levels where impulse responses ring longer. The relative slowness of glides is also evident in the small values of the dimensionless glide slope obtained in Fig. 1(b).

Figure 3 generalizes the results of Fig. 2 across species. The top panel shows IF trajectories $f_{in}(x,t)$ measured in three different species (distinguished in the figure by linewidth). In addition to glides extracted from cat ANF revcor

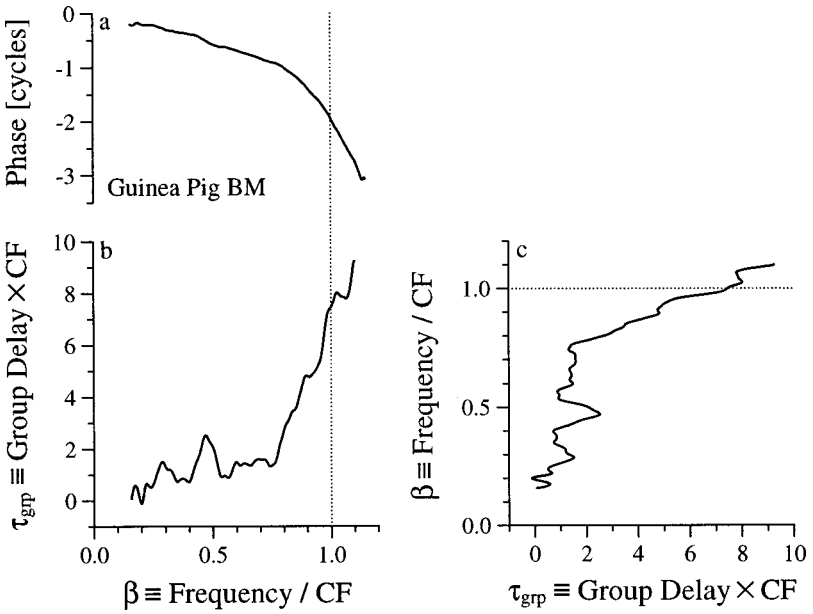


FIG. 4. Basilar-membrane phase and group delay. The graph on the upper left [panel (a)] shows the phase, $\phi(\beta)$, of the basilar-membrane transfer function measured in guinea pig (data from de Boer and Nuttall, experiment 9917 at 70 dB SPL). The graph on the lower left [panel (b)] shows the corresponding normalized group delay, $\tau_{grp}(\beta)$, computed from the slope of the phase. The graph on the right [panel (c)] shows the function obtained by reflecting the graph of $\tau_{grp}(\beta)$ in panel (b) about the diagonal line $y=x$, where y and x are the variables shown along the ordinate and abscissa, respectively. The resulting function, which we call $\beta_{grp}(\tau)$, represents the functional inverse of $\tau_{grp}(\beta)$. In all panels, dotted lines [vertical in panels (a) and (b), horizontal in (c)] mark the frequency $f=f_{CF}$.

functions (Carney and Yin, 1988; Carney *et al.*, 1999),[8] the top panel shows glides obtained from basilar-membrane (BM) velocity responses to acoustic clicks in the chinchilla (Recio *et al.*, 1998) and from derived BM impulse responses (input–output cross-correlation functions) measured with pseudorandom noise in the guinea pig (de Boer and Nuttall, 1997). As in Fig. 2, delays of extra-cochlear origin have been subtracted when necessary. Together, the trajectories span more than a three-octave range of CFs. Although the trajectories are somewhat noisy and uncertain, plotting them in dimensionless form (bottom panel) appears to unify the data across CF and, to a considerable extent, across species as well.[9] More data, however, are needed to make a definitive comparison.

III. FREQUENCY-DOMAIN MANIFESTATION OF GLIDES

Glides are a feature of the time-domain response of the basilar-membrane to acoustic clicks. We seek now to find their frequency-domain manifestation with the goal of illuminating the origin and interpretation of glides. Here, and in the remainder of the paper, we focus on glides as measured in the basal scaling region (i.e., at CFs greater than roughly 1.5 kHz).

Let $T(x,f)$ denote the basilar-membrane transfer function (i.e., basilar-membrane velocity divided by stapes velocity) measured at position x and frequency f. As discussed above, local scaling symmetry implies that $T(x,f) \approx T[f/f_{CF}(x)]$. Figure 4(a) shows the phase, ϕ, of T as measured in the guinea pig (de Boer and Nuttall, 1997, 2000a, b). The curvature of the phase reflects the well-known dispersive properties of cochlear wave propagation. Traveling-wave dispersion results from the spatial variation in the the geometry and mechanics of the cochlea that underlies the frequency-position map; as a consequence of these spatial gradients, the wavelengths and velocities of cochlear traveling waves vary with frequency and position [again, predominantly through the dimensionless ratio $f/f_{CF}(x)$]. The phase curvature indicates that wave components of different frequencies travel with different velocities and reach a fixed observation point (e.g., the measurement location) at different times after launching from the stapes. These delays can be quantified by computing the ''group delay'' from the slope of the phase versus frequency function (e.g., Papoulis, 1962):

$$t_{grp}(x,f) \equiv -\frac{1}{2\pi}\frac{\partial\phi(x,f)}{\partial f}. \tag{2}$$

Since ϕ scales (i.e., since ϕ is a function of f/f_{CF}), the group delay t_{grp} is also naturally expressed in periods of CF, and we therefore define

$$\tau_{grp}(\beta) \equiv f_{CF}(x)t_{grp}(x,f) = -\frac{1}{2\pi}\frac{d\phi(\beta)}{d\beta}, \tag{3}$$

where $\beta(x,f) \equiv f/f_{CF}(x)$. Figure 4(b) shows $\tau_{grp}(\beta)$ computed from the guinea-pig phase data in Fig. 4(a). At location x in the cochlea, the delay $\tau_{grp}[f/f_{CF}(x)]$ gives the approximate latency of waves of frequency f, measured in periods of the local CF. For example, the figure indicates that a tone burst at the characteristic frequency ($f=f_{CF}$, vertical dotted line) applied at the stapes takes roughly 8 stimulus periods to reach its characteristic place.

A. Glides and the inverse group delay

The dispersive character of cochlear traveling waves suggests a simple interpretation of glides. To illustrate, imagine that we measure BM motion at location x after a wideband stimulus has been applied to the stapes at time $\tau=0$. The motion we measure reflects the mechanical response of the cochlear partition to the local driving force. Because of the dispersion, different frequency components in the stimulus travel at different velocities and reach the fixed observation point at different times after application of the stimulus.

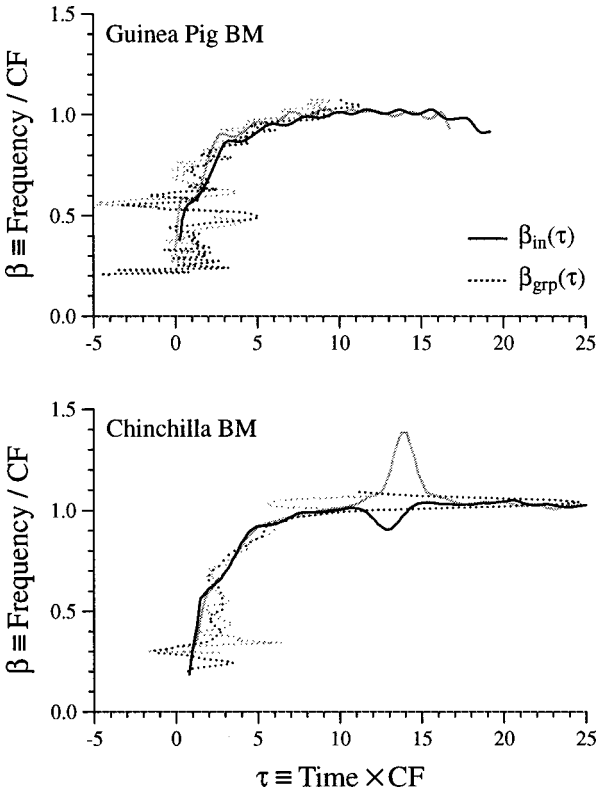


FIG. 5. Comparison between the glide and the inverse group delay. The solid lines represent normalized IF trajectories $\beta_{in}(\tau)$ computed from measured basilar-membrane impulse responses in the guinea pig (top panel) and chinchilla (bottom panel). The dotted lines represent inverse group delays $\beta_{grp}(\tau)$ computed from the phase of the corresponding transfer functions (i.e., from the Fourier transforms of the impulse response). We limit consideration to frequencies near CF and below; values of $\beta_{grp}(\tau)$ greater than 1.05, corresponding to frequencies $f>1.05f_{CF}$, are not shown. As in Fig. 3, the guinea-pig data are from de Boer and Nuttall (experiments 9916 and 9924, indicated by the gray and black lines, respectively). The chinchilla data are from Figs. 3 and 6 of Recio *et al.* (1998, animal L113 with sound levels of 64 and 84 dB SPL indicated by the gray and black lines, respectively). Note that the intensity dependent ''bump'' at $\tau\approx 14$ in the chinchilla IF trajectory results from nonmonotonic decay of the envelope of the impulse response (''waxing and waning'') due to interference between components with different phase (perhaps arising from internal reflection within the cochlea).

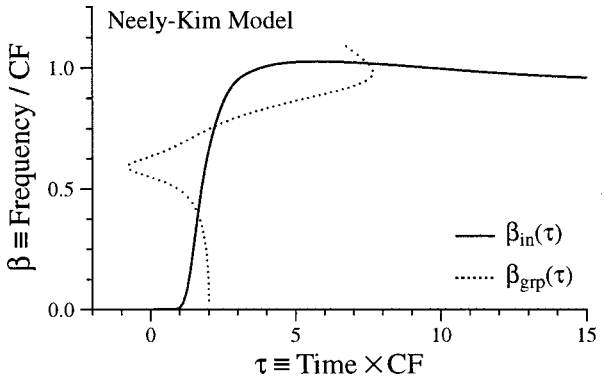


FIG. 6. Comparison between the glide and the inverse group delay for the model of Neely and Kim (1986). The solid line represents the normalized IF trajectory $\beta_{in}(\tau)$ computed from the model basilar-membrane impulse response. The dotted line represents the inverse group delay $\beta_{grp}(\tau)$ computed from the phase of the corresponding transfer function. This model provides an example of a system in which Eq. (4) is not well satisfied.

According to the interpretation of the group delay discussed above, the function $\tau_{grp}(f/f_{CF})$ specifies the approximate latency of frequency component f. In other words, at time $\tau_{grp}(f/f_{CF})$ we expect our observation point to be driven predominantly by components in the stimulus with average frequency f. How does this instantaneous driving frequency vary over time? To answer this question, note that the function $\tau_{grp}(f/f_{CF})$ yields a time (namely, the normalized group delay, τ_{grp}) as a function of normalized frequency, $f/f_{CF}=\beta$; by functionally inverting $\tau_{grp}(\beta)$ we therefore obtain the normalized frequency as a function of time. We call this functional inverse the ''inverse group delay'' and denote it $\beta_{grp}(\tau)$. The functional inverse is defined so that $\beta_{grp}[\tau_{grp}(\beta)]=\beta$; it can be obtained from the graph by reflecting the plot of $\tau_{grp}(\beta)$ about the diagonal line $y=x$, where y and x are the variables shown along the ordinate and abscissa, respectively. The function $\beta_{grp}(\tau)$ specifies the frequency component of the stimulus that we expect to be driving the measurement point at the time τ. At time τ, frequency components in the stimulus less than $\beta_{grp}(\tau)$ have already ''passed by'' en route to the apex, and frequencies greater than $\beta_{grp}(\tau)$ have yet to arrive. So, if glides reflect the time variation of the frequency of the effective driving force at the measurement location, we would expect $\beta_{in}(\tau)\approx\beta_{grp}(\tau)$.

We test this prediction in Fig. 4(c), which shows the inverse group delay obtained by inverting the function $\tau_{grp}(f/f_{CF})$ from Fig. 4(b). In its general form, the function $\beta_{grp}(\tau)$ certainly resembles the normalized IF trajectories, $\beta_{in}(\tau)$, plotted in Fig. 3(b). Figure 5 shows an explicit comparison between $\beta_{in}(\tau)$ and $\beta_{grp}(\tau)$ in guinea pig and chinchilla. No shifts or adjustments of the data along either axis have been performed. Although both functions are somewhat uncertain at the smallest values of τ,[10] the similarity between the two functions is striking. The agreement indicates that

$$\beta_{in}(\tau)\approx\beta_{grp}(\tau) \quad (\beta_{grp}\lesssim 1), \tag{4}$$

as suggested by the conceptual argument outlined in the previous paragraph. The restriction to $\beta_{grp}\lesssim 1$ amounts to limiting consideration to frequencies less than or equal to CF.

We note that the empirical agreement between $\beta_{in}(\tau)$ and $\beta_{grp}(\tau)$ shown here is not the consequence of a mathematical identity that holds in all systems. Based on the logic of our conceptual argument, however, we conjecture that relation (4) holds in cochlear models whose phase response and group delay vary smoothly and monotonically with frequency. If this conjecture is correct, the cochlear model of Neely and Kim (1986), for which approximate equality (4) is not well satisfied, provides an exception that proves the rule. In the Neely–Kim model, the phase of the BM transfer function has an anomalous ''bump'' roughly a half-octave below CF—a feature introduced as a consequence of matching a similar phase bump (the so-called ''π phase shift'') evident in the neural population study of Kim *et al.* (1979). As illustrated in Fig. 6, this bump in the model BM phase produces considerable structure in $\beta_{grp}(\tau)$ with no counterpart $\beta_{in}(\tau)$. Also evident in Fig. 6 is the unusual form of $\beta_{in}(\tau)$ noted by de Boer and Nuttall (1997): Unlike empirical glides at CFs above 1 kHz, the IF trajectory in this model rises quickly and

then decreases towards CF. Providing a more precise definition of the class of systems that satisfies Eq. (4) remains an interesting open question.

IV. A SIMPLE MODEL FOR GLIDES

Equation (4) indicates that glides are closely related to cochlear traveling-wave dispersion and may therefore originate in the time dependence of the effective driving force at the measurement location. We now explore this possibility in the context of a simple model.

In classical point-impedance models of the cochlea (nicely reviewed in, e.g., de Boer, 1980, 1984, 1991; Hubbard and Mountain, 1996), the velocity $V_{\mathrm{BM}}(x,f)$ of the basilar-membrane at position x due to sinusoidal stimulation at frequency f can be written as the product of two factors:

$$V_{\mathrm{BM}}(x,f)=Y_{\mathrm{BM}}(x,f)P(x,f). \tag{5}$$

The first term, $Y_{\mathrm{BM}}(x,f)$, represents the admittance of the cochlear partition, and the second term, $P(x,f)$, represents the driving pressure difference acting across its surfaces. In the time domain, Eq. (5) becomes a convolution:

$$v_{\mathrm{BM}}(x,t)=y_{\mathrm{BM}}(x,t)\times p(x,t), \tag{6}$$

where lower- and upper-case quantities (e.g., v_{BM} and V_{BM}) are related by Fourier transformation. Note that the admittance $y_{\mathrm{BM}}(x,t)$ and driving pressure $p(x,t)$ differ profoundly in character. The time-domain admittance $y_{\mathrm{BM}}(x,t)$ characterizes the response of an isolated section of the cochlear partition to an impulsive force and depends only on the local properties of the partition at position x. The pressure term, $p(x,t)$, by contrast, represents the driving force applied *in situ* and is therefore *global*; since stimuli are usually applied in the ear canal (or, effectively, at the stapes when responses are normalized by stapes motion) and must propagate to the measurement location, $p(x,t)$ depends not only on the form of the stimulus, but also on the mechanics of the entire cochlea, including boundary conditions at the stapes and helicotrema. Although $v_{\mathrm{BM}}(x,t)$ is often referred to as the ''impulse response of the basilar-membrane at location x''—suggesting that it depends solely on local mechanics—the function $v_{\mathrm{BM}}(x,t)$ really represents the ''impulse response of the entire cochlea as measured at location x.''[11]

A. Origin of glides in traveling-wave dispersion

In the two limiting cases, features of the velocity impulse response $v_{\mathrm{BM}}(x,t)$ can arise either because they appear in the local admittance, $y_{\mathrm{BM}}(x,t)$, or as a global effect manifest through the driving pressure, $p(x,t)$. What do simple models predict about the origin of glides? We explore this question with the model obtained by solution to the inverse problem using data from the squirrel monkey cochlea (Zweig, 1991). We adopt this model both for its simplicity and for its ability to reproduce many of the qualitative features of cochlear measurements. In the model, the BM admittance scales and has the form of an harmonic oscillator, with a net *negative* damping, stabilized by a feedback force proportional to the BM displacement at an earlier time.[12] Model parameter values have been modified somewhat from those obtained by Zweig (1991) in order to produce a BM velocity impulse response that peaks after about ten periods of the characteristic frequency (in rough agreement with data at low sound-pressure levels from guinea pig and chinchilla).[13]

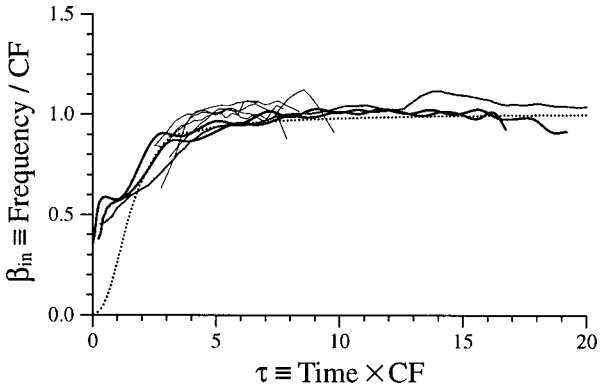


FIG. 7. Model glide compared to glides in three species. The figure shows the model IF trajectory for the BM velocity response (dotted line) together with those obtained from BM velocity impulse responses (guinea pig and chinchilla) and ANF revcor functions (cat). Trajectories for the three species, reproduced from Fig. 3(b), are distinguished by linewidth: guinea pig (thick), chinchilla (medium), and cat (thin).

Figure 7 shows the model IF trajectory[14] for the BM velocity response together with the three-species comparison shown earlier in Fig. 3. Although differences between the model trajectory and the mechanical glides from chinchilla and guinea pig are evident at small values of τ—perhaps suggesting additional structure at frequencies below CF not included in the model admittance—the model clearly captures the qualitative form of the glide.

Figure 8 probes the origin of glides by plotting $v_{\mathrm{BM}}(\tau)$, $y_{\mathrm{BM}}(\tau)$, and $p(\tau)$, corresponding to the three variables in Eq. (6). Note the admittance $y_{\mathrm{BM}}(\tau)$ represents the velocity response to a pressure impulse applied locally; it therefore jumps to a nonzero value at $\tau=0$.[15] The panels on the right show the corresponding IF trajectory for each variable (solid lines). For comparison, the dotted lines give the inverse group delays computed from the corresponding frequency-domain transfer functions; the agreement indicates that Eq. (4) deduced from the data holds also in the model. Note that aside from an onset glitch due to the method of calculation, the IF trajectory of the admittance is essentially constant in this model. Thus, the figure demonstrates how the glide apparent in BM velocity responses, $v_{\mathrm{BM}}(\tau)$, originates in this model not through the local admittance, $y_{\mathrm{BM}}(\tau)$, but through the time dependence of the driving pressure force, $p(\tau)$.

V. SUMMARY AND DISCUSSION

Frequency modulations (or glides), reported in impulse responses of both the auditory nerve (e.g., Møller and Nilsson, 1979; Carney *et al.*, 1999; Lin and Guinan, 2000) and the basilar-membrane (e.g., de Boer and Nuttall, 1997; Recio *et al.*, 1998; Recio and Rhode, 2000), represent a change over time in the instantaneous frequency of oscillation of the response waveform. In this paper, we have combined theory with experimental data to explore the basic phenomenology of glides.

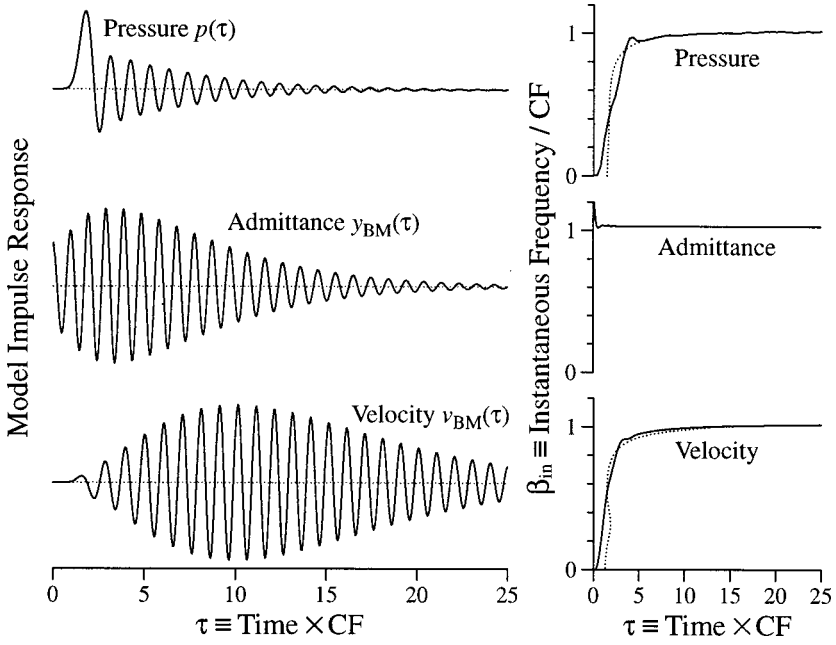


FIG. 8. Model impulse responses and glides. From top to bottom, the three panels on the left show responses $p(\tau)$, $y_{BM}(\tau)$, and $v_{BM}(\tau)$, respectively, corresponding to the three variables (pressure, admittance, velocity) in Eq. (6). For the pressure and velocity responses, the impulses were applied at the stapes; for the admittance, the impulse was applied locally at the measurement location. The panels on the right show normalized IF trajectories computed from the corresponding responses on the left (solid lines). For comparison, the dotted lines (pressure and velocity panels only) give the inverse group delays $\beta_{grp}(\tau)$ computed from the phase of the corresponding frequency-domain transfer functions. The glitch evident at $\tau \lesssim \frac{1}{2}$ in the IF trajectory for $y_{BM}(\tau)$ results from an artifact of the analytic-signal method used to compute the instantaneous frequency.

A. Scaling behavior

In Sec. II we demonstrated the scaling behavior of glides: When expressed in natural dimensionless form, glides have a nearly universal form independent of cochlear location for CFs above approximately 1.5 kHz (see Figs. 1 and 2). Outside the scaling region (i.e., for CF$\lesssim$1.5 kHz), glides appear to depend strongly on CF. The normalized IF trajectory, $\beta_{in}(\tau)$, is obtained by measuring time in units of the CF period at the measurement location. Invariance of glides across CF is a time-domain manifestation of the approximate local scaling symmetry manifest by basilar-membrane transfer functions and neural tuning curves.

Basal glides also appear to have a similar form in cats, guinea pigs, and chinchillas [see Fig. 3(b)], suggesting an approximate invariance across species as well. Evidence from otoacoustic emissions suggests that human glides may prove an exception to the rule. Estimates of human basilar-membrane group delay obtained from measurements of human stimulus-frequency otoacoustic emissions (Shera and Guinan, 2000a,b) indicate that human group delays are roughly a factor of 2–3 larger than group delays obtained at comparable frequencies and levels in cat and guinea pig. These differences suggest that human IF trajectories may differ from those shown in Fig. 3(b), if only by translation along the τ axis. We note, however, that species differences that affect primarily the peak of the transfer function may not be obvious in the form of the IF trajectory, $\beta_{in}(\tau)$. According to approximate equality (4), changes in the transfer-function phase near CF, where the group delay τ_{grp} is large, correspond to values of τ where the slope of the trajectory $\beta_{in}(\tau)$ is small, making differences more difficult to discern.

Although the trend line characterizing the glide-slope data in Fig. 1(b) suggests the existence of an approximate scaling region, the scatterplot also reveals considerable variation in glide slope among fibers with comparable CFs (e.g., at CFs near 2 kHz). Carney *et al.* (1999) attribute some of this variability to pooling data from different cats. Note, however, that the approximate invariance of the basal glide—both across CF and across species—suggests that much of the apparent variability in glide slope may more reflect uncertainty in the estimate of the slope than actual glide variations among fibers. For example, Figs. 2 and 3 indicate that the curvature of the IF trajectory varies with time; the slopes of tangents to the trajectory therefore depend on the precise time interval analyzed, which may differ somewhat from fiber to fiber. In addition, uncertainty in the computed IF trajectory (e.g., due to noise in the revcor function or artifacts of the analytic-signal method) can confound precise determination of the slope. Finally, since uncertainties in the data points are ignored in the standard linear-regression analysis used to estimate the glide slope, the values of best-fit parameters can be skewed considerably by outliers with large errors (e.g., Press *et al.*, 1992).

B. Relation to traveling-wave dispersion

Glides are a feature of the time-domain impulse response. In Sec. III we identified the frequency-domain analog of glides, demonstrating that the normalized IF trajectory, $\beta_{in}(\tau)$, is approximately equal to the inverse group delay of the basilar-membrane transfer function (see Figs. 4 and 5). The inverse group delay, $\beta_{grp}(\tau)$, specifies the frequency component of the stimulus that we expect to be driving the measurement point at the time τ. The empirical relation $\beta_{in}(\tau) \approx \beta_{grp}(\tau)$ thus indicates that glides are closely related to cochlear traveling-wave dispersion and suggests that they reflect the time dependence of the effective driving force at the measurement location. Measurements of intracochlear pressure support this conclusion by providing direct experimental evidence for the dispersive character of the traveling pressure wave (Olson, 1998, 1999).

Traveling-wave dispersion, a consequence of the spatial variations in cochlear geometry and mechanics that underlie the frequency-position map, means that the wavelengths and velocities of cochlear traveling waves vary with frequency

and position. As a consequence of local scaling, spatial and frequency dispersion are intimately related. Scaling relates properties of the mechanical transfer function to those of the traveling wave. In particular, mechanical transfer functions $T[f/f_{CF}(x)]$ measured as a function of f at fixed x also describe the traveling wave as a function of x at fixed f. At fixed position, T is the transfer function; at fixed frequency, the traveling wave. Thus, as a consequence of scaling, phenomena in one domain (e.g., f) have counterparts in the other (e.g., x). In the frequency domain, glides reflect frequency dispersion, specifically the increase in transfer-function group delay as the stimulus frequency approaches CF. Reinterpreted in the spatial domain, glides correspond to the spatial dispersion evident in the decrease in wavelength of the traveling wave as it approaches its characteristic place. In their early work on glides, Møller and Nilsson (1979) sought, but failed to locate, the frequency-domain analog of glides in the amplitude of the transfer function. Our findings here demonstrate that the frequency modulation of the impulse response has its most salient analog not in the amplitude, but in the phase of T.[16]

C. A simple model

To further explore the origin of glides we presented in Sec. IV a simple modeling framework suggested by classical point-impedance models of the cochlea. The framework is based on the decomposition of the velocity response into the convolution of two factors: the impulse response, $y_{BM}(x,t)$, of the local section of the cochlear partition and the driving pressure force, $p(x,t)$, across its surface [see Eqs. (5) and (6)]. One should not infer from this decomposition that $p(x,t)$ and $y_{BM}(x,t)$ are independent.[17] In particular, the pressure $p(x,t)$ depends intimately on the admittance $y_{BM}(x,t)$ and its variation along the length of the cochlea. (In model calculations, for example, one specifies the spatial variation of the admittance and other aspects of cochlear geometry and then solves for the pressure and velocity.) In the two limiting cases, however, glides can arise either because they appear in the local admittance (e.g., as a consequence of local micromechanical modes of vibration with different resonant frequencies) or as a global effect, manifest through the driving pressure, of the spatial variation of mechanics throughout the cochlea.

We explored the origin of glides in a simple one-dimensional transmission-line model based on the solution to the inverse problem in squirrel monkey (Zweig, 1991). Although the one-dimensional model provides an oversimplified representation of the full three-dimensional motion of the cochlear fluids, our qualitative conclusions are expected to apply in more realistic geometries. For example, one-dimensional models capture, qualitatively and at least semi-quantitatively, the essential physics that gives rise to traveling-wave dispersion. In addition, the impedances of the cochlear partition obtained as solutions to the inverse problem in long-wave, short-wave, and fully three-dimensional models are all in remarkable qualitative agreement (e.g., Zweig, 1991; de Boer, 1995a, b; de Boer and Nuttall, 1999).

We used the model results to argue that, in the scaling region above roughly 1.5 kHz, glides arise primarily through the time dependence of the driving pressure (see Fig. 8) rather than via the local admittance of the cochlear partition. Indeed, in this model the instantaneous frequency of the admittance $y_{BM}(\tau)$ is essentially constant. Remarkably, inverse analysis of experimental data that manifest glides [i.e., the empirical squirrel-monkey transfer function analyzed by Zweig (1991)] evidently yields a model admittance without a glide. Within its domain of validity (e.g., at low sound levels), the inverse analysis thus indicates that glides arise entirely through the dispersive character of wave propagation in the cochlea, which in turn reflects the spatial variation in cochlear geometry and mechanics.[18] The inverse analysis that produced the model depends, of course, on many simplifying assumptions (Zweig, 1991). Our results do not, therefore, rule out possible contributions to glides from local micromechanical resonances, especially at higher intensities. Our results do indicate, however, that such resonances are not necessary to explain the basic phenomenology.

Furthermore, the framework implies that to deduce properties of the local admittance of the organ of Corti (e.g., the resonant frequencies of its modes of vibration) from measurements of BM or ANF impulse responses, one needs to deconvolve (''divide out'') the frequency dependence of the driving pressure. Since our results show that the driving pressure can manifest considerable complexity, the necessary deconvolution is nontrivial. Indeed, it is precisely this deconvolution problem that inverse analysis (Zweig, 1991; de Boer, 1995a, b; de Boer and Nuttall, 1999) solves in certain circumstances (e.g., when backwards-traveling waves can be neglected). Along these lines, we note that internal reflection of cochlear traveling waves [arising, for example from wave scattering off distributed micromechanical impedance perturbations (Shera and Zweig, 1993; Zweig and Shera, 1995) and/or reflection from the helicotrema] can make significant contributions to the driving pressure. Although we defer detailed discussion of the topic to a future paper, we note that the mechanisms responsible for otoacoustic emissions can therefore produce important effects on basilar-membrane velocity responses.

Our suggestion that glides arise predominantly from the time dependence of the driving pressure contrasts with existing models for glides, which ignore cochlear hydrodynamics and treat the cochlea essentially as a bank of bandpass filters. As a consequence, such models locate the origin of glides entirely in local resonances (e.g., their differential build-up and decay) or other micromechanical contributions to the impedance of the partition (e.g., de Boer and Nuttall, 1995; Irino and Patterson, 2000; Lin and Guinan, 2000; Mountain, 2000).[19] By contrast, our modeling framework highlights the fact that the velocity impulse response reflects not only the local mechanics at the point of measurement but depends, in fact, on the mechanics of the entire cochlea. The model thus underscores the fundamental difference between the biophysical basis of tuning in the mammalian cochlea—where tuning arises globally and involves the collective action of the outer hair cells in providing coherent amplification of traveling waves (e.g., Neely, 1983; Zweig, 1991; de Boer, 1995a; Olson, 1999)—and its origin in lower vertebrates, where tuning mechanisms local to the sensory hair cell—

such as electrical and mechanical resonances (e.g., Crawford and Fettiplace, 1981; Lewis, 1985; Art *et al.*, 1986; Hudspeth and Lewis, 1988; Weiss and Leong, 1985; Freeman and Weiss, 1990) together with tuned amplification of hair-bundle motion (e.g., Crawford and Fettiplace, 1985; Martin and Hudspeth, 1999)—are believed paramount.

D. Intensity independence

Glides appear nearly independent of stimulus intensity over most of the dynamic range of hearing (e.g., de Boer and Nuttall, 1997; Recio *et al.*, 1998; Carney *et al.*, 1999) and even maintain their general form postmortem (e.g., Recio *et al.*, 1998). The intensity independence of glides follows from a striking symmetry of mechanical and neural impulse responses: near invariance with intensity of the phase of the oscillations in the response waveform. This invariance is manifest, for example, in the approximate level independence of the latency of the peaks of PST histograms obtained from the auditory nerve in response to clicks (e.g., Kiang *et al.*, 1965; Goblick and Pfeiffer, 1969; Lin and Guinan, 2000) or in the near invariance of the zero crossings of the mechanical waveform (e.g., Robles *et al.*, 1976; Ruggero *et al.*, 1992; de Boer and Nuttall, 1997, 2000b; Recio and Rhode, 2000).[20] Although the two are often discussed together, glides acquire their intensity independence through mechanisms distinct from those responsible for the underlying frequency modulation. In a separate paper (Shera, 2001), we apply the concepts and modeling framework developed here to explore both the origins of near invariance of fine time structure and the implications of this symmetry for cochlear mechanics.

E. Apical glides

The division suggested by the data in Fig. 1(b) into basal scaling and apical nonscaling regions adds to the now considerable body of evidence for fundamental differences in the mechanics of the base and apex of the cochlea (e.g., Liberman, 1982; Kiang, 1984; Rhode and Cooper, 1996; Carney *et al.*, 1999). Despite clear differences in the mechanics, relations such as Eq. (4) between the normalized IF trajectory and the inverse group delay may also apply in the apex. For example, the generally downwards path of the IF trajectory obtained from cat revcor functions (Carney *et al.*, 1999) is consistent with neural phase responses measured in the same species (e.g., Kiang, 1984, Fig. 12), which show large group delays at frequencies at and below CF and smaller group delays at frequencies above CF. A similar reversal of the basal pattern is seen in mechanical responses in the apex of the chinchilla cochlea (Rhode and Cooper, 1996). Functional inversion of these apical group delay curves would yield downward glides that approached CF from above.

Since the general modeling framework presented in Sec. IV presumably also applies in the apex, the framework may prove helpful in understanding the origin of inverted glides as well. At a minimum, Eqs. (5) and (6)—combined with the likelihood of anomalous (by basal standards) behavior in the pressure due to end effects caused by proximity of the helicotrema and the narrowing of the cochlear scalae (e.g., Puria and Allen, 1991)—suggest caution in interpreting the unusual features of apical responses as necessarily involving complexity in the local mechanics of the cochlear partition [e.g., in $Y_{\mathrm{BM}}(x,f)$]. Since the discovery of OHC motility and the recognition of the importance of cochlear micromechanics, it has become natural to regard complex or interesting phenomena as invariably indicative of complexity in local micromechanics. Although such complexities surely exist, our results illustrate that far simpler models can often account for much of the data. Indeed, that a capacity for rich and nontrivial signal analysis can emerge as the global consequence of the local interactions of a large number of relatively simple components is one of the great benefits of the traveling-wave architecture (e.g., Zweig, 1991; Sarpeshkar, 2000).

ACKNOWLEDGMENTS

We gratefully acknowledge stimulating conversations with Jont Allen, Egbert de Boer, Laurel Carney, Nigel Cooper, Paul Fahey, John Guinan, Stephen Neely, and George Zweig, and thank Egbert de Boer, Laurel Carney, Alfred Nuttall, Alberto Recio, Bill Rhode, and Mario Ruggero for generously sharing their data for analysis. Finally, we thank William Peake and the two reviewers for helpful comments on the manuscript. This work was supported by Grant No. R01 DC03687 from the NIDCD, National Institutes of Health.

[1]In this paper, the term ''best frequency'' (BF) is used to locate the maximum of the BM frequency response, which may vary with intensity. The ''characteristic frequency'' (CF) is defined as the best frequency measured in the low-level linear limit at sound levels near threshold. By definition, the CF is therefore independent of intensity.

[2]Carney *et al.* (1999) computed instantaneous-frequency trajectories, $f_{\mathrm{in}}(x,t)$, using the standard analytic-signal technique described in footnote 14.

[3]In their report, Carney *et al.* (1999) plot the glide slope versus a mean best frequency computed by averaging over the stimulus intensities studied. Although it may introduce a small bias in the computation—by slightly underestimating the CF in the base and overestimating the CF in the apex (Carney *et al.*, 1999)—we use this mean value as the best available estimate of the fiber CF.

[4]We note that different measures may be more or less sensitive to violations of scaling. For example, the rough estimate of 1.5 kHz for the boundary of the glide-scaling region in the cat differs from the approximate scaling boundary (CFs$\gtrsim$3 kHz) based on the shapes of neural tuning curves (e.g., Kiang and Moxon, 1974; Liberman, 1978). Measurements of stimulus-frequency otoacoustic emission group delay, also in cat, provide evidence for deviations from scaling at frequencies as high as 20 kHz (Shera and Guinan, 2000a, b).

[5]Puria and Allen (1998) argue that the great bulk of the delay associated with middle-ear transmission occurs in the eardrum. Subtracting out umbo delays should therefore remove most of the delay due to the middle ear.

[6]Ideally, BM responses should be normalized by middle-ear motion measured in the same animal (e.g., via deconvolution). Since the necessary middle-ear measurements are not available, we subtract out an average delay. Fortunately, Recio and Rhode's (2000) measurements of umbo vibration suggest that the delay may vary by only a few percent between animals.

[7]Recio and Rhode [2000; Fig. 4(c)] plot their data without correction for middle-ear delay and infer that the relative amount of frequency modulation varies systematically with location in the cochlea. Figure 2(b) indicates, however, that most of the systematic variation evident in their figure arises not from the cochlea but from acoustic and middle-ear delay.

[8]Trajectories were computed from the original revcor data obtained as described in Appendix A of Carney *et al.* (1999). The trajectories shown are from fibers c86100u26-30, c86100u18-1, c86204u28-1, c86204u26-1, and c87154u43-5 at levels of roughly 60 dB. Fiber CFs range from 2.3 to 4.2 kHz.

[9]Some of the scatter may reflect uncertainty or bias in the estimates of CF. For example, Carney *et al.* (1999) generally estimate CF by computing the best frequency (BF) of the revcor at the lowest intensity measured. Since the BF decreases with increasing intensity in the basal half of the cochlea, this value will underestimate the actual CF in some fibers. Underestimating the CF will tend to bias graphs of the normalized trajectory $\beta_{\text{in}}(\tau)$ slightly upwards (along the ordinate) and to the left (along the abscissa).

[10]At the earliest values of τ the IF trajectory, $\beta_{\text{in}}(\tau)$, can be uncertain because (1) the signal energy is small and may be partially obscured by noise; (2) algorithms for computing the IF are subject to end effects; and (3) the IF itself is poorly defined over small time intervals (the uncertainty principle). Because it is obtained by differentiating the BM phase, an operation which amplifies high-frequency measurement noise, the inverse group delay, $\beta_{\text{grp}}(\tau)$, may also be poorly determined in this region. This problem may be most serious at small τ, where the data are obtained from the ''tails'' of the BM velocity transfer functions below CF and signal-to-noise ratios can be poor, especially in responses to clicks.

[11]The global nature of the pressure $P(x,f)$ and its dependence on the admittance $Y(x,f)$ are evident, for example, in the Wentzel–Kramers–Brillouin (WKB) solution for the one-dimensional transmission-line model. If only forward-traveling waves are present,

$$P(x,f)\approx P_0(f)\left[\frac{Z_c(x,f)}{Z_c(0,f)}\right]^{1/2}\exp\left[-i\int_0^x k(x',f)dx'\right],$$

where $Z_c(x,f)$ is the characteristic impedance of the transmission line and $k(x,f)$ is the local wavenumber. The characteristic impedance and wave number are defined in terms of the series impedance $Z(x,f)$ and shunt admittance $Y(x,f)$ by the equations $Z_c\equiv\sqrt{Z/Y}$ and $k\equiv\sqrt{-ZY}$. Note that as a consequence of the integral in the exponential, $P(x,f)$ depends on values of Z and Y at all points $x'\leqslant x$. Furthermore, the existence of only forward-traveling waves depends, in part, on boundary conditions applied at $x'>x$.

[12]The model admittance has the form (Zweig, 1991)

$$Y_{\text{BM}}(\beta_r)\propto\frac{i\beta_r}{1-\beta_r^2+i\delta\beta_r+\rho e^{-2\pi i\mu\beta_r}}.$$

The variable $\beta_r=f/f_r(x)$, where $f_r(x)$ is the undamped resonant frequency of the oscillator (i.e., the resonant frequency in the limit when the damping δ and feedback strength ρ are both negligible). The parameter μ represents the time delay of the stabilizing feedback force, expressed in periods of the resonant frequency $f_r(x)$. Note that β_r differs from β, which here represents the model-independent ratio $f/f_{\text{CF}}(x)$, where $f_{\text{CF}}(x)$ is the characteristic frequency defined by the peak of the transfer function (see footnote 1). The parameter values given in footnote 13 imply that $f_{\text{CF}}(x)$ and $f_r(x)$ are everywhere proportional, with $f_{\text{CF}}/f_r\approx 0.97$.

[13]Model parameter values were determined as detailed in Zweig and Shera (1995, footnote 9) by requiring that the two closely spaced poles that produce the peak in the BM admittance coincide at a distance $\text{Im}\{\zeta_r\}=0.04$ from the real axis. (The variable ζ_r represents the extension of the real variable β_r to the complex plane.) The resulting parameter values are $\{\delta,\rho,\mu\}=\{-0.1024,0.1175,1.7450\}$. The parameter N, which determines the approximate number of wavelengths of the traveling wave on the basilar-membrane in response to sinusoidal stimulation, was given the value $N=2.5$.

[14]We computed instantaneous-frequency (IF) trajectories using the standard analytic-signal technique (e.g., Cohen, 1995; de Boer and Nuttall, 1997; Recio *et al.*, 1998; Carney *et al.*, 1999; Recio and Rhode, 2000). For a real signal $s(t)$, the analytic signal $a(t)$ is defined as the complex function

$$a(t)=s(t)+i\mathcal{H}\{s(t)\},$$

where $\mathcal{H}\{\cdot\}$ is the Hilbert transform. The instantaneous frequency $f_{\text{in}}(t)$ is then given by

$$f_{\text{in}}(t)=\frac{1}{2\pi}\frac{d\angle a(t)}{dt},$$

where $\angle a(t)$ is the unwrapped phase of $a(t)$. As a check on our results, we also computed the instantaneous frequency directly from $s(t)$ using the intervals between adjacent minima, maxima, and zero crossings (both positive- and negative going) of the time waveform. For the model pressure response shown in Fig. 8, the analytic-signal technique gave spurious oscillations in both the envelope and instantaneous frequency of the waveform; we subsequently eliminated these oscillations by smoothing.

[15]The time-domain admittance $y_{\text{BM}}(x,t)$ represents the velocity $v_{\text{BM}}(x,t)$ produced by an applied pressure impulse of unit amplitude $p(x,t)=\delta(t)$. Its nonzero value at small positive times does not reflect any violation of causality. To see this, consider a simpler example: the motion of a mass, m, attached to a spring, k, and subject at time $t=0$ to an impulsive force, $F_0\delta(t)$. The oscillator displacement, $x(t)$, satisfies the equation of motion

$$\ddot{x}(t)=-\omega_0^2x(t)+v_0\delta(t),$$

where $\omega_0=\sqrt{k/m}$ and $v_0=F_0/m$. If the mass is initially at rest at the origin, integrating $\ddot{x}(t)$ over the interval $[-\infty,0^+]$ to obtain the velocity $\dot{x}(0^+)$ yields

$$\dot{x}(0^+)=\int_{-\infty}^{0^+}[-\omega_0^2x+v_0\delta(t)]dt=v_0.$$

The velocity response thus jumps discontinuously to a nonzero value at $t=0$.

[16]A striking consequence of the glide becomes apparent in the amplitude when measurements at different stimulus intensities are compared. As pointed out by Carney (1999), intensity-dependent shifts in the temporal envelope of BM and ANF click responses, when combined with the intensity-independent frequency glide, can produce changes in the best frequency of the response. We discuss this issue further in a separate paper (Shera, 2001).

[17]This point has been emphasized by de Boer (1997), whose analysis underscores the tight connection between cochlear wave propagation and tuning (and thus the artificiality of any division between ''wave travel'' and ''filter build-up'').

[18]Møller and Nilsson (1979) write that ''the reason for [the] frequency modulation is that the impulse response of the inhomogeneous transmission line contains Bessel functions.'' Our results thus provide a physical explanation for Møller and Nilsson's mathematical observation.

[19]Other computational models of glides, although also based on bandpass filters, are more phenomenological in spirit and make no inferences about cochlear biophysics (e.g., Irino and Patterson, 1997; Shekhter and Carney, 1997).

[20]Although an excellent first approximation at low and moderate sound-pressure levels, near intensity invariance of fine time structure breaks down in auditory-nerve responses at high sound levels. The data of Lin and Guinan (2000), for example, show clear evidence for phase reversals and other ''anomalies'' at click levels of 90 dB pSPL (peak equivalent SPL) and above. Although species and methodological issues complicate the comparison, click responses measured on the basilar-membrane show little evidence of comparable features, maintaining near invariance of their zero crossings even at levels exceeding 115 pSPL (e.g., Recio and Rhode, 2000).

Art, J. J., Crawford, A. C., and Fettiplace, R. (**1986**). ''Electrical resonance and membrane currents in turtle cochlear hair cells,'' Hear. Res. **22**, 31–36.

Carney, L. H. (**1999**). ''Temporal response properties of neurons in the auditory pathway,'' Curr. Opin. Neurobiol. **9**, 442–446.

Carney, L. H., and Yin, T. C. T. (**1988**). ''Temporal coding of resonances by low-frequency auditory nerve fibers: Single fiber responses and a population model,'' J. Neurophysiol. **60**, 1653–1677.

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.

Cleveland, W. S. (**1993**). *Visualizing Data* (Hobart, Summit, NJ).

Cohen, L. (**1995**). *Time-Frequency Analysis* (Prentice-Hall PTR, Upper Saddle River, NJ).

Crawford, A. C., and Fettiplace, R. (**1981**). ''An electrical tuning mechanism in turtle cochlear hair cells,'' J. Physiol. (London) **312**, 377–412.

Crawford, A. C., and Fettiplace, R. (**1985**). ''The mechanical properties of ciliary bundles of turtle cochlear hair cells,'' J. Physiol. (London) **364**, 359–379.

de Boer, E. (**1980**). ''Auditory physics. Physical principles in hearing theory. I,'' Phys. Rep. **62**, 88–174.
de Boer, E. (**1984**). ''Auditory physics. Physical principles in hearing theory. II,'' Phys. Rep. **105**, 142–226.
de Boer, E. (**1991**). ''Auditory physics. Physical principles in hearing theory. III,'' Phys. Rep. **203**, 125–231.
de Boer, E. (**1995a**). ''The 'inverse problem' solved for a three-dimensional model of the cochlea. I. Analysis,'' J. Acoust. Soc. Am. **98**, 896–903.
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. (**1997**). ''Cochlear models and minimum phase,'' J. Acoust. Soc. Am. **102**, 3810–3813.
de Boer, E., and Nuttall, A. L. (**1995**). ''Frequency modulations in impulse responses and cross-correlation functions of basilar membrane motion in the guinea pig,'' Assoc. Res. Otolaryngol. Abs. **18**, 747.
de Boer, E., and Nuttall, A. L. (**1996**). ''Cochlear travel time and minimum phase,'' Assoc. Res. Otolaryngol. Abs. **19**, 57.
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. (**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. (**2000a**). ''The mechanical waveform of the basilar membrane. II. From data to models—and back,'' J. Acoust. Soc. Am. **107**, 1487–1496.
de Boer, E., and Nuttall, A. L. (**2000b**). ''The mechanical waveform of the basilar membrane. III. Intensity effects,'' J. Acoust. Soc. Am. **107**, 1497–1507.
Freeman, D. M., and Weiss, T. F. (**1990**). ''Hydrodynamic analysis of a two-dimensional model for micromechanical resonance of free-standing hair bundles,'' Hear. Res. **48**, 37–68.
Goblick, T. J., and Pfeiffer, R. R. (**1969**). ''Time-domain measurements of cochlear nonlinearities using combination click stimuli,'' J. Acoust. Soc. Am. **46**, 924–938.
Gummer, A. W., Smolders, J. W. T., and Klinke, R. (**1987**). ''Basilar membrane motion in the pigeon measured with the Mössbauer technique,'' Hear. Res. **29**, 63–92.
Hubbard, A. E., and Mountain, D. C. (**1996**). ''Models of the cochlea,'' in *Auditory Computation*, edited by H. L. Hawkins, T. A. McMullen, A. N. Popper, and R. R. Fay (Springer, New York), pp. 62–120.
Hudspeth, A. J., and Lewis, R. S. (**1988**). ''A model for electrical resonance and frequency tuning in saccular hair cells of the bull-frog, *Rana catesbeiana*,'' J. Physiol. (London) **400**, 275–297.
Irino, T., and Patterson, R. D. (**1997**). ''A time-domain, level-dependent auditory filter: The gammachirp,'' J. Acoust. Soc. Am. **101**, 412–419.
Irino, T., and Patterson, R. D. (**2000**). ''A gammachirp perspective of cochlear mechanics that can also explain human auditory masking quantitatively,'' in *Recent Developments in Auditory Mechanics*, edited by H. Wada, T. Takasaka, K. Ikeda, K. Ohyama, and T. Koike (World Scientific, Singapore), pp. 230–236.
Kiang, N. Y. S. (**1984**). ''Peripheral neural processing of auditory information,'' in *Handbook of Physiology, Section 1: The Nervous System, Vol. 3 (Sensory Processes)*, edited by I. Darian-Smith (American Physiological Society, Bethesda), pp. 639–674.
Kiang, N. Y. S., and Moxon, E. C. (**1974**). ''Tails of tuning curves of auditory-nerve fibers,'' J. Acoust. Soc. Am. **55**, 620–630.
Kiang, N. Y. S., Watanabe, T., Thomas, E. C., and Clark, L. F. (**1965**). *Discharge Patterns of Single Fibers in the Cat's Auditory Nerve* (MIT Press, Cambridge).
Kim, D. O., Siegel, J. H., and Molnar, C. E. (**1979**). ''Cochlear nonlinearities in two-tone responses,'' Scand. Audiol. Suppl. **9**, 63–81.
Lewis, R. S. (**1985**). ''The ionic basis of frequency selectivity in hair cells of the bull-frog's sacculus,'' Ph.D. thesis, California Institute of Technology.
Liberman, M. C. (**1978**). ''Auditory-nerve response from cats raised in a low-noise chamber,'' J. Acoust. Soc. Am. **63**, 442–455.
Liberman, M. C. (**1982**). ''The cochlear frequency map for the cat: Labeling auditory-nerve fibers of known characteristic frequency,'' J. Acoust. Soc. Am. **72**, 1441–1449.
Lin, T., and Guinan, J. J. (**2000**). ''Auditory-nerve-fiber responses to high-level clicks: Interference patterns indicate that excitation is due to the combination of multiple drives,'' J. Acoust. Soc. Am. **107**, 2615–2630.
Martin, P., and Hudspeth, A. J. (**1999**). ''Active hair-bundle movements can amplify a hair cell's response to oscillatory mechanical stimuli,'' Proc. Natl. Acad. Sci. U.S.A. **96**, 14306–14311.
Møller, A. R., and Nilsson, H. G. (**1979**). ''Inner ear impulse response and basilar membrane modelling,'' Acustica **41**, 258–262.
Mountain, D. C. (**2000**). ''Analysis of auditory nerve fiber responses supports multimode hypothesis,'' Assoc. Res. Otolaryngol. Abs. **23**, 246.
Neely, S. T. (**1983**). ''The cochlear amplifier,'' in *Mechanics of Hearing*, edited by E. Boer and M. A. Viergever (Martinus Nijhoff, The Hague), pp. 111–118.
Neely, S. T., and Kim, D. O. (**1986**). ''A model for active elements in cochlear biomechanics,'' J. Acoust. Soc. Am. **79**, 1472–1480.
Olson, E. S. (**1998**). ''Observing middle and inner ear mechanics with novel intracochlear pressure sensors,'' J. Acoust. Soc. Am. **103**, 3445–3463.
Olson, E. S. (**1999**). ''Direct measurement of intracochlear pressure waves,'' Nature (London) **402**, 526–529.
Papoulis, A. (**1962**). *The Fourier Integral and its Applications* (McGraw-Hill, New York).
Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (**1992**). *Numerical Recipes in C: The Art of Scientific Computing* (Cambridge University Press, Cambridge).
Puria, S., and Allen, J. (**1991**). ''A parametric study of the cochlear input impedance,'' J. Acoust. Soc. Am. **89**, 287–309.
Puria, S., and Allen, J. (**1998**). ''Measurements and model of the cat middle ear: Evidence of tympanic membrane acoustic delay,'' J. Acoust. Soc. Am. **104**, 3463–3481.
Recio, A., and Rhode, W. S. (**2000**). ''Basilar membrane responses to broadband stimuli,'' J. Acoust. Soc. Am. **108**, 2281–2298.
Recio, A., Rich, N. C., Narayan, S. S., and Ruggero, M. A. (**1998**). ''Basilar-membrane responses to clicks at the base of the chinchilla cochlea,'' J. Acoust. Soc. Am. **103**, 1972–1989.
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.
Rhode, W. S., and Cooper, N. P. (**1996**). ''Nonlinear mechanics in the apical turn of the chinchilla cochlea *in vivo*,'' Aud. Neurosci. **3**, 101–121.
Robles, L., Rhode, W. S., and Geisler, C. D. (**1976**). ''Transient response of the basilar membrane measured in squirrel monkeys using the Mössbauer effect,'' J. Acoust. Soc. Am. **59**, 926–939.
Ruggero, M. A., Rich, N. C., and Recio, A. (**1992**). ''Basilar membrane responses to clicks,'' in *Auditory Physiology and Perception*, edited by Y. Cazals, K. Horner, and L. Demany (Pergamon, Oxford), pp. 85–91.
Sarpeshkar, R. (**2000**). ''Traveling waves versus bandpass filters: The silicon and biological cochlea,'' in *Recent Developments in Auditory Mechanics*, edited by H. Wada, T. Takasaka, K. Ikeda, K. Ohyama, and T. Koike (World Scientific, Singapore), pp. 216–222.
Shekhter, I., and Carney, L. H. (**1997**). ''A nonlinear auditory nerve model for CF-dependent shifts in tuning with sound level,'' Assoc. Res. Otolaryngol. Abs. **20**, 617.
Shera, C. A. (**2001**). ''Intensity invariance of fine time structure in basilar-membrane click responses: Implications for cochlear mechanics,'' J. Acoust. Soc. Am. (in press).
Shera, C. A., and Guinan, J. J. (**2000a**). ''Frequency dependence of stimulus-frequency-emission phase: Implications for cochlear mechanics,'' in *Recent Developments in Auditory Mechanics*, edited by H. Wada, T. Takasaka, K. Ikeda, K. Ohyama, and T. Koike (World Scientific, Singapore), pp. 381–387.
Shera, C. A., and Guinan, J. J. (**2000b**). ''Reflection-emission phase: A test of coherent reflection filtering and a window on cochlear tuning,'' Assoc. Res. Otolaryngol. Abs. **23**, 545.
Shera, C. A., and Zweig, G. (**1993**). ''Order from chaos: Resolving the paradox of periodicity in evoked otoacoustic emission,'' in *Biophysics of Hair Cell Sensory Systems*, edited by H. Duifhuis, J. W. Horst, P. van Dijk, and S. M. van Netten (World Scientific, Singapore), pp. 54–63.
Siebert, W. M. (**1968**). ''Stimulus transformations in the peripheral auditory system,'' *Recognizing Patterns*, edited by P. A. Kolers and M. Eden (MIT Press, Cambridge), pp. 104–133.
Sondhi, M. M. (**1978**). ''Method for computing motion in a two-dimensional cochlear model,'' J. Acoust. Soc. Am. **63**, 1468–1477.
Weiss, T. F., and Leong, R. (**1985**). ''A model for signal transmission in an

ear having hair cells with free-standing stereocilia. III. Micromechanical stage,'' Hear. Res. **20**, 157–174.
Zweig, G. (**1976**). ''Basilar membrane motion,'' in *Cold Spring Harbor Symposia on Quantitative Biology* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), Vol. XL, pp. 619–633.
Zweig, G. (**1991**). ''Finding the impedance of the organ of Corti,'' J. Acoust. Soc. Am. **89**, 1229–1254.
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.

The effect of electrode configuration and duration of deafness on threshold and selectivity of responses to intracochlear electrical stimulation

Stephen J. Rebscher,[a] Russell L. Snyder, and Patricia A. Leake
Epstein Laboratory, Department of Otolaryngology, University of California, San Francisco, California 94143-0526

(Received 20 March 2000; revised 13 September 2000; accepted 21 February 2001)

This report examines the effects of intracochlear electrode configuration and mode of stimulation (bipolar or monopolar) on neural threshold and spatial selectivity in the inferior colliculus (IC) of the cat. Single and multiunit IC recordings were made in three groups of animals; acutely deafened adults (controls), neonatally deafened animals studied at 6 to 18 months of age and neonatally deafened cats studied at 2.5 to 6.5 years. Response thresholds were plotted versus IC depth to measure the spatial distribution of responses. The response selectivity for each stimulating configuration was defined as the width of the resulting spatial tuning curve (STC) measured at 6 dB above threshold. Spiral ganglion cell (SG) survival was examined histologically in all neonatally deafened animals and correlated with physiological results. Animals studied at less than 1.5 years had SG densities of 23.5%–64.4% of normal (mean=42.7%) while animals studied at greater than 2.5 years had densities of 5.1%–18.3% of normal (mean=9.9%). Electrophysiological results include the following. (1) Monopolar thresholds were 7–8 dB lower than bipolar thresholds in the same animals. (2) Varying the configuration of bipolar contacts (measured as radial, offset radial and longitudinal pairs) did not systematically affect IC threshold in either controls or short-term neonatally deafened animals. In contrast, the long-term neonatally deafened animals showed a difference in threshold with each configuration. (3) The spatial distributions ($Q_{6\,\mathrm{dB}}$) of responses to bipolar stimulation were approximately 40% more restricted than those for monopolar stimulation. (4) The spatial selectivity of neonatally deafened animals studied at ages up to 1.5 years was equal to that of control animals with normal auditory experience. However, selectivity was degraded in the older animals. (5) Selectivity was decreased in some animals with the longitudinal bipolar configuration and multiple response peaks were seen in several cases using this stimulus configuration. [DOI: 10.1121/1.1365115]

PACS numbers: 43.64.Me, 43.64.Qh [LHC]

I. INTRODUCTION

The performance of cochlear implant subjects has improved substantially during the past two decades. This improvement can be attributed to a number of factors including increases in the number of channels stimulated in each device and to changes in speech processing strategies. In order to make further improvements, we assume that it is desirable to minimize the interaction between individual stimulus sites. We believe that this will require both controlling electrical field summation and maximizing the spatial selectivity of excitation across the auditory nerve fiber array. Intracochlear electrode design, stimulation mode (monopolar or bipolar), electrode contact configuration and anatomical status of the cochlea, i.e., spiral ganglion cell survival and the presence or absence of their peripheral processes, may all be factors that affect both field summation and neural selectivity. Specifying the nature of the interactions among these factors is directly relevant to the development of future intracochlear electrodes. In addition, a better understanding of these fundamental parameters may help to explain the great variability in implant user performance.

Historically, there have been several different approaches to estimating these complex interactions. These have included computer models (simulations) of electrical stimulation, physiological studies measuring activity patterns in implanted animals and inference of spatial neural excitation patterns in behavioral studies in animal and human subjects.

Early computer models attempted to predict voltage field potentials within the cochlea (Strelioff, 1973; Spelman and Clopton, 1987; Finley, 1989) and neural threshold in response to electrical stimuli (Frankenhaeuser and Huxley, 1964; McNeal, 1976; Colombo and Parkins, 1987; Parkins and Colombo, 1987; Reilly and Bauer, 1987; Frijns and Kate, 1994; Frijns *et al.*, 1994). More recently, integrated three dimensional models have combined computed field strength and models of neural activation to predict complex patterns of auditory nerve excitation as a function of electrode placement and geometry (Finley *et al.*, 1990; Frijns *et al.*, 1995, 1996).

Physiological measurements in animals have provided data to test these computer models and to evaluate the effects

[a]Author to whom correspondence should be addressed: Department of Otolaryngology, Box 0526, Room C401, University of California, San Francisco, 533 Parnassus Avenue, San Francisco, CA 94143-0526. Electronic mail: reb@itsa.ucsf.edu

of electrode geometry and placement on neural activation. Previous animal studies have used electrically evoked auditory brainstem responses (EABRs) to measure threshold while varying electrode configuration and intracochlear electrode location (Marsh *et al.*, 1981; Vivion *et al.*, 1981; Lusted *et al.*, 1984; Shepherd *et al.*, 1993; Xu *et al.*, 1993b; Smith *et al.*, 1994). Mapping studies in the auditory nerve (Hartmann *et al.*, 1984; Parkins and Colombo, 1987; van den Honert and Stypulkowski, 1987; Shepherd and Javel, 1997) and inferior colliculus (IC) (Merzenich *et al.*, 1979; Snyder *et al.*, 1990, 1991) have estimated response selectivity, threshold and dynamic range again, while changing electrode type and placement. Measuring the ability to distinguish different stimulus sites in perceptual testing paradigms in both animals (Pfingst and Sutton, 1984; Pfingst *et al.*, 1985, 1995) and humans (Eddington *et al.*, 1978; Shannon, 1983a, b; Tong and Clark, 1985; Townshend *et al.*, 1987; Busby and Clark, 1996; Zwolan *et al.*, 1996; Chatterjee and Shannon, 1998; Chatterjee, 1999; Chatterjee *et al.*, 2000) also provides insight into the selectivity of electrical stimulation in the auditory system.

Several fundamental elements were different among each of the research strategies previously outlined. These variables included the species studied or modeled, status of the auditory system (deafened or hearing) and the electrode design and stimulus parameters used for intracochlear stimulation or simulation of electrically evoked activity. Therefore, it is difficult to directly compare the results from these studies and in some cases, the results appear to lead to conflicting conclusions. For example, van den Honert and Stypulkowski (1987) mapped the distribution of electrical activation across the auditory nerve array in hearing cats. They found that monopolar intracochlear stimulation (activation of an intracochlear contact referenced to an extracochlear return) resulted in almost no spatial selectivity, i.e., auditory nerve fibers close to and far away from the intracochlear contact were excited by nearly identical threshold currents. However, several psychophysical studies in both animals and human deaf subjects routinely report that even closely spaced monopolar electrodes are discriminable (Eddington *et al.*, 1978; Shannon, 1983a, b; Tong and Clark, 1985; Townshend *et al.*, 1987; Busby and Clark, 1996; Zwolan *et al.*, 1996). Results such as these clearly indicate fundamental gaps in our understanding of intracochlear stimulation. Some of these discrepancies are due to differences in methodology, electrode design, electrode position and species studied. Resolution of these differences, and an understanding of the effects which underlie them, will provide a clear direction for the future development of improved cochlear implants.

The goals of this study were to examine the following questions. (1) What is the effect of stimulation mode (bipolar or monopolar) and bipolar electrode configuration on threshold? (2) Is response selectivity affected by stimulation mode and contact configuration? (3) How does the survival of spiral ganglion cells and radial nerve fibers affect threshold and spatial selectivity? (4) Does the duration of deafness, and the progressive loss of spiral ganglion cells and radial nerve fibers, affect the spatial selectivity to electrical stimulation?

To address these points we directly measured response thresholds in penetrations across the tonotopically organized central nucleus of the inferior colliculus (ICC) of deaf cats. Using these measurements, we estimated the spatial distribution of responses to intracochlear electrical stimulation with several different intracochlear contact configurations. To minimize the variation between animals, and between individual IC penetrations in each animal, we compared responses to each configuration selected from a series of contacts on a single, silicone rubber carrier during single penetrations of the IC. This carrier was similar in design to clinically applied devices. In addition, we conducted these experiments in animals of varied duration of deafness in which the anatomical status of the cochlea was documented. Thus, these cats represented a broad range of deafness induced cochlear degeneration and allowed us to infer some effects of neural degeneration on response threshold and selectivity.

II. METHODS

A. Deafening

The deafening and implantation histories of the animals included in this study are presented in Table I. The control group consisted of acutely deafened adult cats with previously normal auditory experience. Four control adults were deafened unilaterally by intracochlear injection of neomycin (Snyder *et al.*, 1990) and five were deafened with a single subcutaneous injection of Kanamycin (400 mg/kg) followed by slow infusion of ethacrynic acid (10–25 mg/kg) (Xu *et al.*, 1993). One adult animal (105) was deafened by a single subcutaneous injection of kanamycin (400 mg/kg) followed by a single subcutaneous injection of aminooxyacetic acid (25 mg/kg) (Leake *et al.*, 1987). During the systemic deafening procedures auditory brainstem responses (ABRs) were measured continuously until no responses were observed at a level of 105 dB SPL. In most cases the control animals were deafened 1 to 2 weeks prior to the physiological experiment and implanted at the time of the experiment.

The experimental animals were neonatally deafened using a single intramuscular injection of neomycin sulfate (50 or 60 mg/kg) daily beginning 24 h after birth and continuing for 16 days. At 16 days of age the ABR for both right and left ears was recorded for 0.1-ms click stimuli presented at a level of 105 dB (SPL). If any residual hearing was observed at this time the administration of neomycin sulfate was continued until day 21 at which time the animal was retested. No ABR responses were seen in this retest and these animals were maintained without complication for periods of up to 6.5 years prior to implantation and the final physiological experiment.

B. Electrode and implantation

The left cochlea of each animal was implanted (see Leake *et al.*, 1991) with one of two multichannel intracochlear electrodes. The UCSF feline electrodes used in this study consist of a cylindrical, silicone carrier which tapers from 1.0 mm at the round window to 0.5 mm at the tip and hold either four or five Pt:Ir electrode contacts 225 μm in

TABLE I. The deafening and implantation histories for each animal in this study. Mean spiral ganglion survival for each animal is shown as the percent of normal. EABR threshold (200 μs click stimulus), ICC threshold (100 Hz sine stimulus) and the number of microelectrode penetrations made through the IC are also given.

Cat No.	Duration of deafness (months/method)	Duration of implant (weeks)	Electrode type	Spiral gang. survival (% normal) 90% of BM	EABR Thr. Bipolar 1,2 200 μs (μA)	ICC Thr. Bipolar 1,2 100 Hz Sine (μA)	No. of ICC recording electrode penetrations
Acutely deafened normal animals							
105	1.5/AOAA	4	UCSF	···	63	36	3
134	Kanamycin	Acute	UCSF	···	50	45	4
138	Kanamycin	Acute	UCSF	···	50	22	3
401	Intracochlear	Acute	UCSF	···	50	63	1
655	Intracochlear	Acute	UCSF	···	63	112	1
755	Intracochlear	Acute	UCSF	···	NA	32	2
865	Intracochlear	Acute	UCSF	···	63	32	2
242	Kanamycin	Acute	UCSF	···	316	56	2
553	Kanamycin	Acute	Wing	···	200	36	1
518	1/Kanamycin	Acute	Wing	···	63	5	3
Neonatally deafened animals (less than 1.5 years of age)							
K11	8/neomycin	Acute	UCSF	NA	100	50	2
K26	14/neomycin	3	UCSF	23.5	200	50	4
K30	7/neomycin	20	UCSF	NA	100	40	4
K44	6/neomycin	Acute	UCSF	64.4	100	14	1
K46	9/neomycin	Acute	UCSF	40.3	100	14	4
Long term neonatally deafened animals (more than 2.5 years of age)							
K03	31/neomycin	2	UCSF	13.1	398	126	3
K16	44/neomycin	Acute	UCSF	10.7	126	45	2
K24	30/neomycin	Acute	UCSF	7.2	251	50	1
K33	51/neomycin	1	UCSF	5.1	251	112	2
K51	78/neomycin	6	Wing	4.85	398	56	3
K73	41/neomycin	2	Wing	18.3	126	14	3

diameter. The original electrode, and the later version with flexible positioning ''wings,'' are shown in Fig. 1 and the average location for each contact is detailed in Table II. In most cases the anatomy of the cat scala tympani limits insertion to approximately 270 degrees from the round window. The apical electrode carrier is approximately space filling and the contacts (1, 1a and 2) are designed to lie adjacent to the spiral ganglion (contact 1) and beneath the osseous spiral lamina (1a and 2). It is important to note that the anatomical constraints of the cat cochlea make it impossible to assure that the apical and basal contacts of a widely separated bipolar pair are located in similar positions within the scala tympani. Because the cat cochlea widens dramatically at the base the basal contact location is inevitably more variable than that of the apical contact. For this reason it should be noted that the bipolar contact configurations compared in this study varied in three ways—radial orientation, longitudinal separation and position within the scala tympani volume. The results reported in this study represent the combined effects of these three factors. In terms of function this may in fact model variability existing in the implanted human cochlea, although it is anatomically less variable than the cat, where current generation intracochlear electrodes often do not assume a consistent position in the scala tympani (Ketten *et al.*, 1998).

C. Electrophysiological experiments

The methods used in the electrophysiological experiments conducted with these animals have been described previously (Snyder *et al.*, 1990, 1991). In brief, animals were sedated with ketamine (22 mg/kg) and acepromazine (2 mg/kg), shaved and an IV catheter was inserted in the cephalic vein. Sodium pentobarbital (40–60 mg/kg) was administered intravenously (I.V.) to induce an areflexic level of anesthesia. The head was mounted in a mouth-bar head holder. Each animal was continuously monitored and supplemental barbiturate was administered as required to maintain a surigical plane of anesthesia throughout the experiment. A craniotomy was performed to access the middle cranial fossa, and a portion of the tentorium was removed to expose the right inferior colliculus.

Electrical stimuli were generated by either of two software controlled digital to analog converters (D/A) based on TMS320 and TMS3200 signal processors. The stimulus for all ICC mapping consisted of three cycles of a 100-Hz sinusoid beginning and ending at zero voltage. All stimuli were capacitively coupled and electrically isolated in a custom low noise voltage to current amplifier (Vurek *et al.*, 1981). Electrical stimuli were switched between the intrachochlear electrode contacts with a custom switching array. The level for each stimulus was set using an audio attenuator. All experiments were performed in an electrically shielded enclosure. Neural signals were recorded using a primary amplifier (gain=1000, Bandpass=100 Hz to 3 KHz, Princeton Applied Research, Princeton, NJ or World Precision Instruments, Sarasota, FL) within the enclosure and a secondary amplifier (gain=100, Tektronix, Inc. Portland, OR). Analog signals were digitized with an analog to digital converter

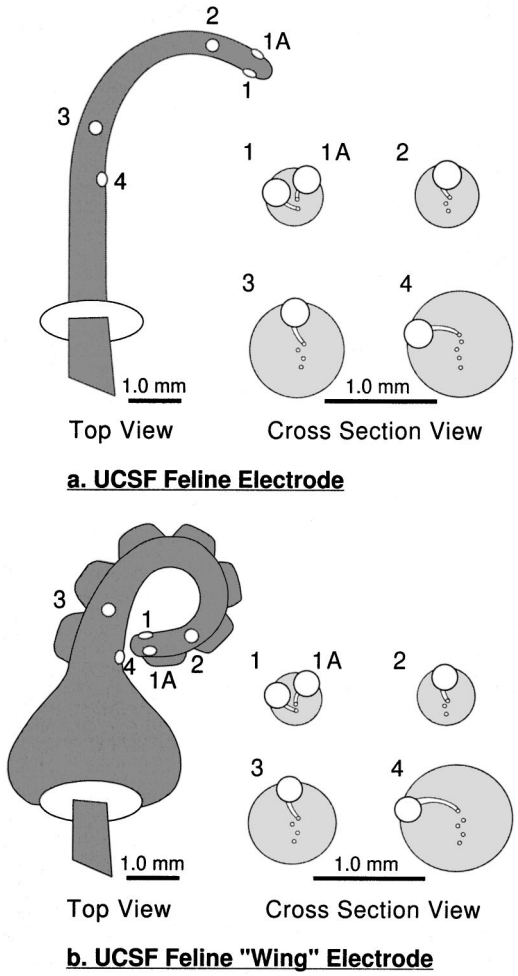


FIG. 1. The UCSF (a) and UCSF ''wing'' (b) electrode designs are shown in top and cross section views above. As in the clinical devices developed at UCSF the electrode lead wires shown in cross section are arranged in a vertically oriented line to increase the stiffness of the complete electrode in the vertical plane and minimize stiffness in the horizontal plane of the cochlear spiral. A surgical dacron fabric cuff was attached to each electrode and secured to the temporal bone near the round window.

(National Instruments, Austin, TX) controlled by custom software in a PC DOS environment. EABR responses were recorded differentially for 500 or 1000 biphasic alternating polarity electrical stimuli (200 μS/ph) using scalp electrodes placed over the ipsilateral bulla (reference), at the vertex of the skull (active) and over the contralateral bulla (ground). Multiunit responses in the IC were recorded using tungsten microelectrodes ($\approx$1 Mohm impedance, BAK Electronics, Germantown, MD) and three cycles of a 100-Hz sinusoid stimulus. The microelectrode trajectories were oriented in the coronal plane, 45 degrees off the sagital plane so that they traverse the IC approximately perpendicular to its cochleotopic organization. Response thresholds for each configuration of stimulating electrodes were determined audiovisually at 100-μm intervals along the recording penetration. Wherever possible, single units were discriminated from multiunit activity and responses were recorded as peristimulus time histograms (PSTHs). Multi- and single unit response thresholds were plotted as a function of IC depth to generate a ''spatial tuning curve'' (STC) allowing the pattern of distributed activity across the auditory nerve array to be inferred. Typical STC plots for two animals are shown in Fig. 2. Peak to peak current measurements are reported in dB (re 100 mA p-p) and in microamps (μA). To compare the relative selectivity of responses to different electrode configurations we measured the width of each STC at a level 6 dB above threshold. Only recordings in the central nucleus of the IC were used for these comparisons because the cochleotopic organization of the external nucleus is too compressed for accurate comparisons.

Statistical comparisons of threshold and spatial tuning (STC width) within an animal group were restricted to matched measurements collected for each electrode configuration during the same penetration through the IC. This procedure controls for variation both between IC penetrations within each animal and for differences between individual animals. Not all electrode configurations were tested in all IC penetrations or in all animals. Most notably, the radial and monopolar conditions were not tested in the neonatally deafened animals studied at less than 1.5 years of age.

This research and all procedures involving live animals were approved by the Committee on Animal Research at the University of California, San Francisco and fully conform to NIH guidelines for animal research.

III. RESULTS

A. Spiral ganglion survival

Eight of the ten control animals in this study were deafened immediately prior to the physiology experiment. For this reason the control group was considered to have normal, or nearly normal, survival of cochlear spiral ganglion neu-

TABLE II. The mean cochlear position of each electrode contact is given both as distance from the cochlear base (mm) and as the calculated frequency for that position. These position data present *in situ* measurement of ten original UCSF feline electrodes and nine UCSF ''wing'' electrodes as previously reported (Leake *et al.*, 2000). The separation (in mm measured contact center to contact center) for each contact pairing used in the study is also given.

	Mean distance from cochlear base (mm) (mean frequency location)				Contact separation for bipolar pairs		
Electrode type	Elect. 1	Elect. 2	Elect. 3	Elect. 4	Radial 1,1a	Offset 1,2	Long. 1,4
UCSF, $n=17$	9.9 (7.5 kHz)	8.8 (9.4 kHz)	7.4 (12.6 kHz)	5.8 (17.3 kHz)	0.47 mm c.c.	1.1 mm c.c.	4.0 mm c.c.
UCSF ''Wing''	11.6 (5.1 kHz)	10.6 (6.4 kHz)	7.4 (12.5 kHz)	7.1 (13.4 kHz)		1.1 mm c.c.	6.0 mm c.c.

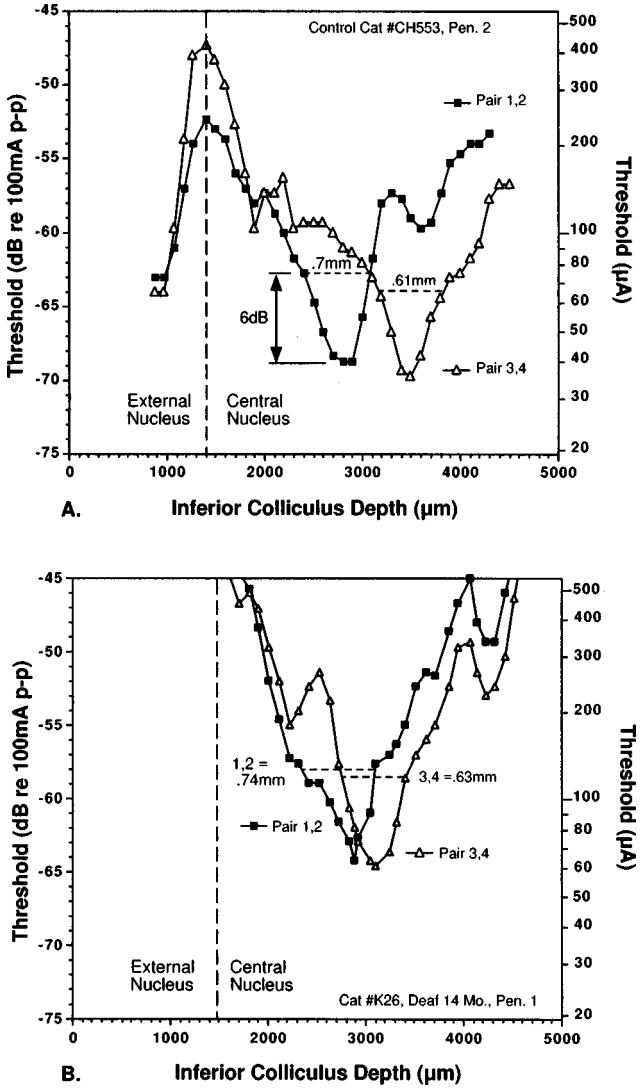


FIG. 2. Plotting the response threshold of single or multiple units in the inferior colliculus (IC) versus the recording depth describes the region of the IC activated by a particular electrode configuration at a given intensity level. The regions of excitation for an apical bipolar electrode pair (contacts 1 and 2) and a basal bipolar pair (contacts 3 and 4) are shown in the plots above. These areas of activation correspond to the expected locations based on the cochleotopic organization of the IC and its reflection of basilar membrane location. To compare the selectivity of the electrode configurations tested we measured the width of each spatial tuning curve (STC) at a level 6 dB above the minimum threshold. The STC widths for each stimulating electrode configuration were then compared to data from other configurations generated during the same microelectrode recording penetration. The upper plot illustrates the spatial tuning for a control animal. The lower plot shows very similar tuning for a neonatally deafened animal studied at an age of 14 months.

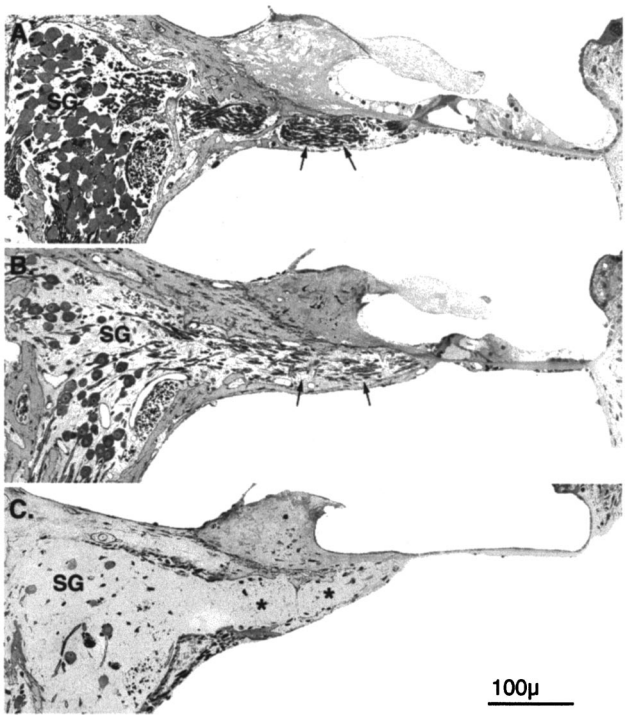


FIG. 3. Typical histological sections from each of the three animal groups are illustrated in this figure. (a) This micrograph illustrates an adult cat cochlea approximately 30 days after deafening with aminooxyacetic acid (AOAA) and kanamycin. Spiral ganglion cells (SG) are visible in nearly normal density in Rosenthal's canal and radial nerve fibers (arrows) can be seen passing through the osseous spiral lamina. (b) Neonatally deafened animals studied at less than 1.5 years of age showed significant loss of spiral ganglion cells (mean = 42.7% of normal). The number of peripheral fibers (arrows) was moderately reduced in these animals. (c) The long term deafened animals (>2.5 yrs.) showed severe degeneration of spiral ganglion cells and nearly complete loss of peripheral fibers (*). The mean spiral ganglion cell density for this group was 9.9% of normal.

rons and was not examined histologically. It should be noted, however, that the functional capacity of some or all of these neurons may have been altered by the deafening procedure. For each neonatally deafened animal the spiral ganglion cell density was evaluated in radial sections of the cochlea and expressed as a percentage of normal cell density using techniques that have been described previously (Leake *et al.*, 1991) (see Table I). The mean spiral ganglion cell survival for the short-term deafened animals (less than 1.5 years) was 42.7%. The mean spiral ganglion cell survival for animals in the long-term deafened group (greater than 2.5 year survival) was 9.9%. Figure 3 illustrates histological sections of the organ of corti and spiral ganglion that are representative of each group of animals. It is important to note the presence of radial nerve fibers in the osseous spiral lamina in the animals deafened for less than 1.5 years. In contrast, the long-term deafened group showed complete loss of these fibers in all cochlear turns in most animals with a few fibers seen in the extreme apex in some animals.

B. Threshold and spatial selectivity

1. Control animals

Table III summarizes the average minimum ICC response threshold and spatial tuning for monopolar, radial bipolar, offset radial bipolar and longitudinal bipolar electrode configurations measured in the acutely deafened, control animals. Since all combinations were not tested in all IC penetrations, or in all animals, these statistical comparisons were restricted to penetrations in which both test configurations were compared directly. For statistical analysis within each group Student's paired *t* test was used to compare values for each combination of stimulating electrodes within a penetration. A *P* value less than 0.05 was considered significant. In this group the mean threshold to monopolar stimulation (-74.9 dB, 18 μA) was lower than the threshold to all bi-

TABLE III. Electrophysiologic data from the ICC in the control group of animals. It should be noted that not all animals were implanted with electrodes holding the same combinations of stimulating contacts. Thus, the number of animals in which a given configuration was compared varies within the group. As the comparison standard the mean threshold and STC values for pair 1,2 represent the mean of all data for pair 1,2. The values for the other configurations represent only the data used in paired comparison with the offset configuration (1,2). Throughout this study all statistical comparisons were restricted to IC penetrations in which the threshold and STC width were measured for each test comparison in the same penetration. STC width is defined as the width of the spatial tuning curve measured at 6 dB above minimum threshold. *P* values are given for Student's paired *t* test compared with electrode pair 1,2.

Stimulation mode	No. of Animals	Contact configuration	Mean threshold (dB)		Mean STC width (mm)	
Monopolar	8	1	−75.8±9.3	$p=0.001$[a]	1.20±0.67	$p=0.005$[a]
Monopolar	8	2	−74.0±9.5	$p=0.03$[a]	1.30±0.63	$p=0.04$[a]
Bipolar	6	1,1a	−68.6±10.0	$p=0.3$[a]	0.86±0.54	$p=0.26$[a]
Bipolar	10	1,2	−67.5±8.0	na	0.78±0.19	na
Bipolar	10	1,4	−68.1±10.2	$p=0.17$[a]	0.97±0.37	$p=0.29$[a]

[a]*p* values for Student's paired *t* test, comparison with electrode pair 1,2.

polar pairs. In contrast, there was no significant difference in thresholds among the three bipolar configurations in this group of animals.

Spatial tuning data for monopolar and bipolar stimulus configurations in these control animals are also shown in Table III. Although individual penetrations in some animals demonstrate quite narrow tuning for monopolar stimulation the mean STC width for this condition was ≈40% greater than that for offset radial stimulation and this difference was statistically significant. The variability seen in different animals in response to monopolar versus bipolar stimulation is illustrated in Fig. 4. Although most of the points in this scatter plot indicate that monopolar stimulation is less selective than bipolar stimulation, many points also lie on or below the iso-width line. There was no significant difference in selectivity measured among the three bipolar configurations within this group. Two examples of STC plots comparing minimum threshold and spatial selectivity for different electrode configurations are shown in Fig. 5. The IC representations for monopolar and bipolar stimulation measured in a single penetration are compared in Fig. 5(a). The similarity in threshold, and the differences sometimes seen in spatial tuning with varied bipolar contact separation, are shown in Fig. 5(b).

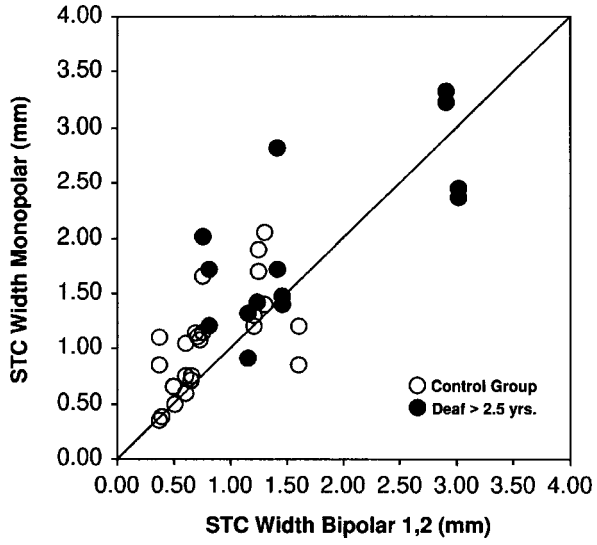


FIG. 4. In this scatter plot the STC width for monopolar stimulation (intracochlear electrode 1 or 2 versus an extracochlear ground electrode) is plotted against the STC width for offset radial stimulation (pair 1,2) for both the control and long deafened groups of animals. In each instance the STC value for the two stimulus conditions was measured within the same IC penetration. Points located above the iso-width line indicate that the monopolar stimulus condition resulted in less selective activation of the IC. Although selectivity was often less in the monopolar mode it is clear that in many observations monopolar stimulation was as restricted, or in some cases even more restricted, than bipolar stimulation as indicated by points below the iso-width line.

2. Short-term neonatally deafened animals (<1.5 years)

Minimum IC thresholds and STC widths for bipolar offset radial (pair 1,2) and longitudinal (pair 1,4) configurations in the neonatally deafened animals studied at less than 1.5 years are shown in Table IV. The monopolar and radial bipolar (pair 1,1a) configurations were not tested in these subjects. There was no significant difference between the threshold using the offset radial and longitudinal configurations in this group of animals. Although the response threshold was not significantly different for offset radial versus longitudinal bipolar stimulation in this group, the spatial selectivity was significantly narrower for the offset radial configuration and the variance between animals was much less, a difference also seen in the control group data.

3. Long-term neonatally deafened animals (>2.5 years)

In contrast to either the control group or the neonatally deafened animals studied at less than 1.5 years the minimum thresholds for the radial configuration (1,1a) and the longitudinal configuration (1,4) were significantly different than those for the offset pair (1,2) in these animals (Table V). In paired comparisons the threshold for the offset radial pair was ≈4 dB lower than that for the radial configuration and the threshold for the longitudinal configuration was again ≈4 dB lower than that for the offset radial configuration.

In general, the spatial selectivity of ICC responses was reduced in the long deafened animals compared to that seen in either the normal group (see Fig. 4) or the short-term

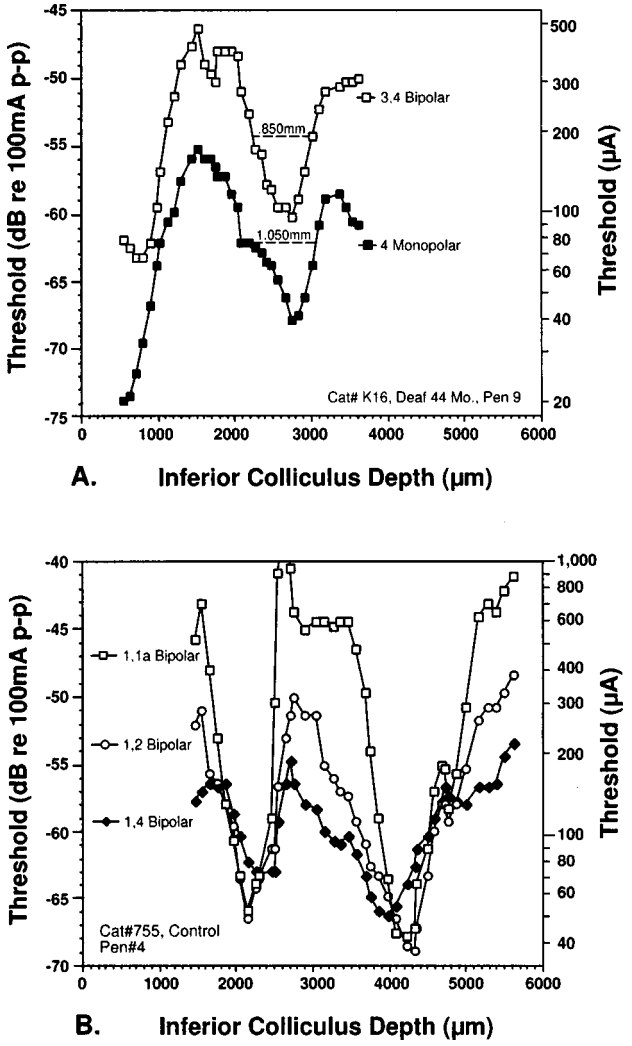


FIG. 5. The spatial distribution of responses in the ICC is illustrated for monopolar and bipolar stimulation modes in a long deafened animal (5a) and for radial, offset radial and longitudinal bipolar configurations in a prior normal control animal (5b). In the second example the STC width increases systematically with increasing interelectrode separation. This effect was seen in some, but not all, animals.

deafened group. There was no significant difference in the STC width, versus pair 1,2, of any of the five electrode configurations tested.

C. Comparisons of animal groups—changes in threshold and spatial selectivity over time

In addition to measuring the effect of contact configuration within each animal group we compared threshold and spatial tuning between these groups. This comparison examined the influence of auditory system degeneration on each measure and provides insight into how progressive degeneration might affect processing of electrically evoked auditory information.

Figure 6 compares the mean response sensitivity for each stimulating condition in the control, neonatally deafened (<1.5 years) and long-term deafened animals (>2.5 years). Interestingly, despite the lack of auditory experience in the deafened animals and very significant differences in spiral ganglion cell and radial nerve fiber survival, the mean IC response threshold for any selected electrode configuration was not significantly different between the three groups.

Figure 7 compares spatial tuning for each electrode configuration across the three groups of animals. *P* values for each statistically significant comparison are indicated in the figure. It is particularly interesting to note that with each electrode configuration tested the STCs from animals deafened less than 1.5 years were equal in selectivity to those of control animals even though these neonatally deafened animals developed without acoustic input and both the dendrites and spiral ganglion cells had degenerated significantly. In contrast, STC widths were broader in the long-term deafened animals. Although the difference in STC widths for monopolar and radial bipolar configurations between the control animals and the long deafened group appears to be relatively large, the differences are not statistically significant, possibly because these configurations were tested in fewer animals.

To examine this progressive loss of spatial selectivity we plotted the STC widths, for electrode pair 1,2, as a function of the duration of deafness for all neonatally deafened animals (Fig. 8). As described above the spatial tuning in animals up to 1.5 years was unchanged. In fact, these animals were all at or below the mean STC width for control animals. In contrast, the STC width in older animals increased with age.

D. Spatial representation with widely spaced bipolar electrodes

The IC representation of widely spaced electrodes (1,4) observed in these experiments produced occasional anomalies in STCs that were not seen in other bipolar configurations or in the monopolar mode. Three of these response patterns are illustrated in Fig. 9 and indicate that widely separated longitudinal pairs may be represented in several, unpredictable forms. These include (1) relatively broad tuning, (2) a single, sharply defined tuning peak that may be located at the cochleotopic location of either one of the

TABLE IV. Electrophysiologic data from the ICC in the neonatally deafened group of animals studied at less than 1.5 years of age. Only electrode pairs (1,2) and (1,4) were studied in this group. Again, all statistical comparisons are paired comparisons of response threshold and STC width measured in the same IC penetration. *P* values are given for Student's paired *t* test.

Stimulation mode	No. of animals	Contact configuration	Mean threshold (dB)		Mean STC width (mm)	
Bipolar	5	1,2	-69.2 ± 8.0	na	0.73 ± 0.09	na
Bipolar	3	1,4	-67.4 ± 10.1	$p=0.20$[a]	1.17 ± 0.33	$p=0.03$[a]

[a]*p* values for Student's paired *t* test, comparison with electrode pair 1,2.

TABLE V. Electrophysiologic data from the ICC in the long term neonatally deafened group of animals studied at more than 2.5 years of age. *P* values are given for Student's paired *t* test compared with electrode pair 1,2.

Stimulation mode	No. of animals	Contact configuration	Mean threshold (dB)		Mean STC width (mm)	
Monopolar	4	1	-73.0 ± 10.4	$p=0.005$[a]	1.97 ± 0.77	$p=0.15$[a]
Monopolar	4	2	-73.3 ± 9.1	$p=0.02$[a]	1.85 ± 0.48	$p=0.25$[a]
Bipolar	4	1,1a	-61.6 ± 9.9	$p=0.001$[a]	1.76 ± 0.99	$p=0.37$[a]
Bipolar	6	1,2	-64.9 ± 7.1	na	1.35 ± 0.58	na
Bipolar	6	1,4	-68.3 ± 8.7	$p=0.03$[a]	1.66 ± 0.94	$p=0.10$[a]

[a]*p* values for Student's paired *t* test, comparison with electrode pair 1,2.

stimulating contacts, or (3) a double peaked representation with threshold minima at the locations of both contacts. One can imagine that a series of such bipolar electrode pairs, each with an idiosyncratic representation, might produce a very distorted multichannel auditory signal or that double peaked representations might be perceived as multiple pitch sensations. It is important to note that there may be several underlying reasons why some responses to widely separated bipolar pairs, as represented by pair 1,4 in this study, differ from responses to more closely spaced pairs. In addition to the separation between contact sites the individual contacts in a widely spaced pair are more likely to vary in their position within the volume of the scala tympani and in their radial location within this volume. Each of these factors may contribute to the unique response patterns observed.

IV. DISCUSSION

These experiments provide a systematic comparison of intracochlear stimulus configurations in an electrode geometry similar to clinical devices and across a range of pathologies similar to that seen in human subjects. Previous animal studies, particularly those by van den Honert and Stypulkowski (1987), and some computer models suggest that current spreads rapidly within the scala tympani for many of the circumstances evaluated. This rapid spread of current will theoretically result in broadly distributed neural excitation and lead to the prediction that patients would experience extensive interactions between electrode sites which might negate, or substantially reduce, the benefits of multiple information channels. Indeed, interactions between adjacent stimulating sites have been observed with all clinically applied cochlear implant systems. However, the overwhelming success of multichannel devices clearly suggests that at least some level of channel independence is achieved with all of these devices.

These experiments demonstrate that intracochlear electrical stimulation produces restricted response regions in the central nucleus of the inferior colliculus that are appropriately matched to their cochleotopic location. Further, in most cases, these responses are clearly affected by the mode of stimulation, by the electrode configuration used and by the cochlear pathology present in each individual.

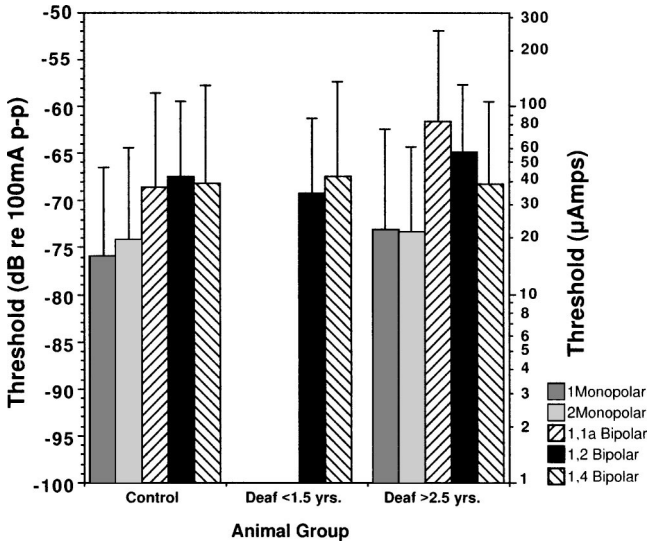


FIG. 6. The mean threshold for each electrode configuration was compared for the three animal groups. Despite very large differences in neural survival no significant difference in threshold was observed between these three groups for any of the electrode configurations tested.

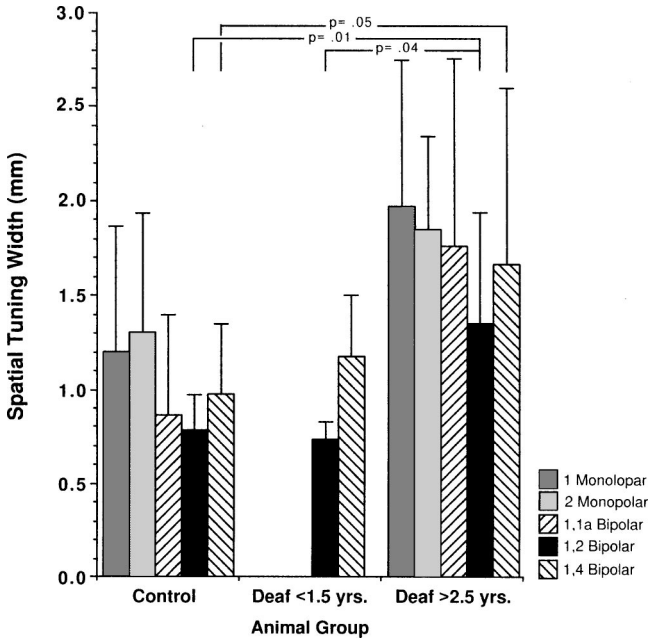


FIG. 7. The relationship between the effect of electrode configuration and the duration of deafness is shown in this figure. The mean STC width is shown for each electrode configuration tested in each animal group. In general STCs for each stimulating configuration in the longest deafened animals were wider than the same configuration in the other animal groups. The selectivity of each configuration was compared statistically between each animal group. Statistically significant comparisons are indicated by connecting lines with the *p* value indicated.

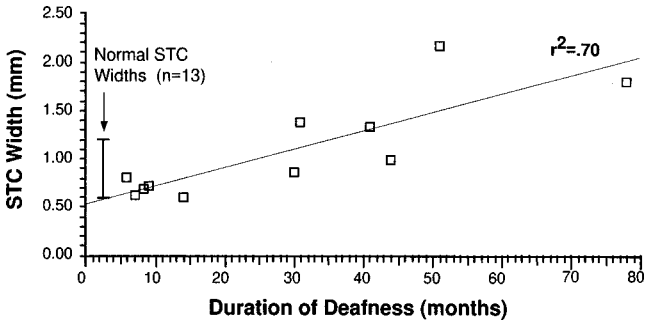


FIG. 8. The selectivity of animals studied with a duration of deafness greater than 2.5 years was diminished. To evaluate how this degradation occurred over time we plotted the STC width for the apical bipolar pair (1,2) versus the duration of deafness for all neonatally deafened animals. We found that despite spiral ganglion cell loss of almost 58%, and the loss of most peripheral dendrites, the spatial selectivity in animals deafened less than 1.5 years was equal to the selectivity of the control group. In contrast, the longer deafened animals showed a clear degradation of selectivity over time.

A. Evaluation of neural survival

It has long been assumed that the loss of the primary afferent neurons of the auditory nerve following deafness significantly affects the performance of individual cochlear implant patients. To model this variable population we measured the effects of electrode configuration in animals grouped into three categories based on their duration of deafness and associated neural survival. As shown in Table I and in the micrographs in Fig. 3 the control, neonatally deafened (<1.5 years) and long-term neonatally deafened animals (>2.5 years) in this study represent three distinct groups based on their anatomical features.

The mean spiral ganglion survival for the neonatally deafened animals examined at less than 1.5 years of age was 42.7% of normal. This value agrees with measurements of aminoglycoside neonatally deafened animals studied at similar ages in previous reports (Leake *et al.*, 1991, 1992, 1995). Histologic examination of these subjects also revealed a moderate number of surviving radial nerve fibers extending from the spiral ganglion to the habenula perforata with a few of these fibers passing into the degenerating tunnel of corti. In contrast, mean spiral ganglion survival in the long-term neonatally deafened animals (>2.5 years) measured only 9.9% of normal and very few radial nerve fibers were seen in these cases. This marked loss of both spiral ganglion neurons and dendrites at durations of deafness exceeding 2.5 years also has been previously documented (Leake and Hradek, 1988). To allow comparison of the pathology observed in these animals with that in the deaf human population we examined the reported histopathology of deaf human subjects.

Table VI summarizes pathology data in 113 human cadaver temporal bones from subjects considered appropriate for cochlear implantation. The mean spiral ganglion cell survival reported in these studies was 56.7% of normal with a range of 41% to 77% (48% to 74% over the basal 20-mm region adjacent to most clinically applied implant electrodes). The range of mean dendrite survival was 3% to 72% overall (0% to 65% in the cochlear base). Thus, the three

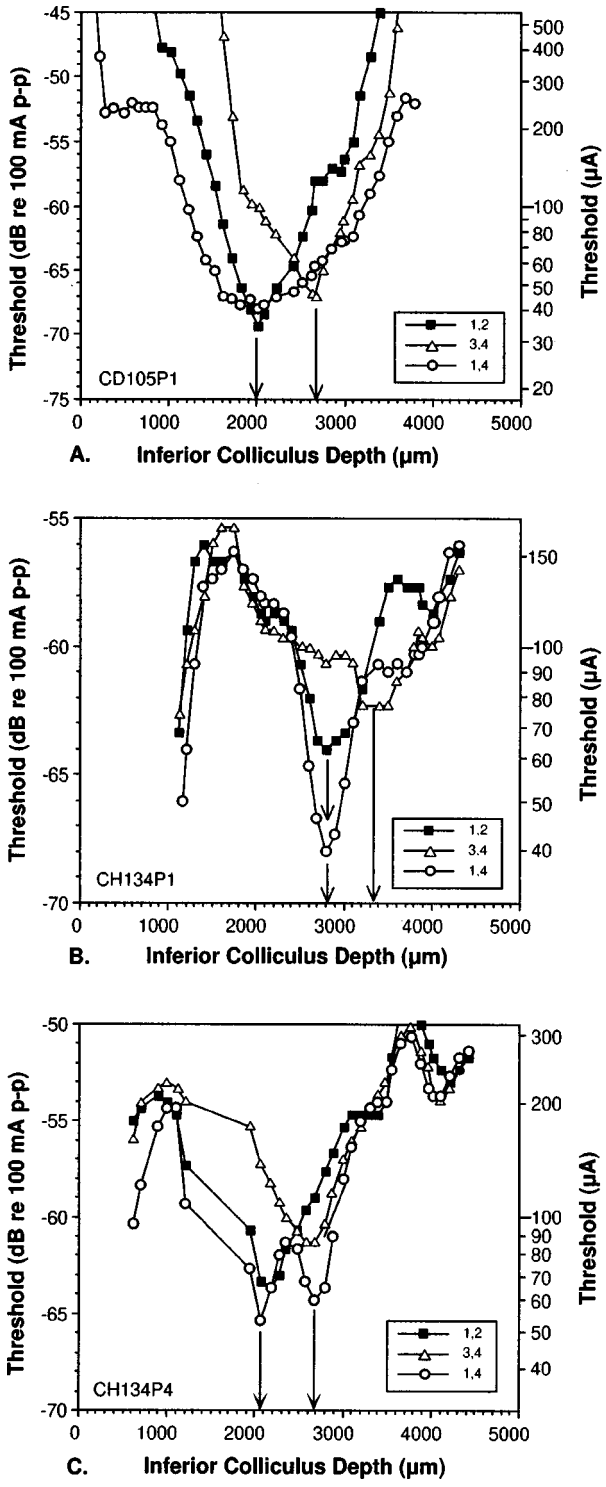


FIG. 9. Three types of representation were seen for widely spaced stimulating electrode configurations, i.e., 4–6 mm. The most characteristic representation is shown in (a). In this case, the tuned response area appears similar to the response for a more closely spaced bipolar pair (apical 1,2 or basal 3,4); however, the spatial tuning curve is broader and includes much, or all, of the area represented by the narrower stimulating pairs. (b) In the second class of representation for widely separated electrodes the response is very sharp and is located at the minimum location for one contact of the stimulating pair. In the example above the STC for pair (1,4) is located near the minimum threshold location for pair (1,2). (c) In the third type of representation the response appears as two sharply tuned regions of activation. This double response peak has been predicted in computer models of longitudinally separated stimulation sites. In an electrode array of longitudinally separated bipolar pairs it is possible that these idiosyncratic representations may generate perceptual distortions.

TABLE VI. Spiral ganglion cell and dendrite survival data from 113 deaf human subjects are presented from five published reports. These data, combined with the neural survival data presented for the animals in this study (see Table I), give some perspective for the interpretation of how these results, and the results of other animal studies and computer simulations, may relate to clinical cochlear implant subjects. The neural survival figures are divided into Basal (mean survival from the cochlear base to 20 mm) and mean overall survival because currently applied intracochlear electrodes are positioned within this basal region of the scala tympani.

Group	No. of Subjects	Hearing loss (dB SPL)	Dendrite survival (% normal)	Basal (to 20 mm) dendrite Survival (% normal)	Ganglion survival (% normal)	Basal (to 20 mm) ganglion survival (% normal)	No. of ganglion cells (total)
Disease/genetic[a]	14	Profound	21.8±27.8	13.0±21.7	47.3±16.7	48.6±19.7	16 084±5672
Disease/genetic[b]	3	Profound	46.8±12.6	30.0±14.4	70.9±2.5	58.7±5.6	23 839±832
Disease/genetic[c]	6	Profound	3.3±8.0	1.1±2.6	54.1±24.7	51.3±8.8	18 196±8307
Ototoxicity[a]	1	≈90 dB	10.0	0.0	57.8	54.0	19 608
Ototoxicity/trauma[b]	10	≈90 dB	71.8±15.4	64.5±17.7	77.0±22.8	73.5±25.2	25 899±7672
Cochlear Implant[d]	13	na	21.4±22.3	na	40.6±15.7	na	12 175±4720
Varied Pathology[e]	66	na	na	na	47.3±19.1	47.9±19.3	13 444±5436

[a]Hinojosa and Marion, 1983.
[b]Suzuka and Schuknecht, 1988.
[c]Hinojosa *et al.*, 1991.
[d]Fayad *et al.*, 1991.
[e]Nadol *et al.*, 1989.

groups of animals compared in this study represent stages in the degenerative processes that parallel those occurring in the patient population.

B. The effect of electrode configuration on threshold

Threshold is an important measure of the efficiency of a cochlear prosthesis. From an engineering perspective a device with lower current thresholds will require lower operating voltages, smaller electronic components and generate lower charge densities at the electrode-to-tissue interface. From a functional standpoint such a design would have greater battery life, reduced package size and potentially greater performance associated with increased dynamic range in some patients.

IC unit thresholds for monopolar stimulation were lower than thresholds for any bipolar configuration in each of the animal groups tested in this study. Comparing the mean monopolar threshold versus the threshold for the 1,2 bipolar pair the difference observed was 7.4 dB in the control group and 8.3 dB in the long deafened group. These values are similar to those predicted by Colombo and Parkins (1987), measured in monkeys (Colombo and Parkins, 1987) and in cats (van den Honert and Stypulkowski, 1987) and observed in human subjects (Shannon, 1983a, b; Brown *et al.*, 1996). No threshold difference was seen in the mean response for contact 1, the medial monopolar contact (see Fig. 1), as compared to contact 2 which was designed to be positioned under the basilar membrane and osseous spiral lamina. The lack of difference for these two radial positions, which is predicted in Frijns *et al.*'s model (1996) and observed in EABR studies (Shepherd *et al.*, 1993), has also been reported in monkeys (Pfingst and Sutton, 1983) and in humans (Shannon, 1983a). It should be noted, however, that the final rotational placement of electrode contacts, particularly near the tip of a flexible carrier, cannot be reliably predicted. In the course of examining the temporal bones in this study we found that often the tip of the electrode, containing the most apical stimulating contact (1), rotated laterally so that the actual radial location of the contact was similar to that of contact 2. It is reasonable to assume that similar variability has occurred in previous studies and in human devices, particularly near the more flexible apex of the electrode, and that this variable confounds attempts to precisely compare the effectiveness of radial contact positioning.

The effect of manipulating bipolar electrode configuration within each group of animals was directly related to cochlear pathology. In the control group and neonatally deafened group less than 1.5 years of age there was no difference in unit thresholds for the contact configurations tested. In contrast, there was a significant difference in the threshold for each configuration in the long-term deafened group (Table V). Similar effects of varying electrode combinations were observed in a number of other animal studies (see Merzenich and White, 1977; Pfingst and Sutton, 1983; van den Honert and Stypulkowski, 1987; Shepherd *et al.*, 1993). Vivion *et al.* (1981) made the distinction that this effect was measurable in animals with low nerve survival but not in animals with minimal pathology, a result that these findings confirm.

Computer models predict that thresholds for bipolar stimulation may be lowest for radial pairs (Finley *et al.*, 1990) or may decrease with increasing contact separation (Frijns *et al.*, 1996) depending on the location of the electrodes and the location of neurons activated. In practice, we observed both conditions. Although the mean threshold for radial stimulation (Table III) was lower than that for the offset radial condition the difference was not statistically significant and we frequently encountered the reverse effect.

C. The effect of pathology on response threshold

The similarity of IC unit thresholds between the three animal groups in this study has important implications for the theoretical modeling of intracochlear electrical stimulation and the design of future devices. Current modeling (Frijns *et al.*, 1996) predicts that the presence or absence of radial nerve fibers may strongly affect threshold and selectivity depending upon the placement of stimulating contacts in the scala tympani. When contacts were located close to the

spiral ganglion these effects were minimized. Because the apical portion of the electrode carrier in this study occupies most of the scala tympani contacts 1 and 2 were positioned close to the spiral ganglion in most cases. As modeled for this position the difference in bipolar threshold observed with the loss of dendrites was minimal ($\approx$2.5 dB for pair 1,2 control versus long deaf). Data from this study confirm the prediction that placing electrodes adjacent to the spiral ganglion will minimize the effects of neural degeneration and is thus a very favorable attribute for a clinical implant system.

In contrast to the similarity of IC thresholds across animal groups with varied pathology, EABR thresholds for the long deafened animals ($>$2.5 years) were higher than those of either the control group or the short-term neonatally deafened animals (see Table I). This change in far field threshold may be due to the ongoing loss of neurons, and thus a decrease in response magnitude. Additionally, it is possible that the increase in threshold may be an effect of degraded temporal synchrony in these animals. This observation may indicate that the difference between perceptual threshold, which is presumably triggered by a small number of unit responses, and EABR threshold, which requires a larger, synchronized response, may be a useful clinical indicator of neural survival.

D. The effect of electrode configuration on spatial selectivity

The ability to produce spatially restricted patterns of neural excitation is presumed to be essential for the effective operation of multichannel cochlear implants capable of implementing diverse speech processing strategies. Prior animal studies have mapped neural responses using microelectrode recording in the auditory nerve (van den Honert and Stypulkowski, 1987) and IC (Merzenich and White, 1977; Snyder *et al.*, 1990, 1991). In addition, areas of the brain responding to intracochlear electrical stimuli were imaged using 2-deoxyglucose (2-DG) autoradiography in the cochlear nucleus and IC (Ryan *et al.*, 1990). In each set of experiments the response patterns observed were restricted with at least some of the electrode configurations used and these regions of activation appropriately reflected the cochlear location of stimulation.

One of the primary goals of this study is to identify electrode configurations or stimulation strategies that produce the most selective patterns of neural activation. As described earlier previous animal studies and computer predictions suggest that a number of variables underlie the complex relationships between electrode configuration and response selectivity. The ways in which these observations and predictions differ from each other, and from those in this study, may offer insights into which electrode design features will most effectively improve performance.

1. Bipolar configurations

Several computer modeling studies (Colombo and Parkins, 1987; Finley *et al.*, 1990; Frijns *et al.*, 1996) predict that radial bipolar electrodes will produce the greatest selectivity among the bipolar configurations modeled. Experimental evidence of these highly restricted representations using the radial bipolar configuration in cats has been reported (van den Honert and Stypulkowski, 1987; Snyder *et al.*, 1990). Although the radial bipolar configuration generated some very narrow STCs in some animals in this study the mean selectivity was greatest for the offset radial configuration. The trend to broader representation with increased interelectrode spacing, seen in limited numbers of animals studied in previous experiments (Merzenich *et al.*, 1979; Snyder *et al.*, 1990), was also observed in all animal groups in this study. However, this difference was statistically significant in only the neonatally deafened animals studied at less than 1.5 years of age. Again, the large volume of the cat basal scala tympani makes the location of contact 4 in this study less reliable than that of the apical electrodes. A portion of the variability in the results with the 1,4 electrode pair may be due to the more variable placement of this contact.

2. Monopolar stimulation strategy

Computer models, physiological experiments and psychophysical studies comparing the selectivity of responses to monopolar and bipolar stimulation report widely varied predictions. In computer models Colombo and Parkins (1987) predicted that monopolar stimulation would be less selective than stimulation with bipolar electrodes. Conversely, Clopton and Spelman (1995) predicted that monopolar excitation could be more restricted than radial bipolar stimulation.

Each physiological study which has measured the activation patterns for bipolar and monopolar electrodes has found that the bipolar configuration produces more restricted excitation, however, the difference observed between the two strategies has ranged from no observed tuning with monopolar electrodes (van den Honert and Stypulkowski, 1987) to minimal tuning (Ryan *et al.*, 1990) to tuning which was broader but very similar to bipolar tuning (Merzenich *et al.*, 1979). Because there are considerable engineering advantages to the design and production of cochlear implant devices based on a monopolar stimulation strategy it is important to better understand the mechanisms which affect spatial selectivity. One clear difference between the observations of van den Honert and the current study is the use of a space filling, insulating carrier in these experiments. This design feature, and proximity of monopolar contacts to the spiral ganglion, are interrelated and may be a necessary condition for selective neural activation in the monopolar mode.

Behavioral studies in animals and human psychophysical studies provide an additional perspective on the question of monopolar versus bipolar selectivity. In behavioral tasks monkeys were able to discriminate between monopolar contacts spaced at 1-mm intervals with nearly 100% accuracy (Pfingst and Sutton, 1983). Human psychophysical studies demonstrate that subjects are able to successfully distinguish between the patterns of activation produced by either bipolar or monopolar stimulation. Indeed, most cochlear implant users are able to identify distinct percepts associated with each electrode site and rank these percepts in appropriate cochleotopic sequence (Eddington *et al.*, 1978; Shannon, 1983a, b; Townshend *et al.*, 1987). Taken one step further, most patients tested have been able to resolve the cochleotopic loca-

tion of neural representations generated with ''current focusing'' techniques to create one or more perceptually distinct ''virtual'' channels between contacts with interelectrode spacing as close as 0.75 mm (Townshend *et al.*, 1987; McDermott and McKay, 1994). Considering the mapping data from our study, and previous animal experiments, which indicate that the central representations of intracochlear electrical stimuli are relatively broad and overlapping over much of their dynamic range, it appears that these subjects are able to correctly discriminate subtle changes in either the fringes of representational fields or shifts in the center of this activity. Thus, it appears that it is not necessary that the response field for each stimulation location be fully independent to produce a unique percept with either bipolar or monopolar stimulation.

However, as mentioned previously, the ability to distinguish locations of electrical stimuli within the cochlea in sequential comparison tests may not accurately assess the practical capability of a device to deliver multiple channels of information simultaneously or nearly simultaneously. Two trends in cochlear implant speech processing make the ability to transfer information within a rapid or simultaneous temporal domain important. First, recent results with improved simultaneous analog processing indicate that some human subjects may achieve superior results with this strategy (Battmer *et al.*, 1999; Osberger and Fisher, 1999). Second, increased pulse rates for continuous interleaved sampling (CIS) processors may improve performance. As the presentation rate in these processors is increased the temporal distinction between two adjacent stimulus channels diminishes. This effect will increase dependence on the distinction of each information channel based on the location of its receptive field. With these developments in mind it is prudent to maintain the design capability to achieve maximum spatial selectivity with the design of each cochlear implant system.

E. The effect of pathology on spatial selectivity

As illustrated in Fig. 8 normal response selectivity was maintained in the neonatally deafened animals less than 1.5 years of age. Morphometric examination of these animals showed a range of spiral ganglion survival of 23.5% to 64.4% of normal (mean=42.7%). These data indicate that relatively few spiral ganglion cells are needed to convey the spatial distribution of electrical signals from the cochlea to the inferior colliculus. However, the long deafened animals (>2.5 years), with presumably greater degeneration throughout the auditory system, have significantly degraded spatial selectivity. These data support the implantation of prospective patients as soon as possible following confirmation of deafness. What is the critical difference in selectivity between these two groups? It seems unlikely that the gradual loss of spiral ganglion cells is the sole cause of broadened tuning in the long deaf animals when no decrease in selectivity was observed with a loss of up to $\approx 77\%$ of spiral ganglion cells in the younger animals.

Does the loss of dendrites account for decreased selectivity in long deafened animals? Computational modeling (Frijns *et al.*, 1996) indicates that the loss of dendrites will lead to a loss in response selectivity. As shown in Fig. 3 some dendrites were seen in the animals <1.5 years while virtually no fibers were observed in the animals at >2.5 years. Modeling predicts that the presence of radial nerve fibers will sharpen spatial representation by responding with very low threshold to a stimulating electrode in close proximity to functional dendrites. In the absence of these fibers the model assumes that excitation will occur only when the voltage field greater than threshold has expanded to include the spiral ganglion cell body. In addition to responding at higher threshold the model predicts that the distribution of responses will be broader at these distant targets, particularly if the stimulating contacts are located more laterally in the scala tympani. The loss of radial nerve fibers, which became nearly complete in the long deafened group of animals during this study, may be the primary factor in the degradation of tuning. The observed difference in selectivity was similar to that predicted by Frijns' model (1996). The fact that this degradation occurs very slowly, and is quite moderate in extent, may again be due to the placement of relatively large stimulating contacts close to the spiral ganglion. Both factors, contact size and proximity, would decrease the sensitivity of the system to the loss of radial nerve fibers.

The anomalous STCs seen with the longitudinal (1,4) contact configuration in some animals (see Fig. 9) confirm the predicted activation of multiple, separate regions of the auditory nerve by several models (Colombo and Parkins, 1987; Finley *et al.*, 1990; Frijns *et al.*, 1996); however, subjects stimulated with widely spaced bipolar electrodes generally do not report unique qualities associated with this stimulus geometry (McDermott and McKay, 1994). Although these distortions may be unnoticeable when perceived independently, it is possible that the interaction of several distorted inputs simultaneously or in rapid sequence might degrade performance.

V. CONCLUSIONS

The results of this study and those of previous reports indicate that the threshold and selectivity achieved with a cochlear implant are the results of complex interactions between electrode design, placement, species specific anatomy and individual pathology. The large variations in threshold and selectivity observed in different animals in this study highlight the system's sensitivity to even small changes in electrode position or individual anatomy. These dramatic effects have been suggested by computer modeling and previous animal studies. Together, these results argue strongly for the development of implant designs which will be positioned nearer to functional spiral ganglion cells, which can be safely implanted in young children and which can be highly customized to meet the needs of each patient. It is important to note that the capability to customize a device must not only be engineered into the system but must also be integrated into clinical fitting protocols to provide maximum benefit for each unique patient.

ACKNOWLEDGMENTS

The authors would like to acknowledge Dr. Ralph Beitel, Dr. Charlotte Moore, Dr. Maike Vollmer and Beth Dwan for their invaluable assistance in data collection and performance of the physiology experiments described in this study and Gary Hradek for the preparation of all histological materials, morphometric analysis of ganglion cell survival in these subjects and assistance in the preparation of this manuscript. This study was supported by NIH Contract No. NO1-DC-7-2105.

Battmer, R., Zilberman, Y., Haake, P., and Lenarz, T. (**1999**). ''Simultaneous Analog Stimulation (SAS)—Continuous Interleaved Sampler (CIS) pilot comparison study in Europe,'' Ann. Otol. Rhinol. Laryngol. **177**(Suppl.), 69–73.

Brown, C. J., Abbas, P. J., Borland, J., and Bertschy, M. R. (**1996**). ''Electrically Evoked Whole Nerve Action Potentials in Ineraid Cochlear Implant Users: Responses to Different Stimulating Electrode Configurations and Comparison to Psychophysical Responses,'' J. Speech Hear. Res. **39**, 453–467.

Busby, P. A., and Clark, G. M. (**1996**). ''Electrode Discrimination by Early-Deafened Cochlear Implant Users,'' Audiology **35**, 8–22.

Chatterjee, M. (**1999**). ''Effects of stimulation mode on threshold and loudness growth in multichannel cochlear implants,'' J. Acoust. Soc. Am. **105**, 850–860.

Chatterjee, M., and Shannon, R. V. (**1998**). ''Forward masking excitation patterns in multielectrode electrical stimulation,'' J. Acoust. Soc. Am. **103**, 2565–2572.

Chatterjee, M., Fu, Q., and Shannon, R. V. (**2000**). ''Effects of phase duration and electrode separation on loudness growth in cochlear implant listeners,'' J. Acoust. Soc. Am. **107**, 1637–1644.

Clopton, B. M., and Spelman, F. A. (**1995**). ''Electrode configuration and spread of neural excitation: compartmental models of spiral ganglion cell,'' Ann. Otol. Rhinol. Laryngol. **104** (Suppl. 166), 115–118.

Colombo, J., and Parkins, C. W. (**1987**). ''A Model of Electrical Excitation of the Mammalian Auditory-Nerve Neuron,'' Hear. Res. **31**, 287–312.

Eddington, D. K., Dobelle, W. H., Brackman, D. E., Mladejovsky, M. G., and Parkin, J. L. (**1978**). ''Auditory prosthesis research with multiple channel intracochlear stimulation in man,'' Ann. Otol. Rhinol. Laryngol. **87**(Suppl. 53), 5–39.

Fayad, J., Linthicum, F. H., Jr., Otto, S. R., Galey, F. R., and House, W. F. (**1991**). ''Cochlear implants: Histopathologic findings related to performance in 16 human temporal bones,'' Ann. Otol. Rhinol. Laryngol. **100**(Suppl. 10), 807–811.

Finley, C. C. (**1989**). ''A Finite-Element Model of Radial Bipolar Field Patterns in the Electrically Stimulated Cochlea—Two and Three Dimensional Approximations and Tissue Sensitivities,'' in IEEE 11th Annual International Conference Eng. Med. Biol. Soc., pp. 1059–1060.

Finley, C. C., Wilson, B. S., and White, M. W. (**1990**). ''Models of Neural Responsiveness to Electrical Stimulation, in *Cochlear Implants: Models of the Electrically Stimulated Ear*, edited by J. M. Miller and F. A. Spelman (Springer-Verlag, New York), pp. 55–96.

Frankenhaeuser, B., and Huxley, A. F. (**1964**). ''The Action Potential in the Myelinated Nerve Fiber of Xenopus laevis as Computed on the Basis of Voltage Clamp Data,'' J. Physiol. (London) **171**, 302–315.

Frijns, J. H. M., and Kate, J. H. t. (**1994**). ''A Model of Mylinated Nerve Fibers for Electrical Prosthesis Design,'' Med. Biol. Eng. Comput. **32**(4), 391–398.

Frijns, J. H. M., Mooij, J., and Kate, J. H. t. (**1994**). ''A Quantitative Approach to Modeling Mammalian Myelinated Nerve Fibers for Electrical Prosthesis Design,'' IEEE Trans. Biomed. Eng. **41**(6), 556–566.

Frijns, J. H. M., Snoo, S. L. d., and Kate, J. H. t. (**1996**). ''Spatial Selectivity in a Rotationally Symmetric Model of the Electrically Stimulated Cochlea,'' Hear. Res. **95**, 33–48.

Frijns, J. H. M., Snoo, S. L. d., and Schoonhoven, R. (**1995**). ''Potential Distributions and Neural Excitation Patterns in a Rotationally Symmetric Model of the Electrically Stimulated Cochlea,'' Hear. Res. **87**, 170–186.

Hartmann, R., Topp, G., and Klinke, R. (**1984**). ''Discharge Patterns of Cat Primary Auditory Fibers with Electrical Stimulation of the Cochlea,'' Hear. Res. **13**, 47–62.

Hinojosa, R., Green, J. D., Jr., and Marion, M. S., (**1991**). ''Ganglion cell populations in labyrinthitis ossificans,'' Am. J. Otol. **12**(Suppl. 3–7), 18–21.

Hinojosa, R., and Marion, M. (**1983**). ''Histopathology of profound sensorineural deafness,'' Ann. N. Y. Acad. Sci. **405**, 459–484.

Ketten, D. R., Skinner, M. W., Wang, G., Vannier, M. W., Gates, G. A., Neely, J. G. (**1998**). ''*In Vivo* measures of cochlear length and insertion depth of nucleus cochlear implant electrode arrays,'' Ann. Otol. Rhinol. Laryngol. **107**, 1–16.

Leake, P. A., and Hradek, G. T. (**1988**). ''Cochlear Pathology of Long Term Neomycin Induced Deafness in Cats,'' Hear. Res. **33**, 11–34.

Leake, P. A., Snyder, R. L., and Schreiner, C. E. (**1987**). ''Cochlear Pathology of Sensorineural Deafness in Cats: Coadministration of Kanamycin and Aminooxyacetic Acid,'' Ann. Otol. Rhinol. Laryngol. **96**(Suppl. 128), 48–50.

Leake, P. A., Hradek, G. T., Rebscher, S. J., and Snyder, R. L. (**1991**). ''Chronic intracochlear stimulation induces selective survival of spiral ganglion neurons in neonatally deafened cats,'' Hear. Res. **54**, 251–271.

Leake, P. A., Snyder, R. L., Hradek, G. T., and Rebscher, S. J. (**1992**). ''Chronic intracochlear electrical stimulation in neonatally deafened cats: Effects of intensity and stimulating electrode location,'' Hear. Res. **64**, 99–117.

Leake, P. A., Snyder, R. L., Hradek, G. T., and Rebscher, S. J. (**1995**). ''Consequences of Chronic Extracochlear Electrical Stimulation in Neonatally Deafened Cats,'' Hear. Res. **82**, 65–80.

Leake, P. A., Snyder, R. L., Rebscher, S. J., Moore, C. M., and Vollmer, M. (**2000**). ''Plasticity in central representation in the inferior colliculus induced by chronic single vs two-channel electrical simulation by a cochlear implant after neonatal deafness,'' Hear. Res. **147**, 221–241.

Lusted, H. S., Shelton, C., and Simmons, F. B. (**1984**). ''Comparison of Electrode Sites in Electrical Stimulation of the Cochlea,'' Laryngoscope **94**, 878–881.

Marsh, R. R., Yamane, H., and Potsic, W. P. (**1981**). ''Effect of Site of Stimulation on the Guinea Pig's Electrically Evoked Brain Stem Response,'' Otolaryngol.-Head Neck Surg. **89**, 125–130.

McDermott, H. J., and McKay, C. M. (**1994**). ''Pitch ranking with nonsimultaneous dual-electrode electrical stimulation of the cochlea,'' J. Acoust. Soc. Am. **1**, 155–162.

McNeal, D. R. (**1976**). ''Analysis of a Model for Excitation of Myelinated Nerve,'' IEEE Trans. Biomed. Eng. **BME-23**(4), 329–337.

Merzenich, M. M., and White, M. (**1977**). ''Cochlear implant: The interface problem,'' in *Functional Electrical Stimulation*, edited by F. T. Hambrecht and J. B. Reswick (Decker, New York), pp. 321–339.

Merzenich, M. M., White, M., Vivion, M. C., Leake-Jones, P. A., and Walsh, S. M. (**1979**). ''Some Considerations of Multi-Channel Electrical Stimulation of Auditory Nerve in the Profoundly Deaf; Interfacing Electrode Arrays with the Auditory Nerve Array,'' Acta Oto-Laryngol. **87**, 196–203.

Nadol, J. B., Jr., 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**(6), 411–416.

Osberger, M., and Fisher, L. (**1999**). ''SAS-CIS preference study in postlingually deafened adults implanted with the Clarion cochlear implant,'' Ann. Otol. Rhinol. Laryngol. **177**(Suppl.), 74–79.

Parkins, C. W., and Colombo, J. (**1987**). ''Auditory-Nerve Single-Neuron Thresholds to Electrical Stimulation from Scala Tympani Electrodes,'' Hear. Res. **31**, 267–286.

Pfingst, B. E., and Sutton, D. (**1983**). ''Relation of cochlear implant function to histopathology in monkeys,'' in *Cochlear Prostheses: An International Symposium*, edited by C. W. Parkins and S. W. Anderson, Ann. N.Y. Acad. Sci. **405**, 224–239.

Pfingst, B. E., and Sutton, D. (**1984**). ''Relation of psychophysical thresholds for electrical stimuli to auditory nerve survival: Summary of results from 18 scala tympani implants,'' Abstr. Assoc. Res. Otolaryngol, p. 10.

Pfingst, B. E., Morris, D. J., and Miller, A. L. (**1995**). ''Effects of Electrode Configuration on Threshold Functions for Electrical Stimulation of the Cochlea,'' Hear. Res. **85**, 76–84.

Pfingst, B. E., Glass, I., Spelman, F. E., and Sutton, D. (**1985**). ''Psychophysical studies of cochlear implants in monkeys: Clinical implications'' in *Cochlear Implants*, edited by R. A. Schindler and M. M. Merzenich (Raven, New York), pp. 305–321.

Reilly, P. J., and Bauer, R. H. (**1987**). ''Application of a Neuroelectric Model to Electrocutaneous Sensory Sensitivity: Parameter Variation Study,'' IEEE Trans. Biomed. Eng. **BME-34**(**9**), 752–754.

Ryan, A. F., Miller, J. M., Wang, Z., and Woolf, N. K. (**1990**). ''Spatial Distribution of Neural Activity Evoked by Electrical Stimulation of the Cochlea,'' Hear. Res. **50**, 57–70.
Shannon, R. V. (**1983a**). ''Multichannel Electrical Stimulation of the Auditory Nerve in Man. I. Basic Psychophysics,'' Hear. Res. **11**, 157–189.
Shannon, R. V. (**1983b**). ''Multichannel Electrical Stimulation of the Auditory Nerve in Man. II. Channel Interaction,'' Hear. Res. **12**, 1–16.
Shepherd, R. K., Hatsushika, S., and Clark, G. M. (**1993**). ''Electrical Stimulation of the Audtitory Nerve: The Effect of Electrode Position on Neural Excitation,'' Hear. Res. **66**, 108–120.
Shepherd, R. K., and Javel, E. (**1997**). ''Electrical stimulation of the auditory nerve. I Correlation of physiological responses with cochlear status,'' Hear. Res. **108**, 112–144.
Smith, D. W., Finley, C. C., Honert, C. v. d., Olszyk, V. B., and Konrad, K. E. M. (**1994**). ''Behavioral and Electrophysiological Responses to Electrical Stimulation in the Cat I. Absolute Thresholds,'' Hear. Res. **81**, 1–10.
Snyder, R. L., Rebscher, S. J., Cao, K., Leake, P. A., and Kelly, K. (**1990**). ''Chronic Intracochlear Electrical Stimulation in the Neonatally Deafened Cat. I. Expansion of Central Representation,'' Hear. Res. **50**, 7–33.
Snyder, R. L., Rebscher, S. J., Leake, P. A., Kelly, K., and Cao, K. (**1991**). ''Chronic Intracochlear Electrical Stimulation in the Neonatally Deafened Cat. II: Temporal Properties of Neurons in the Inferior Colliculus,'' Hear. Res. **56**, 246–264.
Spelman, F. A., and Clopton, B. M. (**1987**). ''Measurement of the Specific Impedance of Bony Tissues in the Guinea Pig Cochlea,'' Abstracts of the 10th Midwinter Research Meeting of the Assoc. for Research in Otolaryngology, p. 41.
Strelioff, D. (**1973**). ''A computer simulation of the generation and Distribution of Cochlear Potentials,'' J. Acoust. Soc. Am. **54**, 620–629.
Suzuka, Y., and Schuknecht, H. F. (**1988**). ''Retrograde cochlear neuronal degeneration in human subjects,'' Acta Otolaryngol. **450**(Suppl.), 1–20.
Tong, Y. C., and Clark, G. M. (**1985**). ''Absolute identification of electric pulse rates and electrode positions by cochlear implant patients,'' J. Acoust. Soc. Am. **77**, 1881–1888.
Townshend, B., Cotter, N., VanCompernolle, D., and White, R. L. (**1987**). ''Pitch perception by cochlear implant subjects,'' J. Acoust. Soc. Am. **82**, 106–115.
van den Honert, C., and Stypulkowski, P. H. (**1987**). ''Single Fiber Mapping of Spatial Excitation Patterns in the Electrically Stimulated Auditory Nerve,'' Hear. Res. **29**, 195–206.
Vivion, M. C., Merzenich, M. M., Leake, P. A., White, M., and Silverman, M. (**1981**). ''Electrode Position and Excitation Patterns for a Model Cochlear Prosthesis,'' Ann. Otol. Rhinol. Laryngol. **90**(Suppl. 82), 19–20.
Vurek, L. S., White, M., Fong, M., and Walsh, S. M. (**1981**). ''Optoisolated stimulators used for electrically evoked BSER,'' Ann. Otol. Rhinol. Laryngol. **90**(**82**), 21–24.
Xu, S. A., Shepherd, R. K., Chen, Y., and Clark, G. M. (**1993**). ''Profound hearing loss in the cat following the single co-administration of kanamycin and ethacrynic acid,'' Hear. Res. **70**, 205–215.
Xu, S. A., McAnally, K. I., Xu, J., and Clark, G. M. (**1993**). ''Comparison of Half-Band and Full-Band Electrodes for Intracochlear Electrical Stimulation,'' Ann. Otol. Rhinol. Laryngol. **102**, 363–367.
Zwolan, T. A., Kileny, P. R., Ashbaugh, C., and Telian, S. (**1996**). ''Patient performance with the Cochlear Corporation '20+2' Implant: Bipolar vs. Monopolar Activation,'' Am. J. Otolaryngol. **17**, 717–723.

Emphasis of short-duration acoustic speech cues for cochlear implant users

Andrew E. Vandali[a)]
Cooperative Research Centre for Cochlear Implant and Hearing Aid Innovation, East Melbourne, 3002, Australia

(Received 6 June 2000; revised 3 November 2000; accepted 1 February 2001)

A new speech-coding strategy for cochlear implant users, called the transient emphasis spectral maxima (TESM), was developed to aid perception of short-duration transient cues in speech. Speech-perception scores using the TESM strategy were compared to scores using the spectral maxima sound processor (SMSP) strategy in a group of eight adult users of the Nucleus 22 cochlear implant system. Significant improvements in mean speech-perception scores for the group were obtained on CNC open-set monosyllabic word tests in quiet (SMSP: 53.6% TESM: 61.3%, $p<0.001$), and on MUSL open-set sentence tests in multitalker noise (SMSP: 64.9% TESM: 70.6%, $p<0.001$). Significant increases were also shown for consonant scores in the word test (SMSP: 75.1% TESM: 80.6%, $p<0.001$) and for vowel scores in the word test (SMSP: 83.1% TESM: 85.7%, $p<0.05$). Analysis of consonant perception results from the CNC word tests showed that perception of nasal, stop, and fricative consonant discrimination was most improved. Information transmission analysis indicated that place of articulation was most improved, although improvements were also evident for manner of articulation. The increases in discrimination were shown to be related to improved coding of short-duration acoustic cues, particularly those of low intensity. [DOI: 10.1121/1.1358300]

PACS numbers: 43.64.Me, 43.66.Ts, 43.71.Ky, 43.72.Ar [SPB]

I. INTRODUCTION

Cochlear implants are well-accepted prosthetic systems providing significant benefits to hearing-impaired adults and children (NIH Consensus Statement, 1995). Most present systems utilize spectral analysis techniques to estimate the signal waveform envelope within a number of adjacent filter bands. Electrodes corresponding in frequency to these bands are stimulated electrically at intensities derived from a mapping between the estimated signal intensities within each band and the user's measured threshold and maximum comfortable loudness levels for electrical stimulation. The spectral maxima sound processor (SMSP) strategy (McKay *et al.*, 1991; McDermott *et al.*, 1992; McKay and McDermott, 1993), the spectral peak (SPEAK) strategy (Skinner *et al.*, 1994; Whitford *et al.*, 1995), and the continuous interleaved sampling (CIS) strategy (Wilson *et al.*, 1991) are all examples of this. Speech perception with these strategies varies widely among individuals. However, some trends are common such as difficulty when listening in background noise (e.g., cafeteria or party noise) (Holden *et al.*, 1995; Kiefer *et al.*, 1996; Dorman *et al.*, 1998; Parkinson *et al.*, 1998; Fu and Shannon, 1999; Zeng and Galvin, 1999), and perception of some consonant and vowels can be difficult (McKay and McDermott, 1993; Whitford *et al.*, 1995). Studies exploring the perception of consonants using information transmission analysis have shown that place of articulation is generally the poorest perceived speech production feature (Dorman *et al.*, 1990; Tyler and Moore, 1992; McKay and McDermott, 1993). Manner of articulation is usually next-best perceived and voicing is typically best perceived.

Acoustic cues used to distinguish place of articulation for some consonants, e.g., unvoiced stops, can often be very short in duration and low in intensity. These cues include spectral properties of the burst of noise accompanying the release of the stop and the frequency loci and trajectories of formants adjacent to the consonant (Liberman, 1956; Stevens and Blumstein, 1978; Kewley-Port and Pisoni, 1983). The burst of noise accompanying a stop is typically only 5 to 50 ms in duration and energy tends to be spread widely over the spectrum with peaks at different frequency regions according to the place of articulation. The formant transitions accompanying articulation of the stop are also relatively short in duration (typically in the order of 5 to 50 ms). For cochlear implant users, these brief, low-intensity cues may not be well coded due to: (a) quantization of the signal in amplitude, frequency, and time; (b) limited dynamic range of the system; and (c) temporal and spectral masking. It was therefore hypothesized that for cochlear implant users, amplification (emphasis) of low-intensity, short-duration acoustic cues may provide improved perception of some consonants.

Some evidence to support this hypothesis can be drawn from a study by Kennedy *et al.* (1998), conducted with hearing-impaired listeners, which indicated that reduction of the consonant-to-vowel intensity ratio (by increasing the consonant intensity) could provide significant improvement to perception of some consonants in a vowel–consonant environment. However, the results indicated that the optimum ratio for best identification varied widely as a function of consonant, vowel environment, and subject. In addition, adjustment of the consonant intensity was carried out over the entire consonantal interval. In the study presented here, however, amplification of short-duration components of the signal is proposed, and thus results different from those of Kennedy *et al.* (1998) might be expected. In a related study by Sammeth *et al.* (1999) it was observed that for voiceless

[a)]Electronic mail: a.vandali@medoto.unimelb.edu.au

stop consonants, reduction of the consonant-to-vowel intensity ratio (by decreasing the vowel intensity) did not improve identification in general. This would imply that the benefits observed by Kennedy *et al.* (1998) were related to improving consonant audibility rather than reduction of the consonant-to-vowel intensity ratio.

In a previous study (Vandali *et al.*, 1995) a prototype of the transient emphasis spectral maxima (TESM) strategy was evaluated with a small group of Nucleus multichannel implant users. Speech comprehension with TESM was compared to that of the subject's standard speech-coding strategy (SMSP). The prototype TESM strategy was identical to the SMSP strategy but also included a multiband gain control that introduced additional gain during periods of rapid rise in the envelope signal of each band. For speech signals, these periods typically corresponded to the onset of consonant noise burst energy or the onset of a vowel formant. The additional gain applied (i.e., above that of the fixed gain of the system) was related to the rate of change of the envelope signal in each band. More precisely, it was adjusted in proportion to the gradient (calculated across a time interval of 20 ms) of the log magnitude of the slow-varying envelope signal (which was averaged over a 20-ms sliding window). In cases where the gradient was negative, no additional gain was applied. The strategy was shown to provide some benefit to open-set sentence perception in noise (multitalker babble) presented at a signal-to-noise ratio (SNR) of +5 dB. However, no benefit to word, consonant, or vowel perception was observed for open-set monosyllabic word tests in quiet.

Investigation of a similar strategy developed by Geurts and Wouters (1999) was conducted with a group of CIS users of the LAURA cochlear implant. A method of enhancing the speech envelope was proposed which was based on emulation of the rapid adaptation seen in the response of auditory nerves to sound stimuli. The enhanced envelope CIS (EECIS) strategy was similar to the CIS strategy but, like the prototype TESM strategy, it included a multiband gain control that operated on the magnitude of the envelope signals in each band. The envelope signals, termed ''fast envelope,'' in each band were low-pass filtered, providing a ''slow-envelope,'' which attenuated high-frequency components such as $F0$ modulation and introduced a delay between the fast and slow envelope signals. Both the fast and slow envelopes were instantaneously compressed as per the nonlinear function used in the subjects' clinical speech processor (CIS). The additional gain applied was then derived by taking the difference between the fast and slow envelopes. For cases of rapid rise in the envelope signal, the fast envelope exceeded the slow envelope and thus additional gain was applied. For negative differences, no additional gain was applied. While the derivation of the additional gain differed from that of the prototype TESM strategy, the overall effect of the algorithm on the envelope signals was similar in that gain was applied to the onset of envelope signals exhibiting a rapid rise in magnitude. The results of speech-perception tests comparing the EECIS strategy to the CIS strategy showed some improvements for closed-set consonant identification (for the vowel /a/ context only). However, open-set monosyllabic word tests showed no significant difference in consonant scores, which was consistent with the results obtained with the prototype TESM strategy. Vowel scores in the word tests were significantly higher, mainly due to better transmission of information about the first and second formant frequencies. Differences in results between the EECIS and prototype TESM strategies may be related in part to differences in the amount and time course of the additional gain applied by the two strategies.

It was speculated that the benefits to sentence perception in noise obtained with the prototype TESM strategy may have been due to greater enhancement of the onset of the target speech envelope compared to that of the background noise. In contrast, the lack of benefit to perception of words in quiet may have been related to factors such as: (a) degradation of the consonant-to-vowel intensity ratio for some consonants, particularly fricatives (Hedrick and Ohde, 1993; Kennedy *et al.*, 1998) and (b) overamplification of the onset of long-duration speech signals, such as vowel formants. It was thus proposed to revise the TESM strategy such that more emphasis was placed on rapid changes in the magnitude of envelope signals that accompanied short-duration speech signals (typically 5 to 50 ms in duration) in contrast to those of long-duration speech signals (typically 100 ms or greater). As a result, short-duration signals such as stop consonant noise bursts and rapid formant frequency transitions in the period adjacent to consonantal onset/offset, would be amplified by a greater amount than the onset of long-duration signals such as vowel formants and fricative noise energy. In doing this, adverse effects associated with amplification of the onset of long-duration signals might be minimized while still offering potential improvements to perception of short-duration signals.

II. PROCEDURE

A. Speech-processing strategies

1. SMSP strategy

The SMSP strategy is identical in many aspects to the commercially available SPEAK strategy. The SMSP strategy evaluated in this study was implemented in the Melbourne University portable research speech processor (P-DSP) (McDermott *et al.*, 1993). The strategy estimated the spectrum of incoming signals using a bank of 16 bandpass filters, in which the first eight had center frequencies spaced linearly over the frequency range 280 to 1780 Hz, and the remaining eight were approximately logarithmically spaced up to 6 kHz. An estimate of the envelope signal in each band was obtained by rectifying and low-pass filtering the band signals.[1] Up to six of these bands, termed ''spectral maxima,'' were then selected on the basis of highest signal level. Electrodes corresponding in frequency to the selected bands were stimulated in a tonotopic basal-to-apical order. The stimulus intensities of the activated electrodes were derived from a mapping between the selected signal intensities and the subject's measured threshold and maximum comfortable loudness levels for electrical stimulation. The above procedure was repeated periodically at an interval of 2.5 ms (i.e., 400-Hz analysis frequency) and thus up to six electrodes were activated at a group stimulation frequency of 400 pulses per second/channel.

2. TESM strategy

The TESM strategy differed from the SMSP strategy in that the envelope signals derived from each filter-bank band were modified by the transient emphasis algorithm prior to selection of the six spectral maxima. For any filter-bank band in which a rapid rise in the slow-varying envelope signal occurred, such as during the onset of consonant noise burst or vowel formant energy, additional gain (i.e., in addition to the gain of the system) was applied to the band. Furthermore, even greater gain was applied in cases where a rapid rise was followed by a rapid fall in the slow-varying envelope, which might correspond to a short-duration (i.e., 5- to 50-ms) consonant noise burst or a rapid shift in frequency across bands associated with a formant transition. In order to differentiate between these two cases, analysis of the slow-varying envelope signal in each band over time was required. For this reason, a running history, which spanned a period of 60 ms, at 2.5-ms intervals, of the envelope signals was maintained in a sliding buffer denoted $S_n(t)$ where the subscript n refers to the band number and t refers to time relative to the current analysis interval. This buffer was divided up into three consecutive 20-ms time windows and an estimate of the slow-varying envelope signal in each window was obtained by averaging across the terms in the window.[2] Averages from the three windows were therefore estimates of the past (E_p), current (E_c), and future (E_f) slow-varying envelope signal with reference to the midpoint of the buffer $S_n(t)$. The amount of additional gain applied was derived from a function of the slow-varying envelope estimates as per Eq. (1)

$$G=\frac{2\times E_c-2\times E_p-E_f}{E_c+E_p+E_f}. \tag{1}$$

The gain factor (G) for each band varied with the behavior of the slow-varying envelope signals such that: (a) Short-duration signals which consisted of a rapid rise followed by a rapid fall (over a time period of no longer than approximately 60 ms) in the slow-varying envelope signal would produce the greatest values of G. For these types of signals, G would be expected to range from approximately 0 to 2. (b) The onset of long-duration signals which consisted of a rapid rise followed by a relatively constant level in the envelope signal would produce lower levels of G which typically ranged from 0 to 0.5. (c) A relatively steady-state or slow-varying envelope signal would produce a negative value of G. (d) A relatively steady-state level followed by a rapid decrease in the envelope signal (i.e., cessation/offset of envelope energy) would produce small (less than approximately 0.1) or negative values of G. Because negative values of G were possible, the result of Eq. (1) was limited such that it could never fall below zero as per Eq. (2)

$$\text{if } (G<0), \quad \text{then } G=0. \tag{2}$$

Another important property of Eq. (1) was that the gain factor was related to a function of relative differences, rather than absolute levels, in the magnitude of the slow-varying envelope signal. For instance, short-duration peaks in the slow-varying envelope signal of different peak level but identical peak-to-valley ratios would be amplified by the same amount. When dealing with speech signals, intensity is typically defined on a log (dB) scale. It is thus convenient to view the gain factor in relation to the gradient of the log magnitude of the slow-varying envelope signal. Equation (1) can be expressed in terms of ratios of the slow-varying envelope signal estimates. Defining the backward magnitude ratio $R_b=E_c/E_p$ and the forward magnitude ratio $R_f=E_f/E_c$ gives Eq. (3)

$$G=\frac{2\times R_b-2-R_b\times R_f}{R_b+1+R_b\times R_f}. \tag{3}$$

The forward and backward magnitude ratios are equivalent to log-magnitude gradients and can be as defined as the difference between log-magnitude terms, i.e., $F_g=\log(E_f)-\log(E_c)$ and $B_g=\log(E_c)-\log(E_p)$, respectively. The relationship between gain factor and forward and backward log-magnitude gradients is shown in Fig. 1.

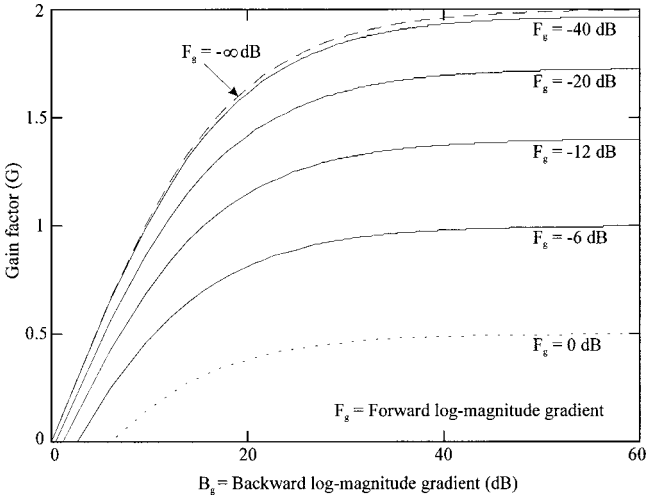


FIG. 1. Gain factor as a function of forward (F_g) and backward (B_g) log-magnitude gradients. Linear gain is plotted on the ordinate, and backward log-magnitude gradient (in dB) is plotted on the abscissa. The gain factor is plotted for different levels of the forward log-magnitude gradient in each of the curves. For any value of F_g, the gain factor reaches some maximum when B_g is approximately 40 dB. The maximum level is dependent on the level of F_g. For the case where F_g is 0 dB, as shown by the dotted line (i.e., at the end a period of rise for a long-duration signal where $E_f=E_c$) the maximum gain possible is 0.5. For the limiting case where F_g is infinitely steep, as shown by the dashed line (i.e., rapid fall in envelope signal where $E_f \ll E_c$), the maximum gain possible is 2.0. Note this limiting case is reached when F_g is approximately -40 dB.

The gain factors for each band (G_n), where n denotes the band number, were used to scale the original envelope signals $S_n(t)$ according to Eq. (4), where t_m refers to the midpoint of the buffer $S_n(t)$

$$S'_n(t_m)=S_n(t_m)\times(1+K_n\times G_n). \tag{4}$$

A gain modifier constant (K_n) was included for adjustment of the overall gain of the algorithm. In this study, $K_n=2$ for all n. During periods of little change in the envelope signal of any band, the gain factor (G_n) was equal to zero and thus $S'_n(t_m)=S_n(t_m)$. However, during periods of rapid change, G_n could range from 0 to 2 and thus a total of 0 to 14 dB of gain could be applied. Note that because the gain was applied at the midpoint of the envelope signals, an overall delay of approximately 30 ms between the time from input to

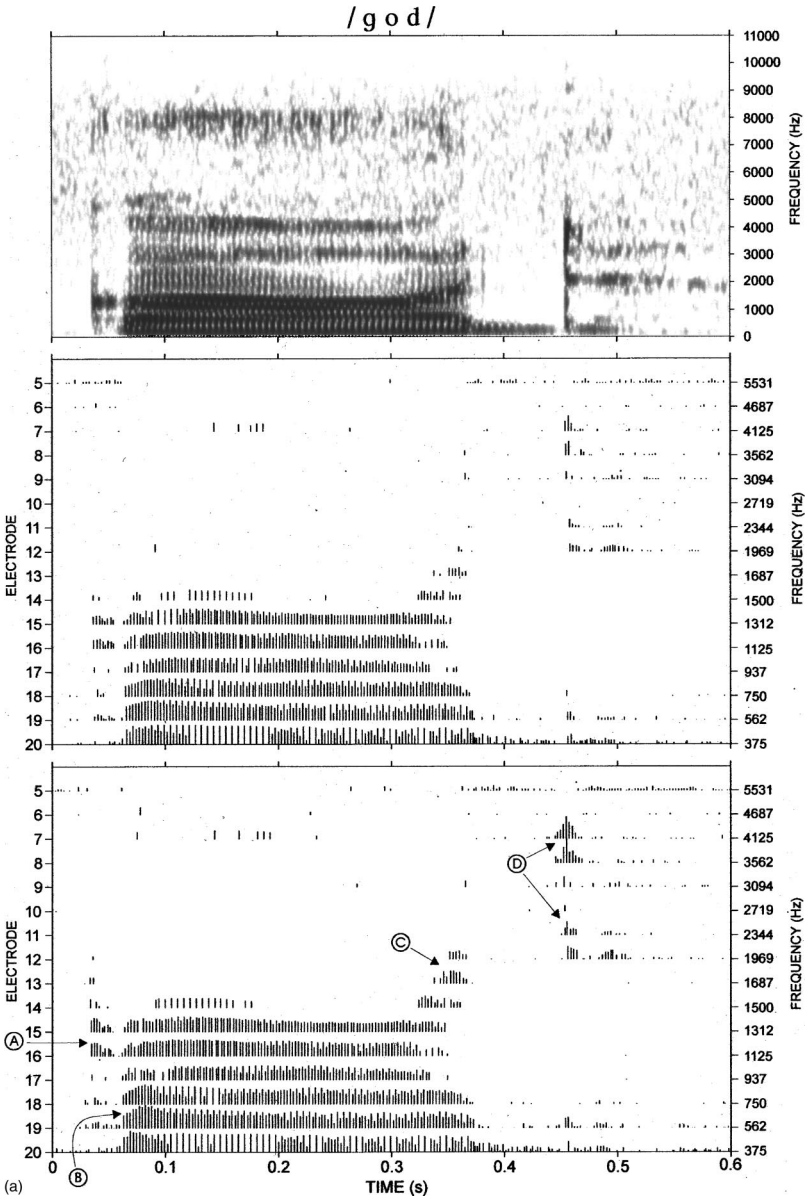


FIG. 2. Spectrogram and electrodogram plots for three speech tokens /god/, /mɪt/, and /fods/ are shown in panels (a), (b), and (c), respectively. For each token the spectrogram is displayed in the upper plot, electrodogram recordings for the SMSP and TESM strategies are shown in the middle and lower plots; respectively. Time is shown along the abscissa and frequency and/or electrode number along the ordinate. For the electrodograms, each stimulus pulse recorded from the output of the processor is shown as a vertical bar at the time and electrode position of the stimulus. The height of the vertical bar represents the stimulus intensity on a logarithmic scale (i.e., stimulus level in clinical units) where minimum amplitude corresponds to threshold, and maximum amplitude corresponds to comfortable level for electrical stimulation. For the token /god/, the effect of the TESM strategy can be seen in the stimulus intensity and number of electrodes representing the noise burst in the initial stop /g/ (point A). The onset of the formant energy in the vowel /o/ has also been emphasized slightly (point B). Stimuli representing the second formant ($F2$) transition from /o/ to the final stop /d/ are also higher in intensity (point C), as are those coding the noise burst in the final stop /d/ (point D). For /mɪt/, stimuli representing the onset of the nasal /m/ have been emphasized slightly (point E). Stimuli representing the $F2$ transition from /m/ to the vowel /ɪ/ are higher in intensity (point F) as are those coding the noise burst energy in the final stop /t/ (point G). For /fods/, the number and intensity of stimuli representing the fricative energy in /f/ have increased slightly (point H). The onset of the vowel /od/ has been emphasized slightly (point I), so too has the $F2$ transition from /od/ to /s/ (point J) and the fricative energy in /s/ (point K).

output of the transient emphasis algorithm was introduced. The modified envelope signals $S_n'(t)$ replaced the original envelope signals $S_n(t)$ derived from the filter bank and processing then continued as per the SMSP strategy. Because the transient emphasis algorithm was applied prior to selection of spectral maxima, bands containing low-intensity, short-duration signals were given a greater chance of selection due to their amplification. These bands might not normally be selected by the SMSP strategy because: (a) they fell below the mapped threshold level of the speech processing system; or (b) higher amplitude steady-state signals were selected in preference to them.

To illustrate the effect of the strategy on the coding of speech signals, stimulus output patterns, known as electrodograms, which plot stimulus intensity per band as a function of time, were recorded for each strategy and are shown in Fig. 2. Three speech tokens, ''god,'' ''mitt,'' and ''force,'' were used in these recordings and were taken from the word test material evaluated in this study.

B. Subjects

Eight adult users of the Nucleus 22 cochlear implant system participated in this study. Subject details are summarized in Table I. At commencement of the study all subjects had been using their implant for at least 5 years. Subjects 1 and 2 were experienced users of the SMSP strategy as implemented in the P-DSP speech processor. Subjects 4 through 8 were experienced users of the SPEAK strategy implemented in the Nucleus Spectra 22 processor, while subject 3 was an experienced user of the multipeak strategy implemented in the Nucleus MSP speech processor (Dowell *et al.*, 1990; Skinner *et al.*, 1991).

C. Strategy fitting

All subjects were initially fitted with the (P-DSP) SMSP strategy. Parameters, such as the stimulation level for threshold and maximum comfortable loudness on each electrode, were adjusted so that subjects were comfortable with the

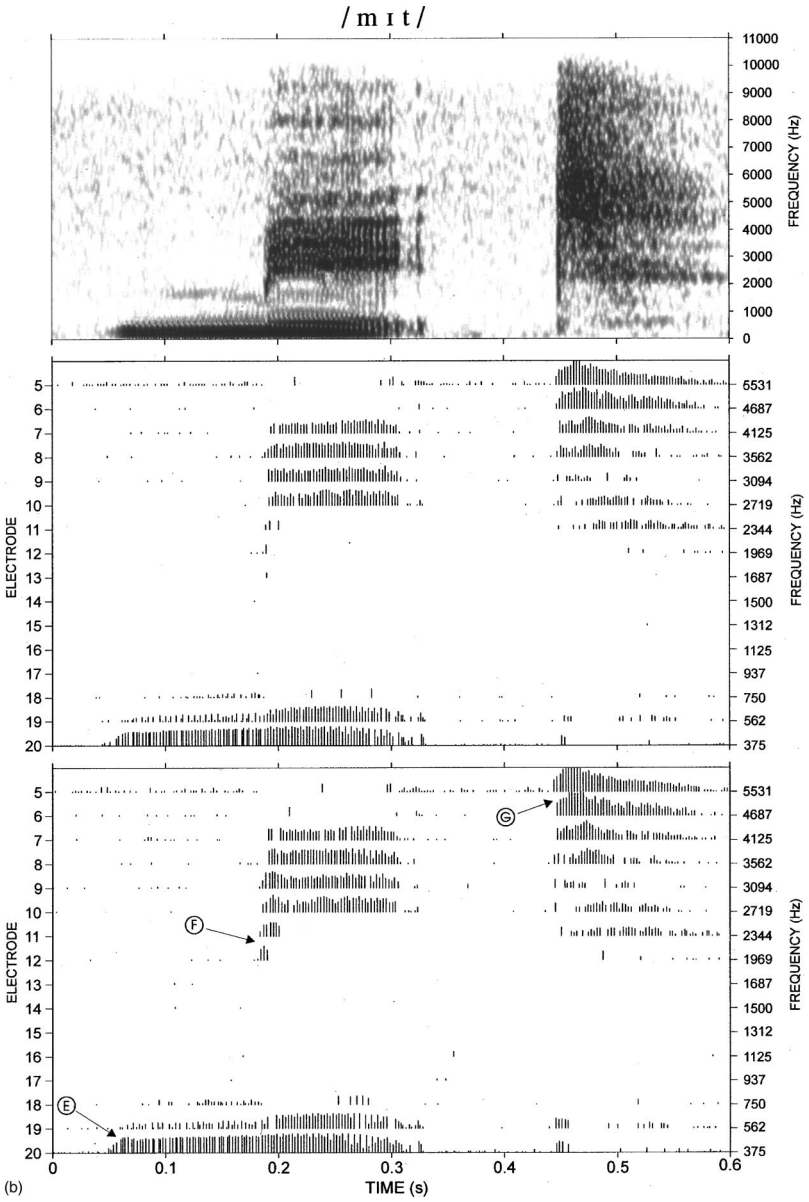


FIG. 2. (*Continued.*)

perception and quality of speech in everyday listening conditions. The base level of the loudness growth mapping function, which is used to convert input intensity to stimulus intensity, was also adjusted for individual subjects. The base level defines the minimum input level coded by the strategy and thus affects the mapped dynamic range of the system. Users of the P-DSP processor (SMSP) were accustomed to a base level of 2, which is equivalent to a mapped dynamic range of approximately 37.5 dB. However, users of the Spectra 22 (SPEAK) and MSP (multipeak) processors were accustomed to a base level of 4, which is equivalent to a mapped dynamic range of approximately 32 dB, ignoring the effects of the automatic-gain control. Thus, when fitting the SMSP strategy, the default base level of 2 was increased, if necessary, so that subjects were comfortable with the loudness of speech and environmental sounds (see Table II).

D. Study design

An AB (strategy) evaluation protocol was employed and the order of evaluation was alternated between subjects so as to allow for a randomized complete-block analysis of the results. Subjects 2, 3, 4, and 8 were evaluated with the SMSP strategy initially (A–B), while subjects 1, 5, 6, and 7 were evaluated with the TESM strategy initially (B–A); see Table II. For each subject, a training period using the first strategy was provided, the duration of which depended on their previous experience with the P-DSP speech processor and the SMSP strategy. Subjects 1 and 2, who were experienced users of the (P-DSP) SMSP strategy, were only provided with 2 weeks of training. Subjects 3 through 8 required a longer training period so as to adjust to the speech processor and strategy. During this training period subjects used the strategy on a take-home basis. Training sessions were provided weekly and speech material similar to the evaluation material was presented in each session. Perceptual scores were monitored over this period so as to help estimate when processor/strategy/task learning, if any, was complete. The training period continued until an approximate plateau in perceptual scores was observed (i.e., when scores were within approxi-

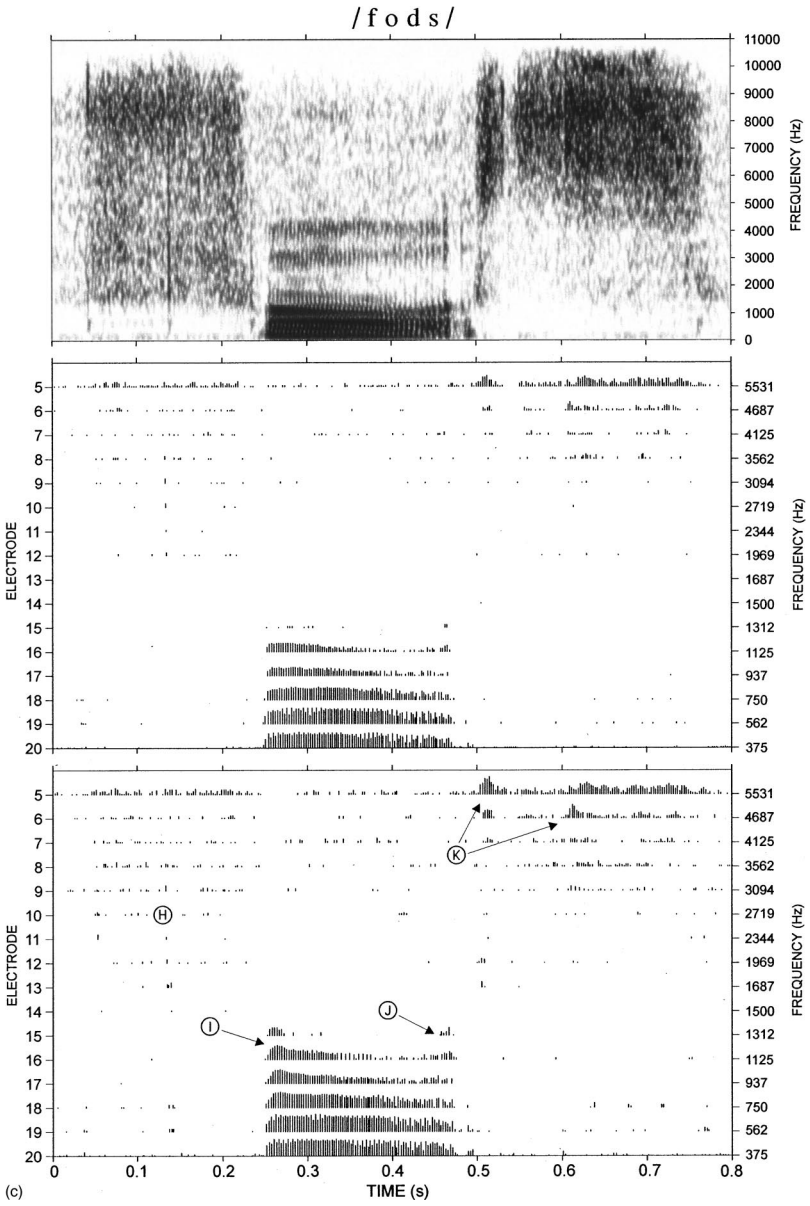


FIG. 2. (*Continued.*)

mately ±5% of those in the previous session). Following the training period, 4 consecutive weeks of data collection with the first strategy were carried out. Another training period with the second strategy, for a 2-week duration, then followed. A 2-week period was chosen because subjects were, at that point, familiar with the P-DSP speech processor and parameters of the speech-coding strategy, apart from those specific to the second speech-coding strategy. In addition, the initial training period for subjects 1 and 2 was only 2 weeks. Furthermore, in changing to the second strategy, subjects reported minimal perceptual differences between the two strategies.[3] Following the second training period, a 4-week evaluation period with the second strategy was conducted to complete the study.

TABLE I. Subject details: age in years at commencement of study, duration of severe-to-profound deafness, duration of implantation, etiology of deafness, and speech-coding strategy used prior to study.

Subject	Age (yrs)	Duration of severe-to-profound deafness	Years implanted	Etiology	Pre-evaluation strategy
1	45	5 months	9	Sudden-onset unknown	SMSP
2	71	35–40 years	6	Congenital progressive	SMSP
3	67	1 year	7	Otitis media	Multi-Peak
4	25	10 years	7	Maternal rubella	SPEAK
5	33	1 year	5	Congenital progressive	SPEAK
6	60	4 years	6	Menierés	SPEAK
7	51	1 year	12	Congenital progressive	SPEAK
8	45	1 year	13	Meningitis	SPEAK

TABLE II. Evaluation details: base level and equivalent input dynamic range for loudness growth function mapping, strategy evaluation order (A=SMSP, B=TESM), duration of initial training period in weeks, presentation level for CNC word tests, and signal-to-noise ratio for MUSL sentence in noise tests.

Subject	Loudness growth function base level (mapped dynamic range)	Evaluation order	Duration of initial training period (wks)	Presentation level of CNC words (dB SPL)	SNR for sentence in noises (dB)
1	2 (37.5 dB)	B–A	2	70	+10
2	2 (37.5 dB)	A–B	2	70	+10
3	3 (34 dB)	A–B	7	70	+10
4	3 (34 dB)	A–B	4	70	+10
5	4 (32 dB)	B–A	4	70	+10
6	4 (32 dB)	B–A	6	70	+10
7	4 (32 dB)	B–A	4	60	+5
8	4 (32 dB)	A–B	4	60	+5

E. Speech-perception evaluation

In each weekly evaluation session, two lists of monosyllabic CNC words (Peterson and Lehiste, 1962) were presented in quiet, and two lists of MUSL sentences [Melbourne University Sentence Lists derived from CUNY sentences (Boothroyd *et al.*, 1985)] were presented in background noise (Auditec, St. Louis, catalog no. C146-MT MULTITALKER). Thus, over the 4-week evaluation period, a total of eight lists of CNC words and eight lists of MUSL sentences were presented with each strategy. The speech-perception tests were conducted in a sound-attenuated room using prerecorded material played back via a loudspeaker at an azimuth of zero degrees. The presentation level of the speech material (including the multitalker noise) was measured using a Simpson hand-held sound-level meter (model no. 886). All measurements were made using dBA weighting with slow response metering for the sentence and multitalker noise material and fast response metering, which was visually averaged, for the word material. The CNC words were spoken by a female speaker, whereas the MUSL sentences were spoken by a male speaker.

Initially, only six subjects (1–6) participated in the study. For these subjects the presentation level of the target speech material (CNC words and MUSL sentences) was 70 dB SPL. The presentation level of the multitalker noise used in the MUSL sentence tests was 60 dB SPL, providing a SNR of +10 dB. However, ceiling effects were observed in results for subject 5, who was an excellent performer, which may have confounded possible strategy effects. It was thus decided to include two additional subjects (7 and 8) in the study, both of whom were good to excellent performers. For these subjects, ceiling effects in the word tests were avoided by reducing the presentation level of the CNC words to 60 dB SPL. Similarly for the sentence in noise tests, the noise level was increased to 65 dB SPL, thus reducing the SNR to +5 dB.

III. RESULTS

A. MUSL sentence in noise tests

Results for each subject for the MUSL sentence in noise tests are shown in Fig. 3(a). Seven of the eight subjects ob-

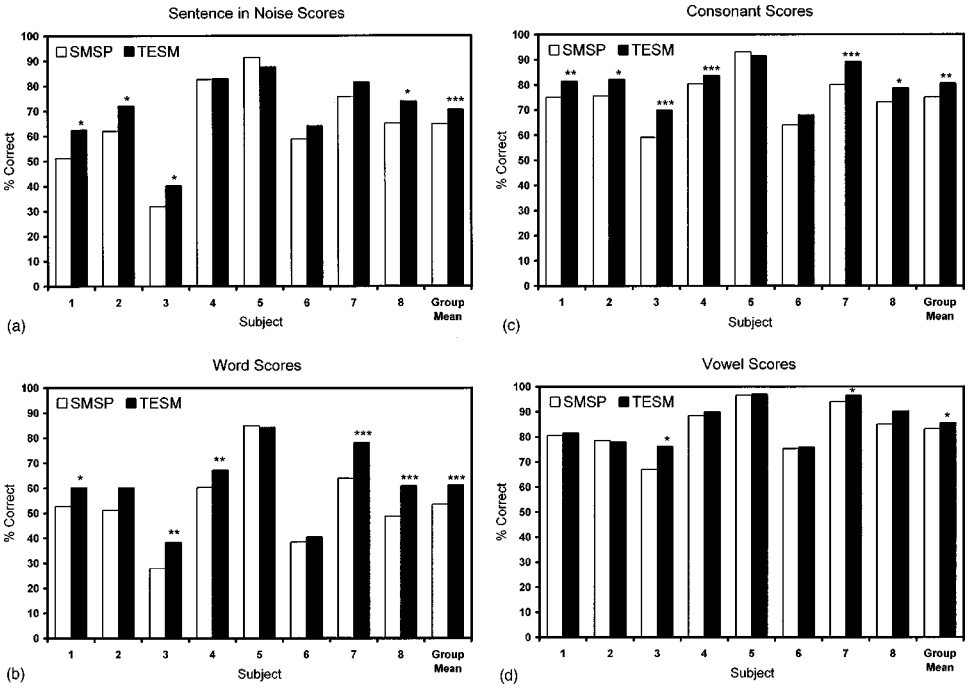


FIG. 3. Results for the speech perception tests. Panel (a): mean scores for the MUSL sentence tests in noise are shown for each subject and for pooled subject scores. Panels (b), (c), and (d): mean scores for the CNC word tests shown as percentage of whole words (b), consonants (initial and final positions combined) (c) and vowels (d) correct for each subject and pooled subject scores. Statistically significant differences (using ANOVA) are indicated by * at the 5% level, ** at the 1% level, and *** at the 0.1% level.

tained equal or better scores with the TESM strategy as compared to the SMSP strategy. For each subject, a two-way analysis of variance (ANOVA) was conducted to determine the effect of strategy and session. A significant effect of strategy was observed in four of the eight subjects. Mean scores for the TESM strategy were significantly higher than those of the SMSP strategy by 11.3 percentage points ($\Delta\%$)[4] for subject 1 [$F(1,8)=8.65$; $p<0.05$], 10.0 $\Delta\%$ for subject 2 [$F(1,8)=5.35$; $p<0.05$], 8.2 $\Delta\%$ for subject 3 [$F(1,8)=5.69$; $p<0.05$], and 8.8 $\Delta\%$ for subject 8 [$F(1,8)=5.84$; $p<0.05$]. No significant effect of session or interaction between strategy and session was observed for any of the subjects.

Pooling the scores of subjects 1 through 8, a two-way ANOVA was conducted for strategy and subject effects using subjects as a random factor within the design model. Group mean scores for the TESM strategy were significantly higher than those for the SMSP strategy by 5.7 $\Delta\%$ [$F(1,119)=19.8$; $p<0.001$]. A significant subject effect was also observed [$F(7,119)=84.7$; $p<0.001$], as would be expected from the variation in performance observed across subjects.

B. CNC word tests

Results for each subject for the CNC word tests are presented separately for percentage of whole words, consonants, and vowels correct in Fig. 3 (b), (c), and (d), respectively. As was the case for the sentence in noise tests, seven of the eight subjects obtained equal or better word scores with the TESM strategy as compared to the SMSP strategy. For each subject, a two-way ANOVA was conducted to determine the effect of strategy and session. With the exception of subjects 5 and 6, all subjects obtained significantly higher scores using the TESM strategy in one or more of the word test categories. For subject 1, word and consonant scores for the TESM strategy were significantly higher by 7.5 $\Delta\%$ [$F(1,8)=5.84$; $p<0.05$] and 6.5 $\Delta\%$ [$F(1,8)=14.86$; $p<0.01$], respectively. Consonant scores for subject 2 were significantly higher by 6.6 $\Delta\%$ [$F(1,8)=5.89$; $p<0.05$]. For subject 3, word scores were significantly higher by 10.5 $\Delta\%$ [$F(1,8)=11.31$; $p<0.01$], consonant scores by 10.9 $\Delta\%$ [$F(1,8)=34.56$; $p<0.001$], and vowel scores by 9.2 $\Delta\%$ [$F(1,8)=9.07$; $p<0.05$]. For subject 4, word and consonant scores were significantly higher by 7.0 $\Delta\%$ [$F(1,8)=12.65$; $p<0.01$] and 3.2 $\Delta\%$ [$F(1,8)=21.12$; $p<0.001$], respectively. Results for subject 7 were significantly higher for word, consonant, and vowel scores by 14.2 $\Delta\%$ [$F(1,8)=104.81$; $p<0.001$], 9.2 $\Delta\%$ [$F(1,8)=80.53$; $p<0.001$], and 2.5 $\Delta\%$ [$F(1,8)=8.33$; $p<0.05$], respectively. For subject 8, word and consonant scores were significantly higher by 12.3 $\Delta\%$ [$F(1,8)=24.75$; $p<0.001$] and 5.6 $\Delta\%$ [$F(1,8)=17.02$; $p<0.01$], respectively.

Of the eight subjects evaluated, only two (subjects 1, 4) exhibited session effects in word test results. It was found that these effects were the result of either obtaining higher scores in the first session or lower scores in the last session with each strategy and showed no evidence of positive learning. For subject 1, a significant effect of session was observed in the consonant scores [$F(3,8)=5.61$; $p<0.05$]. *Post hoc* analysis for the session effect using a pairwise Tukey test showed that scores in session 1 were significantly higher than session 4 by 9.2 $\Delta\%$ ($p<0.05$). A similar result was observed for subject 4. For word and consonant scores, a significant effect of session was observed [$F(3,8)=4.92$; $p<0.05$] and [$F(3,8)=12.04$; $p<0.01$], respectively. *Post hoc* analysis for the word scores revealed that scores in session 1 were significantly higher than those in session 4 by 5.2 $\Delta\%$ ($p<0.05$). Similarly for the consonant results, scores in session 1 were significantly higher than those in sessions 2, 3 and 4 by 4.2 $\Delta\%$($p<0.05$), 5.0 $\Delta\%$($p<0.01$) and 5.2 $\Delta\%$($p<0.01$), respectively.

Subject 7 was the only subject to exhibit interactions between strategy and session in the word tests. In this case the interaction was the result of a strategy effect in one or more of the initial sessions but not in the final sessions. For word, consonant, and vowel scores, a significant interaction between strategy and session was observed [$F(3,8)=9.15$; $p<0.01$], [$F(3,8)=4.41$; $p<0.05$], and [$F(3,8)=5.00$; $p<0.01$], respectively. *Post hoc* analysis for the word scores showed a significant effect of strategy in sessions 1 to 3 (i.e., TESM significantly higher than SMSP) but not in session 4. A similar finding was observed in the consonant scores, where a strategy effect was observed in sessions 1 to 3 but not in session 4. The vowel scores showed a significant strategy effect in session 1 but not in sessions 2, 3, and 4.

The word test results were pooled and a two-way ANOVA was conducted for strategy and subject effects using subjects as a random factor within the design model. Group mean scores for the TESM strategy were significantly higher than those of the SMSP strategy for words by 7.7 $\Delta\%$ [$F(1,119)=53.6$; $p<0.001$], for consonants by 5.5 $\Delta\%$ [$F(1,119)=57.0$; $p<0.001$], and for vowels by 2.6 $\Delta\%$ [$F(1,119)=6.7$; $p<0.05$]. In all cases subject effects were significant [$F(7,119)=121.1$; $p<0.001$], [$F(7,119)=78.6$; $p<0.001$], and [$F(7,119)=45.4$; $p<0.001$], respectively, as expected due to the large variation in scores among subjects.

C. Consonant confusions and information transmission analysis

For the pooled CNC word test results, scores for consonants were analyzed separately for perception of initial and final consonant positions and grouped according to their manner of articulation (i.e., stop, affricative, fricative, nasal, and semivowel); see Fig. 4. The mean data indicate that increased consonant scores using the TESM strategy could be accounted for mainly by improved perception of nasal, fricative, and stop consonants. Confusion matrices for the pooled consonant scores were also constructed (see Fig. 5), which were subjected to information transmission analysis (Miller and Nicely, 1955), the results are displayed in Fig. 6. The consonants were classified according to the features: manner of articulation, place of articulation, and voicing, as indicated in Table III. The analysis indicates that transfer of information related to place of articulation was most improved, particularly for consonants in the final position. Some improvement to transfer of information related to manner of articulation was also observed. These findings are consistent

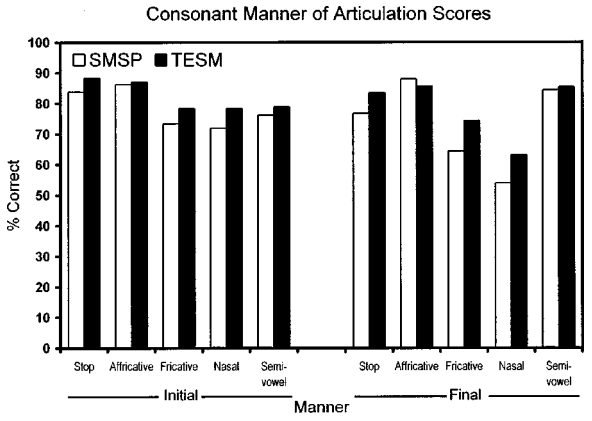


FIG. 4. Consonant perception scores for CNC word tests grouped according to initial and final consonant positions and according to the manner of articulation classifications: stops, affricatives, fricatives, nasals, and semivowels.

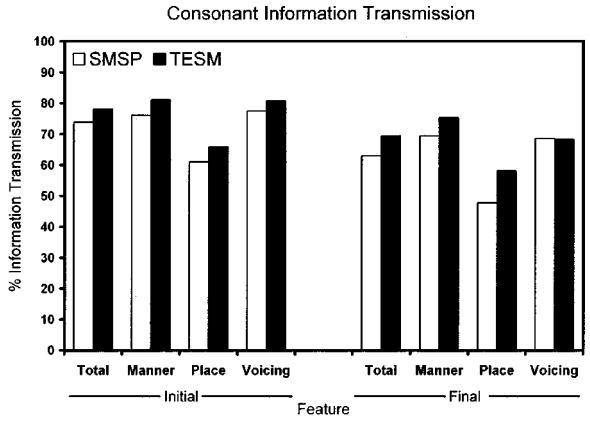


FIG. 6. Results of information transmission analysis for consonant perception in the CNC word tests. For initial and final consonant positions, the graphs show the total information transfer, and the transfer of information about the features: manner and place of articulation, and voicing.

with those reported in the literature (Dorman *et al.*, 1990; Tyler and Moore, 1992; McKay and McDermott, 1993).

Consonant confusion matrices for each of the subjects were also constructed, and together with those for the pooled data, trends in confusions common to many of the subjects were sought. Due to differences in the degree of confusions within subjects and variations in confusions across subjects, the following criteria were used in this analysis. First, for the pooled data only confusions of approximately 10% or greater were considered. Second, of the considered confusions, only those that showed at least a 10% difference between strategies (in individual scores) for approximately half the subjects or more where accepted. The number of subjects used for this criterion was in some cases reduced to three if at least two subjects not satisfying the criterion displayed ceiling effects. These criteria were adopted to restrict the analysis of

(a)

	p	t	k	b	d	g	tʃ	dʒ	f	s	ʃ	θ	v	h	m	n	w	j	l	r	[-]	#
p	136	3	20	3	0	1	0	1	2	0	0	1	0	23	0	0	0	0	0	0	2	192
t	1	184	4	0	0	0	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	192
k	14	5	169	0	0	0	1	0	1	0	0	0	0	2	0	0	0	0	0	0	0	192
b	3	0	1	158	1	3	0	0	8	0	0	2	1	1	0	0	0	0	0	0	14	192
d	0	1	0	0	176	11	0	3	0	0	0	0	0	0	0	0	0	0	0	0	1	192
g	1	0	7	4	15	89	0	1	3	0	0	1	0	2	0	0	0	0	0	0	5	128
tʃ	0	13	3	0	0	0	109	0	0	0	3	0	0	0	0	0	0	0	0	0	0	128
dʒ	0	5	0	0	2	0	8	112	0	0	1	0	0	0	0	0	0	0	0	0	0	128
f	0	0	1	9	1	0	0	0	93	7	1	2	4	5	0	1	0	0	0	0	4	128
s	1	0	0	3	1	0	0	0	24	149	2	10	0	0	0	1	0	0	0	0	1	192
ʃ	0	0	0	0	0	0	4	1	0	0	123	0	0	0	0	0	0	0	0	0	0	128
θ	1	2	0	3	2	0	0	0	10	11	0	34	0	1	0	0	0	0	0	0	0	64
v	1	0	1	16	3	0	0	1	15	0	0	0	16	2	1	3	0	0	3	1	1	64
h	8	1	5	5	0	0	0	0	8	1	0	3	0	149	0	0	3	0	0	3	6	192
m	0	0	0	2	0	0	0	0	0	0	1	0	0	0	0	123	45	2	0	9	9	1	192
n	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	15	107	0	0	4	1	1	128
w	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	177	0	3	11	0	192
j	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	4	0	51	1	0	6	64
l	1	0	0	0	0	0	0	0	0	0	0	0	0	0	31	52	2	2	149	18	1	256
r	0	0	0	2	0	0	0	0	0	0	0	0	0	0	8	5	19	0	12	208	2	256

(c)

	p	t	k	b	d	g	tʃ	dʒ	f	s	ʃ	θ	v	z	m	n	ŋ	l	r	[-]	#
p	124	12	42	4	2	4	1	0	1	0	0	0	0	0	1	0	1	0	0	0	192
t	3	348	13	0	8	1	4	0	1	2	0	1	0	0	0	0	0	1	0	2	384
k	12	16	217	0	1	9	0	0	0	0	0	0	0	0	0	0	0	0	0	1	256
b	7	1	3	33	4	9	1	0	0	0	0	0	0	0	3	1	0	0	0	2	64
d	4	12	5	2	193	9	0	0	2	0	0	0	6	1	3	8	0	0	0	11	256
g	5	3	28	6	11	68	0	0	1	0	0	0	0	0	0	1	0	1	0	4	128
tʃ	0	11	0	0	0	0	115	2	0	0	0	0	0	0	0	0	0	0	0	0	128
dʒ	0	6	0	0	2	0	1	54	0	0	0	0	0	0	0	0	1	0	0	64	
f	5	8	4	1	2	1	3	1	80	13	0	2	2	1	2	2	0	0	0	1	128
s	0	8	4	0	1	1	3	0	30	135	0	5	0	1	1	1	0	0	0	2	192
ʃ	0	0	0	0	0	0	2	0	0	2	60	0	0	0	0	0	0	0	0	0	64
θ	4	7	4	0	2	0	0	0	5	6	0	31	0	0	0	2	0	0	0	3	64
v	3	2	1	3	7	1	0	0	2	0	0	1	23	5	0	7	0	1	0	8	64
z	0	2	1	0	2	1	0	1	3	5	0	3	11	83	0	6	0	1	0	9	128
m	0	0	1	3	10	3	0	0	1	0	0	0	5	3	80	59	7	2	2	16	192
n	1	2	3	0	18	8	0	1	3	1	0	0	5	6	27	205	10	6	2	22	320
ŋ	0	0	0	0	3	0	0	0	1	0	0	0	0	0	5	29	26	0	0	0	64
l	0	1	3	0	4	1	0	0	2	0	0	0	1	2	9	10	4	250	14	19	320
r	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	4	182	4	192

(b)

	p	t	k	b	d	g	tʃ	dʒ	f	s	ʃ	θ	v	h	m	n	w	j	l	r	[-]	#
p	156	2	17	0	0	0	1	0	1	0	0	0	0	15	0	0	0	0	0	0	1	192
t	1	184	3	0	1	0	2	0	0	0	0	0	0	0	0	0	0	0	1	0	0	192
k	5	5	174	0	1	1	1	0	0	0	0	0	0	4	0	1	0	0	0	0	0	192
b	3	0	1	156	1	2	0	0	11	0	0	3	1	0	1	0	1	0	0	0	12	192
d	0	0	0	1	187	3	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	192
g	1	1	4	5	12	104	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	128
tʃ	1	10	1	0	0	1	110	3	0	2	0	0	0	0	0	0	0	0	0	0	0	128
dʒ	0	3	0	0	3	0	9	113	0	0	0	0	0	0	0	0	0	0	0	0	0	128
f	0	0	0	2	0	0	0	0	101	16	0	4	1	2	0	0	0	0	0	0	2	128
s	0	0	0	1	0	0	0	0	20	164	1	4	0	1	0	0	0	0	0	0	1	192
ʃ	0	0	0	0	0	0	1	0	0	4	123	0	0	0	0	0	0	0	0	0	0	128
θ	0	0	0	2	1	1	0	0	12	17	0	31	0	0	0	0	0	0	0	0	0	64
v	1	1	0	11	1	0	0	0	11	1	0	3	21	3	2	1	0	1	1	4	2	64
h	7	1	0	5	0	1	1	0	2	1	1	1	0	162	0	0	7	1	2	0	0	192
m	0	0	0	0	0	0	0	0	0	0	0	0	0	1	140	42	0	0	6	3	0	192
n	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14	111	0	0	3	0	0	128
w	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	1	172	0	4	13	0	192
j	0	0	0	0	0	0	0	1	0	0	0	0	0	2	0	1	0	58	0	0	2	64
l	0	0	0	1	0	0	0	0	0	0	0	0	0	0	23	42	1	6	174	9	0	256
r	0	0	0	1	0	0	0	0	0	0	0	0	0	0	9	4	24	0	13	202	3	256

(d)

	p	t	k	b	d	g	tʃ	dʒ	f	s	ʃ	θ	v	z	m	n	ŋ	l	r	[-]	#
p	144	7	28	4	0	7	0	0	1	0	0	0	0	0	0	0	0	0	0	1	192
t	2	361	11	1	6	2	0	0	0	0	0	0	0	1	0	0	0	0	0	0	384
k	8	7	232	0	1	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	256
b	2	0	6	33	6	10	0	0	1	0	0	0	0	0	4	2	0	0	0	0	64
d	6	12	12	2	200	9	0	0	0	2	0	1	2	0	0	2	1	1	0	6	256
g	3	2	19	0	4	99	0	0	1	0	0	0	0	0	0	0	0	0	0	0	128
tʃ	0	7	0	0	1	0	116	4	0	0	0	0	0	0	0	0	0	0	0	0	128
dʒ	0	5	1	0	1	1	3	52	0	0	1	0	0	0	0	0	0	0	0	0	64
f	2	3	4	0	1	0	0	0	90	22	2	1	1	1	0	0	0	0	0	1	128
s	0	3	0	0	0	0	2	0	16	161	1	1	4	1	1	0	0	0	0	2	192
ʃ	0	0	0	0	0	0	0	0	0	2	62	0	0	0	0	0	0	0	0	0	64
θ	2	7	4	0	4	0	0	0	3	7	0	36	0	0	0	1	0	0	0	0	64
v	1	3	3	0	5	1	0	0	2	0	0	2	35	2	3	1	0	0	0	6	64
z	0	2	0	0	6	2	0	0	3	10	0	2	7	92	1	1	0	0	0	2	128
m	1	1	0	1	8	2	0	0	2	0	0	0	7	3	117	35	2	1	3	9	192
n	3	4	3	1	28	6	0	1	0	1	0	2	3	4	29	207	9	2	2	15	320
ŋ	0	0	2	0	2	1	0	0	1	1	0	0	0	1	3	13	40	0	0	0	64
l	0	1	5	0	5	5	0	0	0	0	0	0	2	0	5	7	4	265	7	14	320
r	0	3	0	0	0	0	0	0	0	1	0	2	0	0	0	1	0	11	172	2	192

FIG. 5. Consonant confusion matrices from the CNC words for pooled subject data. Confusions for consonants in the initial position are shown in panel (a) for the SMSP strategy and panel (b) for the TESM strategy. Similarly, confusions for consonants in the final position are shown in panel (c) for the SMSP strategy and panel (d) for the TESM strategy. The presented stimuli are indicated by the consonants listed vertically in the first column. The subject responses are listed horizontally across the top of the table. The response column [−] denotes absent consonants in the response words. The last response column # indicates the number of presentations of each stimuli. Confusions of greater than 10% (according to the first criteria) have been enclosed in a rectangular box. Confusions that decreased (according to the second criteria) have been shaded in.

TABLE III. Classification of articulation features used in information transmission analysis for consonants in CNC word tests.

Feature	p	b	t	d	k	g	tʃ	dʒ	f	v	θ	s	z	ʃ	h	m	n	ŋ	w	j	l	r
Manner	1	1	1	1	1	1	2	2	3	3	3	3	3	3	3	4	4	4	5	5	5	5
Place	1	1	4	4	6	6	5	5	2	2	3	4	4	5	7	1	4	6	1	5	4	4
Voicing	1	2	1	2	1	2	1	2	1	2	1	1	2	1	1	2	2	2	2	2	2	2

confusions to a manageable number that was representative of confusions most common to the group of subjects and not biased by outlying results of individuals.

For the pooled data, improved results (i.e., decreases in confusions) using the TESM strategy were mainly observed for confusions in which place of articulation errors for consonants in the final position occurred. This was consistent with the findings of the information transmission analysis. By far, the largest improvements were observed for nasal consonants in which a 12.5 Δ% decrease was observed for /m/ perceived as /n/, and a 25.0 Δ% decrease for /ŋ/ perceived as /n/. For stop consonants a 7.3 Δ% decrease was observed for /p/ perceived as /k/, and for fricatives a 7.3 Δ% decrease was observed for /s/ perceived as /f/. Improved transfer of manner of articulation using the TESM strategy was also observed. A 7.8 Δ% decrease was observed for /v/ perceived as /b/ in the initial position, and a 9.4 Δ% decrease was observed for /v/ perceived as /n/ in the final position.

Some trends that were less consistent among the pool of subjects than those discussed above (i.e., common to only three of the subjects) were also observed. For consonants in the initial position, a 4.2 Δ% decrease for /p/ perceived as /h/, and a 3.9 Δ% decrease for /l/ perceived as /n/ were observed. For consonants in the final position, a 7.8 Δ% decrease for /b/ perceived as /p/, and a 7.0 Δ% decrease for /g/ perceived as /k/ were observed. Only one case was found in which decreased performance (i.e., increased confusions) using the TESM strategy was observed. In this case a 9.4 Δ% increase in /θ/ perceived as /s/ in the initial position was observed in results for three of the subjects. It should also be noted that while a 7.0 Δ% increase in confusions for /f/ perceived as /s/ in both the initial and final positions was observed, the result was only common to two of the subjects.

A number of cases were observed in which confusions of 10% or greater in the pooled data occurred for which there was little difference in mean scores between strategies. Most of these confusions involved errors in place of articulation for consonants in the initial position such as /m/ perceived as /n/, /θ/ as /f/, /n/ as /m/, /g/ as /d/, /p/ as /k/, and /s/ as /f/. For place confusions in the final position, little effect of strategy was observed for /b/ perceived as /g/ and /θ/ as /s/. Manner confusions for consonants which showed little effect of strategy included /tʃ/ perceived as /t/, and /l/ perceived as /m/ in the initial position and /θ/ perceived as /t/, and /v/ perceived as /d/ in the final position.

IV. DISCUSSION

All subjects, with the exception of subjects 5 and 6, obtained a significant improvement in speech comprehension using the TESM strategy in comparison to the SMSP strategy for one or more of the perceptual tests. In the case of subject 5, ceiling effects in the data might account for the lack of a strategy effect in results. Perhaps a lower presentation level for the word tests and a lower SNR for the sentence in noise tests, as was administered to subjects 7 and 8, may have provided more sensitivity to strategy effects. For subject 6 mean scores for the TESM strategy were higher than those of the SMSP strategy in all test categories. For this subject, the lack of a strategy effect may have been due to the small difference in mean scores between strategies.

The improvements to consonant perception in the CNC word tests support the hypothesis of this study that amplification of low-intensity, short-duration acoustic cues may provide improved perception of some consonants. From the consonant confusion matrices it was found that perception of nasal, stop, and fricative consonants were most improved. Transfer of information related to place of articulation was most improved, although some improvement was also evident for manner of articulation. Detailed analysis of the consonant confusion matrices revealed a number of trends that were consistent across many of the subjects. While some of these trends may have been specific to the speaker used in this study, generalizations independent of the speaker can still be drawn from the results in most cases.

First, it was observed that for confusions involving errors in place of articulation, the largest improvements to perception using the TESM strategy were obtained for nasals in the final consonant position. Some improvement was also observed for stops in the final position and for fricatives in both positions. For nasal and stop consonants, place can be discriminated via the frequency spectrum of formants adjacent to the onset/offset of the consonant, and via the trajectory of the formants during the transition from/to the adjacent vowel. For stops, the spectrum of the noise burst energy may also be used as a cue to place. These cues, especially the dynamic ones such as rapid formant transitions, can be relatively short in duration. In addition, the spectral energy associated with stop consonant bursts and nasal consonant murmur in the region of the second formant frequency ($F2$) can be relatively low in intensity compared to that of vowels. Thus, given that a strategy effect was observed for confusions involving /m/, /ŋ/, and /p/ in the final position, it would be reasonable to conclude that better perceptual coding of aspects of these short-duration cues, especially those of low intensity, may have indeed been provided by the TESM strategy [see Fig. 2(a), points A, C, and D and Fig. 2(b), points F and G].

Second, in contrast to the above findings, place perception for nasal and stop consonants in the initial position showed little effect of strategy. For both strategies, nasal and stop consonant scores in the initial position were higher than those in the final position (see Fig. 4). Consonant perception

results from other studies with cochlear implant users (McKay and McDermott, 1993; Geurts and Wouters, 1999) have also shown similar trends, but in their case, the magnitude of the differences in scores between consonant positions was smaller. It may be the case that intensity differences between consonants in the initial and final position may account for these results. This was confirmed by examining spectrograms which showed a small but consistent decrease in intensity, of approximately 6 dB in many cases, over the duration of the word. Thus, we might conclude that for low-intensity (short-duration) consonant cues, little benefit of amplification is provided when the cues are already sufficiently coded (i.e., loud enough), as was observed for nasal and stop consonants in the initial position. This argument is also consistent with overall scores for consonants in the initial and final position.

Third, for the fricative consonants, improvements to information transfer of both place and manner were observed for cases in which the intensity of fricative noise energy was very low, i.e., near the minimum (mapped) input level of the system. For instance, the fricative energy in /v/ and /f / both have a relatively broad spectrum and are very low in intensity compared to other fricatives. Depending on the signal level and particular speech utterance, it was observed in electrodograms that the fricative energy was often poorly coded by the SMSP strategy (i.e., very few stimuli at near-threshold level of stimulation). The TESM strategy provided more stimuli at higher intensities for coding of the onset, and random fluctuations in the envelope of the fricative noise energy [see Fig. 2(c), point H]. The improvements to fricative confusions such as /v/ perceived as /n/ (final position), or /v/ perceived as /b/ (initial position) might thus be accounted for by better coding of spectral energy in the fricative. In addition, for the case where /v/ was perceived as /b/, which is a stop consonant, the duration of the fricative energy, which is a cue to its manner, may also have been coded more reliably.

Another common fricative confusion that improved was /s/ perceived as /f/. For the female speaker used in this study, the fricative noise energy in /s/ had a spectrum which extends upward in frequency from approximately 4 kHz and had a spectral peak in the vicinity of 8.5 kHz [see Fig. 2(c)]. The frequency response of the speech processor was limited to an upper frequency of 6 kHz, due to the sampling rate of the system and the response of the ear-level microphone (Knowles XL-9924 directional two-port microphone). Thus, for the fricative energy in /s/, little spectral energy may have been coded by the SMSP strategy for cases in which the intensity of the utterance was low. Given this to be the case, spectral/amplitude contrast between the fricative energy in /s/ and that in /f/ may have been poor. Amplification of the onset and random fluctuations in the respective fricative noise energy by the TESM strategy may have provided better spectral contrast and thus fewer confusions. This result, however, is rather specific to the speaker used in the study and we might expect less confusions for speakers in which the spectral peak in the fricative energy of /s/ was lower in frequency or if the upper frequency range of the system was higher.

As a consequence of amplifying parts of the fricative noise burst, a small decrease in perception of place was also observed for some consonants. This may have been due to disruption of relative amplitude differences between the intensities of the noise energy and the accompanying vowel formants (Hedrick and Ohde, 1993; Kennedy *et al.*, 1998). However, this was observed in only a few cases (/θ/ perceived as /s/ for three subjects and /f/ perceived as /s/ for two subjects), and in general the increases to fricative perception outweighed the decreases. Regardless, it highlights the fact that excessive amplification of fricative noise energy may be detrimental. On the other hand, for short-duration cues, such as those that accompany nasal and stop consonants, too little gain may not provide the benefits to perception that were observed in this study. A balance between the amount of gain applied in these two cases may be required to provide improved perception of nasal and stop consonants without adversely affecting fricative perception. In this study the gain applied by the TESM strategy to short-duration acoustic signals (such as for nasals and stops) ranged from approximately 0 to 14 dB, in contrast to that applied to long-duration signals (such as fricative noise energy) which ranged from approximately 0 to 6 dB. While improved results were observed with little negative effects, this ratio (or gain rule) may not be optimal and may be investigated in future studies.

Finally, it was surprising to see a small improvement in perception of some semivowels, which have no rapid formant transitions or noise burst cues. In these cases it would appear that emphasis of the onset of poorly coded formant frequencies may have provided an advantage. For instance, /l/ was often perceived as the nasal /n/. In some cases, the intensity of $F2$ in /l/ can be quite low and poorly coded by the SMSP strategy, thus reducing the contrast between it and /n/, which has virtually no $F2$ energy due to a spectral null in the vicinity of $F2$. It is possible that amplification of the onset of $F2$ in /l/ may have provided additional contrast between the two consonants. In addition, the first formant frequency ($F1$) (which can be a cue to nasality for /n/) for these two consonants did not differ by much for the particular speaker used in this study. It was typically observed that $F1$ was coded in the two most apical bands and differed only slightly in intensity in the most basal of these two bands. Amplification of the onset of $F1$ may have assisted correct identification of its frequency and thus contributed to fewer confusions.

Given these findings, it is reasonable to suggest that many of the confusions made by subjects were the result of inadequate coding of less-intense acoustic cues in consonants. In our speech-processing system, the practical input dynamic range at the processor was approximately 50 to 55 dB. However, this was limited in the mapping from input intensity to stimulus intensity to a range of approximately 32 to 37.5 dB, depending on the individual subject. Attempts to use a larger dynamic range (more amplitude compressive mapping) introduced background noise, which some subjects found uncomfortable. In addition, Fu and Shannon (1999) demonstrated that although a high level of compressive amplitude mapping provided slightly better perception in quiet, less compressive mapping provided better perception in

noise. Thus, while the use of more compressive amplitude mapping techniques might offer better coding of low-intensity acoustic cues, the problems of increased background noise levels and reduced intensity contrast between some consonants need to be considered. The TESM strategy addressed these problems by applying additional gain based on the time profile of the slow-varying envelope in each band, rather than based on some function of its level as is normally done in multiband compressors. Emphasis of short-duration acoustic cues in particular was targeted. Thus, short-duration acoustic cues, particularly those of low intensity, were made more audible without greatly affecting overall consonant-to-vowel intensity contrast and/or increasing the level of background noise significantly.

It follows from the improved consonant scores that we should also expect increased perception of whole words, as was indeed the case. The small but significant increase in vowel scores could also be accounted for by improved consonant perception if one assumes that a proportion of vowel confusions is the result of language constraints imposed by incorrect consonant identification in the word tests. In addition, given that the onset of signals received some gain, we might expect a small improvement in the coding of vowel formant frequencies in cases where their intensities are low or during rapid formant glides as might occur for some diphthongs. This result is consistent with the findings of Geurts and Wouters (1999). However, in their case a bigger effect of strategy was shown for vowels which may be accounted for by other differences between the TESM (SMSP) and EECIS (CIS) coding strategies.

For the improvements to sentence comprehension in noise, a similar mechanism to that responsible for improved perception in quiet may have existed. However, it may also have been the case that onset envelope cues in the sentence material were, on average, enhanced by a greater degree than those of the background noise. Smaller fluctuations in the amplitude of the slow-varying envelope signal would be expected for the multitalker noise compared to that of a single speaker (i.e., target sentence material), even for a 0-dB SNR. Given that the gain applied by the TESM strategy was related to relative differences between peaks and valleys in the slow-varying envelope signal, less gain would thus be expected for fluctuations in the noise compared to that of the target speech. In addition, the spectrum of the noise in any analysis interval was on average broader than that of the target speech spectrum. Thus, for bands containing speech and noise, on average, the relative contribution of the noise to the gain would be less than that due to the target speech since the average signal level within the band was dominated by the speech. Therefore, better contrast between the speech and noise may have been provided by the TESM strategy.

While in this study the TESM strategy was compared to the SMSP, the findings may also be of relevance to other multichannel speech-processing strategies, such as SPEAK. Both the SMSP and SPEAK strategies are very similar, in fact, development of SPEAK was based on the SMSP strategy. Both strategies provide similar speech intelligibility, as was confirmed in comparative speech-perception tests conducted during the development of SPEAK which showed no significant differences in scores (Whitford *et al.*, 1993); unfortunately these data were never published. Given this to be the case, we would expect that application of the transient emphasis algorithm to the SPEAK strategy might provide similar perceptual benefits to those observed in this study.

V. CONCLUSION

The results obtained with the TESM strategy indicated that emphasis of short-duration acoustic cues, particularly those of low intensity, could provide improved speech perception for users of the SMSP strategy and possibly other strategies such as SPEAK. Six of the eight subjects obtained a significant improvement in speech comprehension using the TESM strategy in comparison to the SMSP strategy for one or more of the perceptual tests. For the sentence in noise tests, group mean scores were significantly higher, as were individual results for four of the eight subjects. For the CNC word tests, group mean scores were significantly higher for words, consonant, and vowel scores. Individual results were significantly higher for word scores in five subjects, for consonant scores in six subjects, and for vowel scores in two subjects.

An analysis of the consonant scores within the CNC word tests showed that perception of nasal, stop, and fricative consonants were most improved by the TESM strategy. Information transmission analysis indicated that the improvement to consonant perception could be accounted for mainly by better perception of place of articulation, although some improvement to perception of manner of articulation was also observed. Furthermore, the improvements to perception were found to be greatest for consonants such as nasal and stops. It was suggested that these improvements were the result of emphasis (amplification) of acoustic cues that were short in duration and low in intensity, and thus were normally poorly coded and not well perceived. In addition, it was found that some emphasis of the onset of long-duration acoustic signals resulted in fewer confusions for fricatives and semivowels. It was suggested that these improvements were mainly related to providing improved spectral information in cases where it was low in intensity and poorly coded. In a few cases, increases in fricative confusions were observed, which highlighted the importance of preserving relative amplitude differences between vowels and fricatives. Finally, the improvements to speech perception in noise indicated that: (a) the mechanisms providing improvements to consonant/word perception in quiet may have, in part, been robust to the effects of background noise, at least for the positive SNRs used in this study, and/or (b) better contrast between the (onset) envelopes of speech and noise may have been provided by the TESM strategy.

ACKNOWLEDGMENTS

The author wishes to thank the eight subjects who participated in this study. Thanks also to Richard Van Hoesel, Robert Cowan, Peter Busby, Hugh McDermott, Colette McKay, and Peter Blamey for their valuable contributions towards the preparation of this paper.

[1]Implementation of the filter bank was carried out using a fast Fourier transform (FFT). A 64-point Hanning window was employed, thus providing a 187.5-Hz FFT bin spacing and a low-pass filter cutoff frequency of 270 Hz. The magnitude of the first eight filter-bank bands was derived from FFT bins 2 to 9. The remaining eight filter-bank bands were derived from FFT bins 10 to 32 by summation of power (square root of magnitude squared terms) in adjacent FFT bins.

[2]The averaging window provided approximate equivalence to a second-order low-pass filter with a cutoff frequency of 45 Hz and was primarily used to smooth fine envelope structure, such as voicing frequency modulation, and unvoiced noise modulation.

[3]Even though some subjects reported perceptual differences between the two strategies, they quickly adapted to them. For instance, the 30-ms delay in processing introduced by the TESM strategy was noticeable to some subjects who commented that their own speech was followed by an echo but that other people's voices and external sounds were normal. Given that the time delay from production to perception of one's own voice is substantially less than 30 ms and remains approximately constant, we would expect this 30-ms delay to be noticeable. However, when listening to other sounds, time delays vary and can be longer than 30 ms and thus the processing delay may be less noticeable. In any case, subjects became accustomed to the delay within a matter of hours. In addition, subjects reported that the quality of speech was not noticeably different between the two strategies. This, too, was not surprising in that modifications made by the TESM strategy typically occurred only at transitional points within the signal over a short interval of time (typically $\leqslant$30 ms) and only on a few channels at any one time. Thus, only a small proportion of the total stimulus signal was modified.

[4]Percentage points (Δ%) denote the difference between mean percent-correct scores. For instance, if mean scores increased from 40% to 50%, this would be an increase of 10Δ%.

Boothroyd, A., Hanin, L., and Hnath, T. (**1985**). ''A sentence test of speech perception: Reliability, set equivalence, and short-term learning (Internal Report RCI 10),'' Speech & Hearing Sciences Research Center, City University of New York.

Dorman, M. F., Soli, S., Dankowski, K., Smith, L. M., McCandless, G., and Parkin, J. (**1990**). ''Acoustic cues for consonant identification by patients who use the Ineraid cochlear implant,'' J. Acoust. Soc. Am. **88**, 2074–2079.

Dorman, M. F., Loizou, P. C., and Fitzka, J. (**1998**). ''The identification of speech in noise by cochlear implant patients and normal-hearing listeners using 6-channel signal processors,'' Ear Hear. **19**, 481–484.

Dowell, R. C., Whitford, L. A., Seligman, P. M., Franz, B. Z. H., and Clark, G. M. (**1990**). ''Preliminary Results with a Miniature Speech Processor for the 22-electrode Melbourne/Cochlear Hearing Prosthesis,'' *Proceedings of the XIV World Congress of Otorhinolaryngology, Head and Neck Surgery, Madrid, Spain, 10–15 Sept. 1989* (Kugler and Ghedini, Amsterdam, 1990), pp. 1167–1173.

Fu, Q.-J., and Shannon, R. V. (**1999**). ''Phoneme recognition by cochlear implant users as a function of signal-to-noise ratio and nonlinear amplitude mapping,'' J. Acoust. Soc. Am. **106**, L18–L23.

Geurts, L., and Wouters, J. (**1999**). ''Enhancing the speech envelope of continuous interleaved sampling processors for cochlear implants,'' J. Acoust. Soc. Am. **105**, 2476–2484.

Hedrick, M. S., and Ohde, R. N. (**1993**). ''Effect of relative amplitude of frication on perception of place or articulation,'' J. Acoust. Soc. Am. **94**, 2005–2026.

Holden, L. K., Skinner, M. W., and Holden, T. A. (**1995**). ''Comparison of the normal and noise-suppression settings on the spectra 22 speech processor of the Nucleus 22-channel cochlear implant system,'' Am. J. Audiol. **4**(3), 55–58.

Kennedy, E., Levitt, H., Neuman, A. C., and Weiss, W. (**1998**). ''Consonants–vowel intensity ratios for maximizing consonant recognition by hearing-impaired listeners,'' J. Acoust. Soc. Am. **103**, 1098–1114.

Kewley-Port, D., and Pisoni, D. B. (**1983**). ''Perception of static and dynamic acoustic cues to place of articulation in initial stop consonants,'' J. Acoust. Soc. Am. **73**, 1779–1793.

Kiefer, J., Müller, J., Pfennigdorff, Th., 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 J. Otorhinolaryngol. Relat. Spec. **58**, 127–135.

Liberman, A. M. (**1956**). ''Some results of research on speech perception,'' J. Acoust. Soc. Am. **29**, 117–123.

McDermott, H. J., McKay, C. M., and Vandali, A. E. (**1992**). ''A new portable sound processor for the University of Melbourne/Nucleus Limited multielectrode cochlear implant,'' J. Acoust. Soc. Am. **91**, 3367–3371.

McDermott, H. J., Vandali, A. E., van Hoesel, R. J. M., McKay, C. M., Harrison, J. M., and Cohen, L. T. (**1993**). ''A portable programmable digital sound processor for cochlear implant research,'' IEEE Eng. Med. Biol. Soc. **1**(2), 94–100.

McKay, C. M., McDermott, H. J., Vandali, A. E., and Clark, G. M. (**1991**). ''Preliminary results with a six spectral maxima sound processor for the University of Melbourne/Nucleus multiple-electrode cochlear implant,'' J. Oto-Laryngol. Soc. Austra. **6**, 354–359.

McKay, C. M., and McDermott, H. J. (**1993**). ''Perceptual performance of subjects with cochlear implants using the Spectral Maxima Sound Processor (SMSP) and the Mini Speech Processor (MSP),'' Ear Hear. **14**, 349–367.

Miller, G. A., and Nicely, P. E. (**1955**). ''An analysis of perceptual confusions among some English consonants,'' J. Acoust. Soc. Am. **27**, 338–352.

NIH Consensus Statement (**1995**). ''Cochlear Implants in Adults and Children,'' National Institutes of Health, Office of the Director **13**(2), 15–17 May, 1995.

Parkinson, A. J., Parkinson, W. S., Tyler, R. S., Lowder, M. W., and Gantz, B. J. (**1998**). ''Speech perception performance in experienced cochlear-implant patients receiving the speak processing strategy in the Nucleus spectra-22 cochlear implant,'' J. Speech Lang. Hear. Res. **41**, 1073–1087.

Peterson, G., and Lehiste, I. (**1962**). ''Revised CNC lists for auditory tests,'' J. Speech Hear. Disord. **27**, 62–70.

Sammeth, C. A., Dorman, M. F., and Stearns, C. J. (**1999**). ''The role of consonant–vowel intensity ratio in the recognition of voiceless stop consonants by listeners with hearing impairment,'' J. Speech Lang. Hear. Res. **42**(1), 42–55.

Skinner, M. W., Holden, L. K., Holden, T. A., Dowell, R. C., Seligman, P. M., Brimacombe, J. A., and Beiter, A. L. (**1991**). ''Performance of post-linguistically deaf adults with the Wearable Speech Processor (WSP III) and the Mini Speech Processor (MSP) of the Nucleus multielectrode cochlear implant,'' Ear Hear. **12**, 3–22.

Skinner, M. W., Clark, G. M., Whitford, L. A., Seligman, P. A., Staller, S. J., Shipp, D. B., Shallop, J. K., Everingham, C., Menapace, C. M., Arndt, P. L., Antogenelli, T., Brimacombe, J. A., Pijl, S., Daniels, P., George, C. R., McDermott, H. J., and Beiter, A. L. (**1994**). ''Evaluation of a new spectral peak (SPEAK) coding strategy for the Nucleus 22 channel cochlear implant system,'' Am. J. Otol. **15** (Suppl. 2), 15–27.

Stevens, K. N., and Blumstein, S. E. (**1978**). ''Invariant cues for place of articulation in stop consonants,'' J. Acoust. Soc. Am. **64**, 1358–1368.

Tyler, R. S., and Moore, B. C. J. (**1992**). ''Consonant recognition by some of the better cochlear-implant patients,'' J. Acoust. Soc. Am. **92**, 3068–3077.

Vandali, A. E., Harrison, J. M., Huigen, J. M., Plant, K., and Clark, G. M. (**1995**). ''Multichannel cochlear implant speech processing: Further variations of the Spectral Maxima Sound Processor strategy,'' Ann. Otol. Rhinol. Laryngol. **104**, (9 part 2), Suppl. 166, 378–381.

Whitford, L. A., Seligman, P. M., and McDermott, H. J. (**1993**). ''Preliminary Testing of the SPEAK strategy in the SPEAK Processor,'' Cochlear Ltd, Australia, Internal Report, 25 Jan. 1993.

Whitford, L. A., Seligman, P. M., Everingham, C. E., Antognelli, T., Skok, M. C., Hollow, R. D., Plant, K. L., Gerin, E. S., Staller, S. J., McDermott, H. J., Gibson, W. R., and Clark, G. M. (**1995**). ''Evaluation of the Nucleus Spectra 22 processor and new speech processing strategy (SPEAK) in postlinguistically deafened adults,'' Acta Oto-Laryngol. (Stockholm) **115**, 629–637.

Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., and Rabinowitz, W. M. (**1991**). ''Better speech recognition with cochlear implants,'' Nature (London) **352**, 236–238.

Zeng, F.-G., and Galvin, J. J. (**1999**). ''Amplitude mapping and phoneme recognition in cochlear implant listeners,'' Ear Hear. **20**, 60–74.

The relationship between frequency selectivity and overshoot

Elizabeth A. Strickland[a)]
Department of Audiology and Speech Sciences, Purdue University, West Lafayette, Indiana 47907

(Received 19 May 2000; revised 7 August 2000; accepted 25 January 2001)

Under some conditions, threshold for a brief tone is higher at the onset of a broadband masker than it is if it is delayed from the onset of the masker. Evidence suggests that this ''overshoot'' is related to active processing in the auditory system. The present experiments examined this question, by measuring frequency selectivity under the same conditions in which overshoot was measured. The first experiment demonstrated that the growth of masking with masker level was approximately linear for a 1-kHz signal with or without a precursor (which was identical to the masker), and a 4-kHz signal with a precursor. For the 4-kHz signal with no precursor, an elevation in signal-to-masker ratio was seen at mid masker levels, relative to the other conditions. Frequency selectivity was then measured for a fixed-level signal, with and without a precursor. Relative frequency selectivity was highest for the 4-kHz signal with no precursor, lower for the 1-kHz signal with no precursor, and lowest for the 1- or 4-kHz signal with a precursor. The overshoot results and the frequency selectivity results would be consistent with stronger active processing at 4 kHz than at 1 kHz, and a decrease in active processing following a broadband noise precursor. [DOI: 10.1121/1.1357811]

PACS numbers: 43.66.Ba, 43.66.Dc, 43.66.Mk [SPB]

I. INTRODUCTION

Numerous experiments have shown that the threshold for a brief tone decreases as the onset of the tone is delayed from the onset of a noise masker (e.g., Zwicker, 1965; Elliott, 1965). This change in threshold has been called overshoot. Although it has been demonstrated for various spectral configurations of the signal and masker, for the sake of simplicity, the overshoot discussed in this paper will primarily be for a tonal signal in a broadband masker. Overshoot is an interesting phenomenon, because it suggests that stimulating the auditory system with sound actually changes its response to sound. Different theories have focused on different levels of the auditory system at which this change might take place. The explanation that has most often been proposed is that overshoot is due to adaptation at the level of the auditory nerve (Champlin and McFadden, 1989; Bacon, 1990; McFadden and Champlin, 1990; Bacon and Healy, 2000). If a peristimulus time histogram (PSTH) is recorded for repeated stimulation of an auditory-nerve fiber, it will show a large onset response, and then a decrease in firing. An increment to the stimulus causes the same increase in firing whether it is presented at the onset of the longer stimulus, or is delayed from the onset (Smith and Zwislocki, 1975). Therefore, the proportional response to the increment is larger when the increment is delayed from the onset of the stimulus. Although neural adaptation may play a role in overshoot, it cannot account for all aspects of overshoot, as has been summarized by Bacon and Healy (2000), and as will be discussed below.

Another explanation is that the change is at the level of the cochlea. A growing body of data suggests that the high signal threshold at masker onset is associated with active processing in the auditory system. This active processing is thought to be due to the action of the outer hair cells, and is associated with, among other things, high absolute sensitivity, sharp tuning, and a nonlinearity in the input/output function of the basilar membrane. Studies have shown that ingestion of aspirin (McFadden and Champlin, 1990), the presence of hearing loss greater than 30 dB (Bacon and Takahashi, 1992), and temporary threshold shift induced by exposure to a high-intensity tone (Champlin and McFadden, 1989) reduce or abolish overshoot by decreasing thresholds at masker onset. These same conditions have been found to decrease frequency selectivity and absolute sensitivity (Wightman *et al.*, 1977; Feth *et al.*, 1980; Evans *et al.*, 1981; Carlyon and Butt, 1993; Beveridge and Carlyon, 1996; Hicks and Bacon, 1999b), suggesting that they decrease the active processing in the cochlea. In addition, overshoot is larger for tones above 2 kHz than for lower frequencies (Bacon and Takahashi, 1992; Carlyon and White, 1992). There is evidence that active processing is also stronger at high frequencies than at low (Cooper and Yates, 1994; Cooper and Rhode, 1995; Hicks and Bacon, 1999a, 1999b).

The specific aspect of active processing related to overshoot could be the nonlinearity in the input/output function on the basilar membrane, as has been suggested by von Klitzing and Kohlrausch (1994). This function has a shallow slope at mid input levels, and a steeper slope at low and high input levels (e.g., Ruggero and Rich, 1991). If a signal is detected by causing some criterion change in output above the masker, then the relative signal level should need to be higher for midlevel signals. This would be consistent with the fact that overshoot is largest at mid masker levels (Bacon, 1990; Carlyon and White, 1992; von Klitzing and Kohlrausch, 1994; Overson *et al.*, 1996). As pointed out by von Klitzing and Kohlrausch, the fact that threshold decreases as the signal is delayed from masker onset would suggest that

a)Electronic mail: estrick@purdue.edu

the nonlinearity decreases as the system is stimulated by the masker.

One way to explore the active processing hypothesis and the neural adaptation hypothesis would be to look for other psychophysical measures that may vary under overshoot conditions. As noted above, both the nonlinear input/output function and sharp tuning are linked to active processing. If active processing is high at masker onset, and decreases with delay from the onset of the masker, this suggests that tuning should be sharper at onset, and become broader with delay. The neural adaptation hypothesis, on the other hand, does not predict any change in tuning associated with overshoot. Thus, examining frequency selectivity under conditions of overshoot could help determine the role of active processing.

The temporal course of frequency selectivity has been examined in studies that have varied the frequency relationship between the masker and the signal, and compared signal thresholds when the signal is near masker onset to thresholds when signal onset is delayed from masker onset. If the signal and masker are tonal, masker effectiveness decreases with signal delay for masker frequencies above the signal frequency (Bacon and Viemeister, 1985; Bacon and Moore, 1986; Kimberley *et al.*, 1989). This occurs for signals from 500 to 3000 Hz, and for maskers whose relative frequencies in units of $|f-f_s|/f_s$, where f is the masker frequency and f_s is the signal frequency, are 0.2 to 0.6 units above the signal frequency. Essentially the same results have been found in studies using notched-noise maskers (Carlyon, 1989; Wright, 1989; Hicks and Bacon, 1992), although one study found no significant change (Moore *et al.*, 1987). These results have been taken to suggest that auditory filtering sharpens as the system is stimulated, and although it has been noted that it could not account for the entire overshoot effect, it has been mentioned as a contributing factor (Scholl, 1962; Elliott, 1965, 1967; Green, 1969). If filtering narrows with stimulation, less masker energy will pass through the filter when the signal is presented after a delay, and thus threshold will be lower.

Thus, there seems to be a discrepancy between the active processing hypothesis and the studies of the temporal course of frequency selectivity. In the above studies, however, the stimulation preceding the tone has been the narrowband masker, which has been varied in frequency. In contrast, in measuring overshoot, the stimulation preceding the tone has been a broadband noise. If whatever precedes the tone affects active processing, the stimulation preceding the tone should be held constant when measuring tuning. Also, in some of the studies the masker level has been fixed and the signal level adjusted to determine threshold. If overshoot is related to the nonlinear basilar-membrane input/output function, it is important that the signal level be fixed to keep its position on the function constant. Finally, in order to determine whether tuning is related to overshoot, frequency selectivity must be measured under conditions where overshoot varies. The purpose of the present study is to determine whether there is a relationship between frequency selectivity and overshoot, and if so, whether frequency selectivity increases or decreases with preceding stimulation.

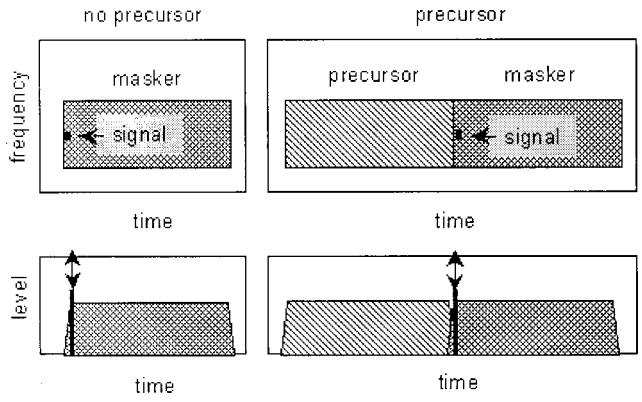


FIG. 1. Schematic showing the temporal and spectral characteristics of the stimuli for measuring overshoot.

II. METHOD

A. Overshoot stimuli and procedure

The signal was a sinusoid with a frequency of 1 or 4 kHz. Signal duration was 10 ms, including 5-ms $\cos^2$ onset and offset ramps (all stimulus durations are specified by 0-voltage points). This duration was chosen because it is long enough to preclude effects of spectral splatter (Bacon and Viemeister, 1985). The masker was a flat-spectrum broadband noise from 10 to 10 000 Hz. It had a duration of 200 ms, including 5-ms $\cos^2$ onset and offset ramps. The signal onset always occurred 2 ms after masker onset. Rather than manipulating the signal onset delay, a separate noise was added before the masker, which will be called a precursor. The precursor had the same long-term spectrum as the masker, but was 205 ms in duration. When present, the precursor offset overlapped the masker onset by 5 ms.[1] Thus, the precursor also overlapped the signal by 3 ms. Precursor and masker spectrum levels were equal, and were fixed at 0 to 30 dB across blocks. In Fig. 1 the spectral (top row) and temporal (bottom row) parameters of the stimuli are shown schematically.

The signal, masker, and precursor were digitally generated and produced through three separate D/A channels at a rate of 25 kHz (TDT DA1). The stimuli were low-pass filtered at 10 kHz (TDT FT5 and TDT FT6-2). The levels were adjusted by programmable attenuators (TDT PA4). Stimuli were presented through one Etymotic ER-2 insert earphone to a listener seated in a double-walled sound-attenuating booth.

The initial signal level was set above estimated threshold. Signal level was adjusted using a three-interval forced-choice (3IFC) task with a two-down, one-up stepping rule to estimate the 71%-correct point on the psychometric function (Levitt, 1971). Time intervals and feedback were marked visually on a computer screen, and the listener responded via a computer keyboard. The initial step size was 5 dB, and decreased to 2 dB after the first two reversals. Fifty trials were presented, and thresholds were taken as the average of the last even number of reversals at the smaller step size. Blocks for which the standard deviation was 5 dB or greater were discarded. Data points are the average of at least four blocks, although in three cases only two or three blocks were aver-

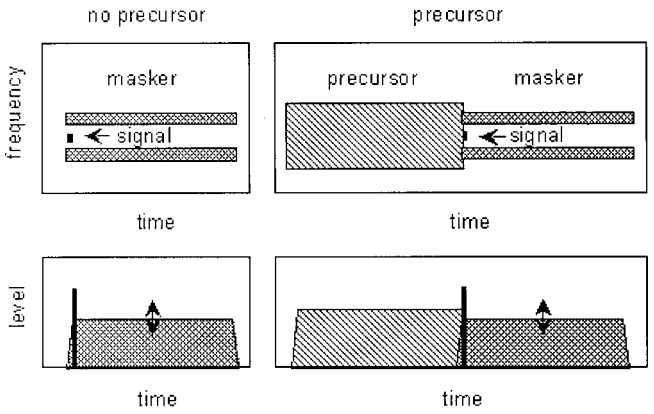


FIG. 2. Schematic showing the temporal and spectral characteristics of the stimuli for measuring frequency selectivity.

aged due to high standard deviations in the others. Overshoot may decrease with practice, so in the condition where the most overshoot was seen (signal frequency of 4 kHz, masker spectrum level of 20 dB, no precursor), each listener completed at least eight blocks. For two listeners (L3 and L6), substantial improvement was seen with practice in some of the conditions with no precursor, so only the last four blocks were included in the average. L4 was only tested at masker spectrum levels of 10 and 20 dB, and L6 only at 20 dB. L3 was tested with the 1-kHz signal, but her data are not included because her thresholds had not stabilized before testing ended.

B. Frequency selectivity stimuli and procedure

The signals and precursor were the same as those used in the overshoot conditions. The masker was a flat-spectrum noise with a notch around the signal frequency. The outer spectral edges of the noise were fixed at $f_s \times (1 \pm 0.8)$, where f_s is the signal frequency. Notch edges are expressed in normalized units of $|f - f_s|/f_s$, where f is the edge of the notch. This normalized unit will be referred to as Δf. Notch edges were placed symmetrically around the signal, and asymmetrically, with one edge 0.2 units further from the signal frequency than the other edge. For signal levels below 70 dB SPL, Δf was set at values of 0.0, 0.1, 0.2, 0.3, and 0.4 for the symmetric case, and 0.0, 0.1, 0.2, and 0.3 for the closer edge for the asymmetric case. For higher signal levels, the nearer notch edge was set at 0.0 to 0.2 in 0.05 steps. Larger Δf's were not used at these levels in general, because they required higher masker levels to reach threshold. This would have exceeded the limit of the system (95 dB SPL) and been uncomfortably loud for the subjects. The precursor, if present, and signal were fixed at the level chosen from the results in the overshoot experiment. Stimulus generation was the same as for the overshoot stimuli. Stimuli are shown schematically in Fig. 2.

The initial masker level was set below estimated threshold. Masker level was adjusted using a 3IFC task with a two-up, one-down stepping rule to estimate the 71%-correct point (Levitt, 1971). The initial step size was 5 dB, and decreased to 2 dB after the second reversal. Fifty trials were presented, and thresholds were taken as the average of the last even number of reversals at the smaller step size. Blocks for which the standard deviation was 5 dB or greater were discarded. Data points are the average of at least three blocks, although in three cases only two blocks were averaged. In these cases, the standard deviation of the blocks was 1.15 dB or less. For L3, one block was discarded because it was more than 2 standard deviations from the mean of the other (four) blocks in that condition.

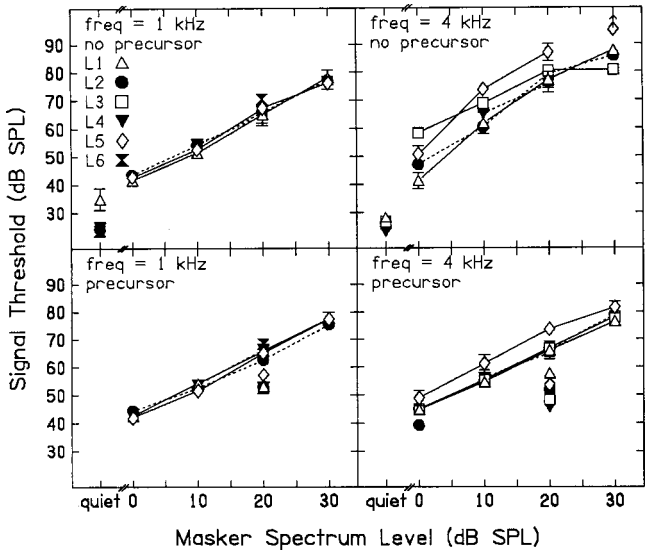


FIG. 3. Signal detection threshold as a function of masker spectrum level for the six listeners. Error bars are one standard deviation about the mean. The symbol with the arrow indicates that threshold could not be measured without exceeding the limits of the system. Unconnected symbols at the left in the top row are thresholds for the signals in quiet. Unconnected symbols in the bottom row are thresholds for the signal with the precursor at the level shown on the abscissa, but no masker.

C. Listeners

Two males and four females served as listeners. They had thresholds within laboratory norms for 200-ms pure tones at octave frequencies from 250 to 8000 Hz. The left ear was the test ear for all listeners. Three listeners were 23, and the other ages were 24, 27, and 40. Not all listeners participated in all conditions. All listeners except L2, who is the author, were paid for their participation.

III. RESULTS

A. Overshoot

In Fig. 3, thresholds are shown for the six listeners for the 1-kHz signal (left column) and 4-kHz signal (right column), with no precursor (top row), and with a precursor (bottom row). Listeners are represented by different symbols. The symbol with the upward arrow indicates that threshold could not be measured because the signal level reached the limit of the system. The thresholds for the 4-kHz signal with no precursor stand out from the other conditions in several ways. Variability among subjects was greatest in this condition, and in general, for each masker level, thresholds were highest in this condition. In the other three conditions, the increase in signal level with masker level was approximately linear. These results are consistent with other studies (Bacon,

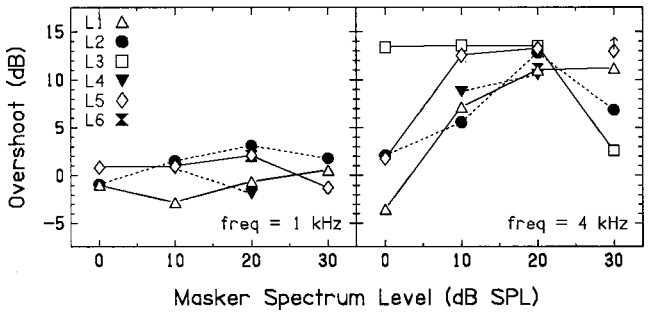


FIG. 4. Overshoot as a function of masker spectrum level for the six listeners.

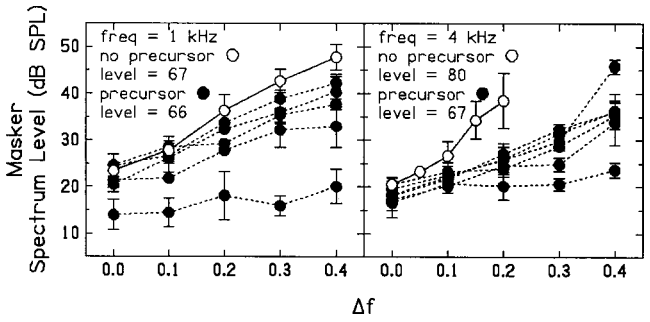


FIG. 5. Masker level needed to mask a fixed-level signal as a function of Δf. The precursor data are for individual listeners, and the error bars are one standard deviation about the mean. The no-precursor data are the mean data for all listeners, and the error bars are one standard deviation about this mean. The signal level was fixed at threshold level from the overshoot experiment when the masker spectrum level was 20 dB.

1990; Bacon and Takahashi, 1992; Carlyon and White, 1992; von Klitzing and Kohlrausch, 1994). Least-squares fits to the individual data yielded average slopes of 1.16 (rms err=1.38) and 1.13 (rms err=1.21) for the 1-kHz signal with no precursor and with a precursor, and 1.05 (rms err=0.80) for the 4-kHz signal with a precursor. The data for the 4-kHz no-precursor condition were not well fit by a single line (slope=1.22, rms err=5.34). Therefore, these data were fit with two lines, with the break point at a masker spectrum level of 20 dB. Slopes for the lower masker levels ranged from 1.05 to 1.81, with an average of 1.47 (rms err=1.54). The average slope for the higher masker levels was 0.58, with a range from 1.06 to 0.03.

Thresholds for the signal alone are shown by the unconnected symbols on the left side of the top two panels. Thresholds for the signals in the presence of the precursor and no masker are shown by the unconnected symbols in the bottom row. The difference between these values shows the amount of forward masking produced by the precursor, which will be discussed in Sec. IV.

For a given signal frequency and masker level, the difference between threshold with no precursor and threshold with a precursor was defined as overshoot. In Fig. 4, the amount of overshoot is shown as a function of masker level for the six listeners. The signal frequency was 1 kHz in the left panel, and 4 kHz in the right panel. Listeners showed little or no overshoot at a signal frequency of 1 kHz, regardless of masker level. At a signal frequency of 4 kHz, all listeners showed overshoot. The maximum average amount was 13 dB at a masker spectrum level of 20 dB. The frequency effects are consistent with the data of Bacon and Takahashi (1992), who also found a greater amount of overshoot at 4 kHz than 1 kHz. The level effects are consistent with previous studies (Bacon, 1990; Carlyon and White, 1992; von Klitzing and Kohlrausch, 1994; Overson *et al.*, 1996), which also found that overshoot is largest at mid-masker levels.

B. Frequency selectivity

In this condition, frequency selectivity was measured for the same signals used in the overshoot experiment. Listeners showed the maximum overshoot when the signal was 4 kHz and the masker spectrum level was 20 dB. Therefore, the results from this condition were used for the overshoot condition. Most listeners showed almost no overshoot when the signal was 1 kHz and the masker spectrum level was 20 dB, so this was chosen as a comparison no-overshoot condition. For two listeners, L1 and L2, who showed little overshoot when the signal was 4 kHz and the masker spectrum level was 0 dB, this was chosen as a second comparison no-overshoot condition. L1 was also tested using results from the condition where the masker spectrum level was 0 dB and the signal was 1 kHz.

The first condition was set up to examine frequency selectivity with the signal set at threshold level from the overshoot experiment. The precursor (if present) was set at the same spectrum level as in the overshoot experiment, and the signal was fixed at the threshold level for a given condition. For example, for L1, when the masker spectrum level was 20 dB and there was no precursor, the threshold for the 4-kHz signal was 77 dB SPL (see Fig. 3, open triangle in upper-right panel). Therefore, the signal was set at 77 dB SPL when there was no precursor in the frequency selectivity condition. Likewise, for L1, when the masker spectrum level was 20 dB and it was preceded by a precursor at the same level, threshold for the 4-kHz signal was 66 dB SPL (Fig. 3, open triangle in lower-right panel). In the frequency selectivity condition, the signal level was set at 66 dB SPL and a precursor at a spectrum level of 20 dB preceded the signal. The level of a notched-noise masker was varied to find the level that just masked the signal.

The results for the six listeners are shown in Fig. 5 for the 1-kHz signal (left panel) and the 4-kHz signal (right panel), for conditions chosen from the overshoot experiment when the masker was at a spectrum level of 20 dB. Threshold masker spectrum levels as a function of Δf are shown without a precursor (open circles) and with a precursor (filled circles). Individual data are shown for the precursor condition, because there was some individual variability. The no-precursor results are average results. For clarity, only results from the symmetric notch conditions are shown. When $\Delta f = 0.0$, the threshold masker level was close to a spectrum level of 20 dB, as expected, because the signal was set at threshold from the overshoot experiment when the masker was at 20 dB spectrum level. It is clear that tuning is broader after a precursor rather than narrower, and that for most listeners the change in tuning is greater at 4 kHz than at 1 kHz.

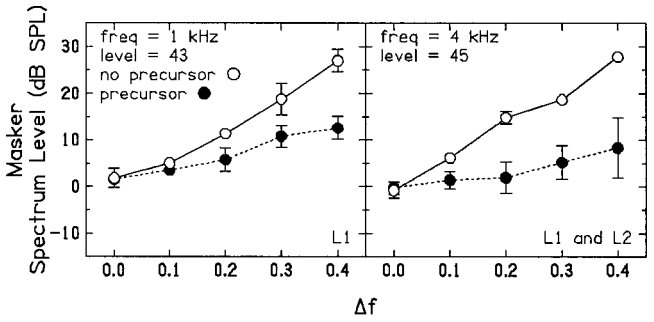


FIG. 6. Masker level needed to mask a fixed-level signal as a function of Δf, for L1 in the left panel, and the mean of L1 and L2 in the right panel. Error bars are one standard deviation about the individual data in the left panel and the mean data in the right panel. The signal was fixed at the threshold level from the overshoot experiment with a masker spectrum level of 0 dB.

In Fig. 6, results are shown in the same manner as in Fig. 5, but for conditions chosen from the overshoot experiment when the masker was at a spectrum level of 0 dB. Recall that there was little to no overshoot at this masker level for either frequency. L1 was the only listener tested in this condition with a 1-kHz signal, and this listener's results are shown in the left panel. In the right panel the average results for L1 and L2 are shown for a 4-kHz signal. The results are extremely similar to those in Fig. 5, in that tuning is broader after a precursor, and the change in tuning is slightly greater for the 4-kHz signal than for the 1-kHz signal. In Fig. 6, however, the signal was set at nearly the same level with and without a precursor for 4 kHz as well as 1 kHz.

In Fig. 5, the level of the 4-kHz signal was higher with no precursor than with a precursor, because all of the listeners showed overshoot. In the next condition, the signal level was fixed, in order to look at the results in terms of the effects of the precursor. From the first condition, data were already available for one level with a precursor, and a higher level with no precursor. Therefore, these same two levels were used, but for the lower level thresholds were measured without a precursor, and for the higher level, they were measured with a precursor at 20 dB spectrum level. Data for these two new comparison conditions were obtained for L2, L3, and L6, and are shown in Fig. 7. In the left column, the average signal level was 66.7 dB SPL (see Fig. 3). Masker thresholds with no precursor are shown by the open circles, while thresholds with a precursor are shown by the filled circles. In the right column, results are shown with the signal set to 80 dB SPL for all three listeners. At both levels, the effect of the precursor was to increase the threshold masker level for small Δf, and to decrease threshold masker level for large Δf. The transition from increase to decrease occurred at a Δf between 0.1 to 0.2 at the lower level, and for a slightly larger Δf at the higher level. At the higher level, the largest Δf's were only tested for L3, due to concern about exposure to higher masker levels (see Sec. II and the following section). For this listener, the precursor had a much smaller effect at the larger Δf at the higher level than it did at the lower level. This was also true for L2 and L6 for a Δf of 0.2.

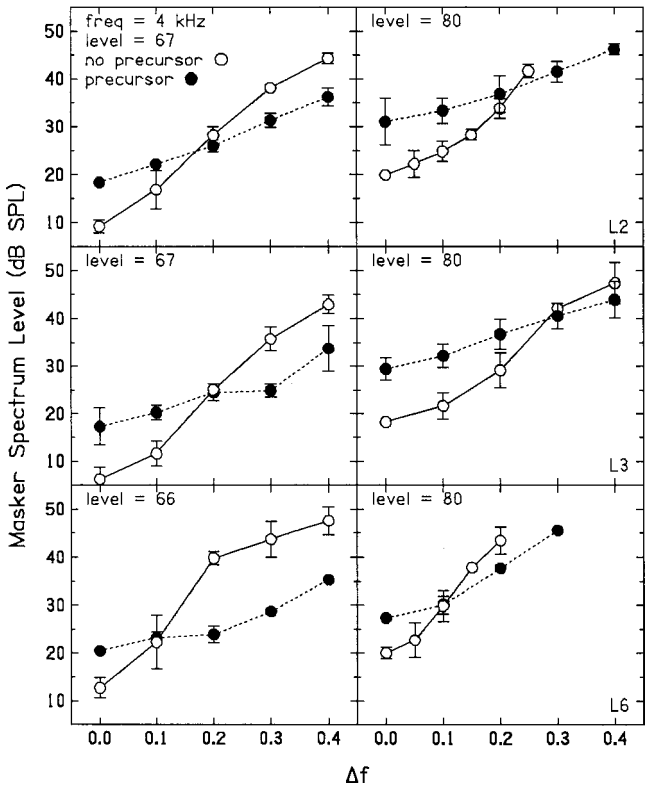


FIG. 7. Masker level needed to mask a fixed-level signal as a function of Δf, for three listeners. Error bars are one standard deviation about the mean.

In order to quantify the frequency selectivity, filter functions were fit to the individual and the mean data using an adaptive polynomial fitting program called POLYFIT (Rosen *et al.*, 1998). Thresholds measured using symmetric and asymmetric (not shown in figures) notches were used in fitting the filter functions. The function used to fit the filters to the no-precursor data was a rounded exponential, with a weighting function for the upper skirt as described by the following equation:

$$W(g)=(1-r_u)(1+p_u g)\exp(-p_u g)+r_u, \tag{1}$$

and the weighting function for the lower skirt described by the following equation:

$$W(g)=(1-r_l)(1+p_l g)\exp(-p_l g)+r_l, \tag{2}$$

where g is frequency normalized to the center frequency, p_u and p_l are the slopes of the upper and lower skirts of the filter, and r_u and r_l are absolute thresholds (Patterson *et al.*, 1982). For the precursor data, r_u and r_l were set to zero because this minimized the error per degree of freedom. The other parameter used in fitting was filter efficiency (k). The values of $Q_{3\text{dB}}$ (center frequency/bandwidth) are shown for the different listeners and conditions in Table I. $Q_{3\text{ dB}}$ is clearly higher, i.e., tuning is sharper, with no precursor than with a precursor. When there is no precursor, tuning is sharper at 4 kHz than at 1 kHz, when compared at equal signal levels. Also, the decrease in $Q_{3\text{ dB}}$ when a precursor is added is greater at 4 kHz than at 1 kHz. The decrease in $Q_{3\text{ dB}}$ is not directly related to overshoot, because it is seen at 1 kHz, where there was little or no overshoot, and at 4 kHz, where there was overshoot at the two higher levels but not at

TABLE I. Filter parameters for roex filters fit to individual and mean data. Where two signal levels are listed, the first is the level with no precursor, and the second with a precursor.

Listener	Signal freq. (kHz)	Signal level (dB SPL)	Q_{3dB}		k	
			No pre	Pre	No pre	Pre
L1	1	43	5.42	2.72	18.06	15.11
	1	65/66	7.56	4.26	22.41	17.23
	4	42/45	7.47	1.00	15.16	10.56
	4	66		0.68		10.85
	4	76	9.89		26.89	
L2	1	65/68	6.07	2.86	18.21	13.29
	4	47/45	5.85	2.61	17.44	12.42
	4	67	7.81	4.15	33.65	18.86
	4	80	9.20	3.76	30.63	19.84
L3	4	67	6.39	2.82	33.46	17.60
	4	80	9.19	3.48	34.08	20.39
L4	1	65/66	6.53	4.30	19.03	19.05
	4	65		4.76		18.53
	4	76	8.89		26.57	
L5	1	68/65	6.50	1.63	20.70	21.95
	4	74		4.86		27.42
	4	88	8.80		39.41	
L6	1	68	4.75	3.07	22.06	21.20
	4	66	11.64	3.20	26.78	14.77
	4	80	10.90	5.60	33.77	24.49
Mean	1	43	5.42	2.72	18.06	15.11
	1	67/66	6.44	3.27	20.24	18.66
	4	45	6.58	1.95	16.22	12.43
	4	67	9.79	3.36	30.36	17.83
	4	80	8.71	3.58	31.98	20.90

the lower level. Values of the signal-to-noise ratio at threshold, k, are also shown in Table I. The values are large overall because of the short duration of the signals. The amount of overshoot is largely reflected in the change in k between the no-precursor and the precursor condition.

The $Q_{3\,\mathrm{dB}}$ values in Table I are comparable to the few data available for similar conditions. Hant *et al.* (1997) derived filter estimates by measuring threshold for a bandpass noise signal in a broadband noise masker as a function of signal bandwidth. The signal was 10 ms, and was temporally centered in the 750-ms masker, so these results may be compared to the precursor data in the present study, although the masker level was fixed (at a spectrum level of 36 dB) rather than the signal level. They reported equivalent rectangular bandwidths (ERBs) of approximately 800 and 1800 Hz for signals centered at 1 and 4 kHz, respectively. This would correspond to a $Q_{3\,\mathrm{dB}}$ of 1.49 at 1 kHz and 2.64 at 4 kHz, values which are in the same range as the values in the precursor condition in Table I. Wright and Dai (1994) measured threshold for a 10-ms, 2.5-kHz signal at the onset of a notched-noise masker fixed in spectrum level at 25 dB. They reported ERBs of 397 to 563 Hz, which correspond to $Q_{3\,\mathrm{dB}}$ values of 5.27 to 7.5. These values are within the range of values reported for the no-precursor condition in Table I.

In the left column of Fig. 5 and the left column of Fig. 7, the signal levels were approximately equal. The difference is that the signal is 1 kHz in Fig. 5, and 4 kHz in Fig. 7. For larger Δf, the effect of the precursor is very similar for signal frequencies of 1 and 4 kHz. The difference is that when

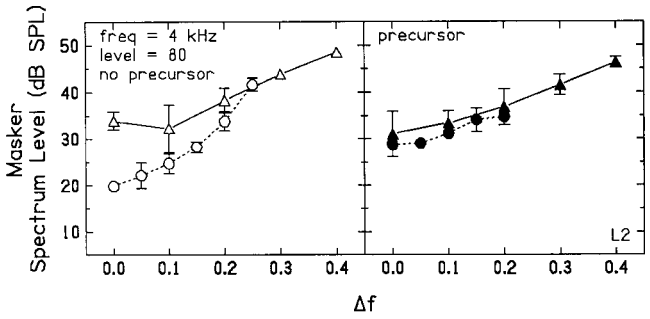


FIG. 8. Masker level needed to mask a fixed-level signal for a small (circles) and a large (triangles) range of Δf, for L2. Error bars are one standard deviation about the mean.

the signal is 4 kHz, adding the precursor also *increases* threshold masker level for Δf less than 0.2. At the lower signal level, shown in Fig. 6, the effect of the precursor is the same for both frequencies.

On average, there was no particular tendency for the decrease in filter slope with the addition of the precursor to be greater for one side than the other. This may be compared to results for aspirin ingestion, another condition which decreases filter slope. Hicks and Bacon (1999b) found on average a greater decrease in slope on the low-frequency side, using a fixed masker level and a notched-noise masker. Beveridge and Carlyon (1996) found a slight tendency for the slope to decrease more on the high-frequency side, using a fixed signal level and a narrow-band masker.

C. Effects of range of Δf at a high signal level

As noted above, when the signal was 4 kHz at a level of 80 dB SPL, and there was no precursor, L3 was tested at Δf's from 0.0 to 0.4. L2 was also originally tested over this same wide range of Δf. Thresholds for L2 are shown by the open triangles in the left panel of Fig. 8. The results were paradoxical in that when $\Delta f = 0.0$, threshold spectrum levels were approximately 30 dB rather than the 20 dB that would be expected from the overshoot experiment. At $\Delta f = 0.3$ and 0.4, threshold masker spectrum levels were near 45 and 50 dB, which means the overall level was near 81 to 85 dB SPL at threshold. During a run, of course, some trials would be above this level. On inspection of the data, it happened that threshold measurements for small Δf's were always preceded by threshold measurements for large Δf's, although the conditions were randomized. Therefore, it seemed possible that by using intense masker levels, overshoot was in fact being abolished. Champlin and McFadden (1989) demonstrated that exposure to intense sounds can produce a temporary threshold shift (TTS) and also decrease overshoot.

To test this hypothesis, thresholds for listener L2 were remeasured for this condition using only $\Delta f = 0.2$ and less. Thresholds when $\Delta f = 0.25$ were then measured separately, to further define the function. The results are shown by the open circles in the left panel of Fig. 8. The threshold masker spectrum level for $\Delta f = 0.0$ was 20 dB, as expected. Note also that tuning was sharper when the narrower range was used. Champlin and McFadden (1989) noted that TTS would be accompanied by a decrease in frequency selectivity,

which was accidentally tested here. TTS was not measured in this experiment, but exposure to high masker levels did increase threshold masker level for this signal. This is certainly a direct argument against filter *narrowing* as an explanation for overshoot, which was also noted by McFadden and Champlin (1990).

The right panel of Fig. 8 shows that testing over a wide range of Δf had little effect on performance when there was a precursor. When a precursor was present, thresholds were similar to those measured with no precursor over a wide range of Δf. Both decreased tuning, by increasing threshold masker level for Δf less than 0.2. This is consistent with the results above, that a broadening of tuning at small Δf is associated with lower thresholds for the signal in the masker. Also note that the effects were not additive, as the range of Δf had no effect when there was a precursor. Clearly L3 was not affected in the same way by the broader range of Δf, suggesting that there is individual variability in this effect.

IV. DISCUSSION

As was noted in the Introduction, it has become clear that overshoot is related to active processing in the auditory system, and this could suggest that active processing decreases with acoustic stimulation. The present experiments have shown that frequency selectivity also decreases with acoustic stimulation by a precursor, which adds support to this hypothesis.

In the first experiment, it was shown that threshold for detection of a short-duration signal at the onset of a broadband masker was similar for a 1-kHz signal with no precursor, a 1-kHz signal with a precursor, and a 4-kHz signal with a precursor. In these conditions there was relatively low variability across subjects, and the growth of signal threshold with masker level was essentially linear with a slope slightly greater than 1. In contrast, detection threshold for a 4-kHz signal with no precursor was variable across subjects, and showed a steeper growth of signal threshold with masker level for low masker levels, that for some listeners became shallower at higher masker levels. Thus, for the 4-kHz signal, thresholds were higher with no precursor than with a precursor, and this overshoot was greatest at mid masker levels. This is consistent with other studies (Bacon, 1990; Carlyon and White, 1992; von Klitzing and Kohlrausch, 1994; Overson *et al.*, 1996).

When frequency selectivity was measured by fixing the signal level and varying the level of a notched-noise masker, the 4-kHz no-precursor condition again stood out because, for a given signal level, relative tuning was sharpest in this condition. This suggests that active processing is strongest in this condition, because it is associated with sharp tuning. Tuning was somewhat broader for the 1-kHz no-precursor condition, which could suggest that active processing is weaker at this frequency. This is consistent with other psychophysical and physiological data showing that active processing is stronger at higher frequencies than lower frequencies (Cooper and Yates, 1994; Cooper and Rhode, 1995; Moore *et al.*, 1996; Rhode and Cooper, 1996; Hicks and Bacon, 1999a, 1999b).

For a given signal level, relative tuning was broader with a precursor than with no precursor. The manner in which this broadening occurred depended on signal frequency and level, as well as precursor level. For the 4-kHz signal, the precursor had no effect for small Δf's at the lowest signal level (with a precursor spectrum level of 0 dB, Fig. 6, right panel), but at higher signal levels (and a precursor level of 20 dB), the precursor increased masker threshold (Fig. 7). For large Δf's, the precursor decreased masker threshold substantially at the lowest signal level, but this effect became smaller as signal level increased. For the 1-kHz signal, adding the precursor decreased masker threshold for $\Delta f = 0.2$ and greater, for both the low and mid signal levels, but did not affect threshold for smaller Δf's (Figs. 5 and 6, left panel). This broadening in tuning is similar to that seen following exposure to intense sound (Feth *et al.*, 1979) or ingestion of aspirin (Beveridge and Carlyon, 1996; Hicks and Bacon, 1999b). Beveridge and Carlyon used slightly different signal levels with and without aspirin, so the results cannot be compared exactly, but their data for simultaneous masking also seem to show a pattern of an increase in masker level for maskers near the signal frequency, and a decrease for those farthest from the signal frequency.

Both the change in threshold and the broadening in tuning after a precursor would be consistent with a decrease in active processing. In order to explain the various results, it would be helpful to consider one aspect of active processing, the nonlinear input/output function on the basilar membrane. The input/output function at the base of the cochlea, where high frequencies are represented, has been well measured. In a healthy cochlea, the growth of response to a tone presented at characteristic frequency (CF) for the place of measurement has a relatively high slope at low and high input levels, and has a shallower slope at mid input levels (Rhode, 1971; Ruggero and Rich, 1991; Ruggero *et al.*, 1997). After death, the response to the same tone becomes linear, with a slope of 1. The response to tones above or below CF is shifted to the right (i.e., a higher input level is necessary to achieve the same output), and becomes more linear as the frequency diverges from CF.

An approximation to the basilar-membrane input/output function for a tone at CF was used by Moore *et al.* (1996) to fit data for detection of increments and decrements as a function of level. A modification of this function is shown in Eq. (3):

$$l_{\text{out}} = (0.5 + (x \times 0.1)) \times l_{\text{in}} + 50 - (x \times 10) + 8.5 \times \left(1 - \left(\frac{1}{1 + \exp(-0.09 \times (l_{\text{in}} - 60))}\right)\right), \quad (3)$$

where l_{in} and l_{out} are input and output levels in dB, and x is a free parameter. Moore *et al.* (1996) found that $x = 0$ would fit results for signal frequencies from 500 to 4000 Hz, but that x had to be increased to 2 (i.e., the function was made more linear) to fit results for a signal frequency of 250 Hz. Therefore, this function with $x = 0$ will be used for the results at 4 kHz with no precursor. This function is shown by the solid line in Fig. 9. For comparison, a linear response is shown by the diagonal dotted line. It can be seen that in a

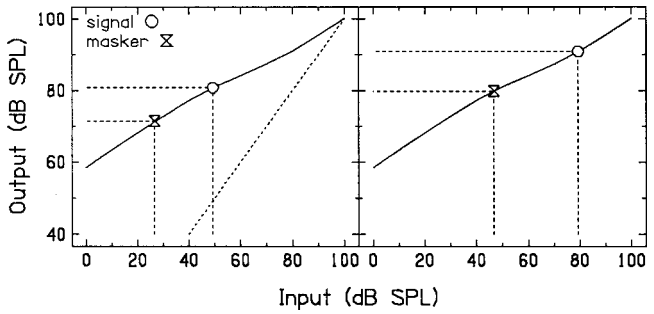


FIG. 9. Schematic showing the relationship between the input/output function on the basilar membrane and overshoot. The solid line is Eq. (3) when $x=0$. The dotted function is linear. The input signal and masker levels are calculated from the average data shown in Fig. 3. The output levels were calculated using Eq. (3). In the left panel, the masker spectrum level was 0 dB, and in the right panel, 20 dB.

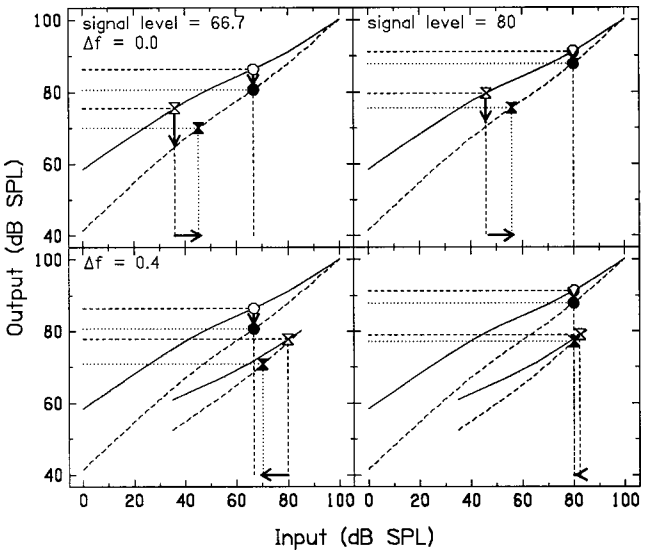


FIG. 10. Schematic showing the effects of a change in the input/output function on the masker level needed to mask a fixed-level signal. The signal is shown by the circles, the masker by the hourglasses, open symbols are data with no precursor, and filled symbols are data with a precursor. Input signal and masker levels are calculated from the average of the data shown in Fig. 7. The upper solid line is Eq. (3) when $x=0$. The upper dashed line is Eq. (3) when $x=1.69$, the best fit to the data. The solid and dashed lines at the bottom are the input/output function for the masker. See the text for details. Vertical arrows show the effect of the precursor. Horizontal arrows show the change in masker level with a precursor from Fig. 7.

healthy cochlea the response is amplified at low input levels, amplified less at mid levels, and not amplified at all as it approaches high levels.

Using this figure, it is possible to illustrate how overshoot is related to active processing. In modeling the results, the approach taken will be to begin with the results at 4 kHz, begin with the conditions in which the fewest assumptions need to be made, and use the results from these conditions to proceed to conditions which require more assumptions. Then, the results at 1 kHz will be considered. As mentioned above, the basilar-membrane input/output functions which are being approximated in the modeling are responses to tones, either at CF for a given place or away from CF. Since the present experiment used a noise masker, some assumptions will have to be made about the effective level of the noise in different frequency regions. It is also assumed that the listener is listening at CF, which is probably an oversimplification.

The first result to be modeled is why the signal-to-masker ratio for a 4-kHz signal at masker onset is larger at mid masker levels than at low or high masker levels. It is assumed that the threshold level for a tone in the masker corresponds to some criterion change on the y axis, which is in arbitrary output units. Although this change would be between the masker alone and the signal plus the masker, for simplicity responses are presented for the signal alone, because the addition of the masker would cause a negligible change in the level. The input values in the figure are taken from Fig. 3, while the output values are calculated using Eq. (3), with $x=0$. The mean threshold for the 4-kHz tone in a masker with a spectrum level of 0 dB, and no precursor, is 49.3 dB SPL (Fig. 3, shown by the circle in the left panel of Fig. 9). Assuming that frequencies near 4 kHz in the noise are masking the signal, the masker would have the same input/output function as the signal. The effective level of the noise was estimated as that which would pass through a filter centered at 4 kHz. Assuming an equivalent rectangular bandwidth (ERB) of 456 Hz (Glasberg and Moore, 1990), the effective level of the noise would be 26.6 dB SPL (shown by the hourglass in the left panel).[2] At this masker level, the signal must be 22.7 dB above the masker (difference between dashed lines on the abscissa) to achieve the threshold change in output (difference between dashed lines on the ordinate). This threshold change turns out to be about 10 dB, so this will be the criterion for all subsequent modeling. The large difference between the signal and masker input level is due to the short duration of the signal. In the right panel, results are shown when the masker level is increased by 20 dB. Now, the input signal level must be more than 30 dB above the masker in order to produce the same 10-dB change in output, because of the decrease in slope of the input/output function at this signal level. As was suggested by von Klitzing and Kohlrausch (1994), this would explain why there is an elevation in signal threshold at masker onset for midlevel maskers for a 4-kHz signal. Because real data values were used, note that the output increment is slightly larger in the right panel than the left. The function actually fits the average data quite well. The output increment was on average 10 dB, and varied by less than 1 dB across levels.

The next question to be addressed is why there is overshoot; that is, why does the signal-to-masker ratio decrease at these mid levels when there is a precursor? From the above explanation, this would be consistent with an increase in the slope of the input/output function between the signal and masker for mid masker levels. It is known from physiological data that this type of change occurs when active processing is decreased (Murugasu and Russell, 1995; Ruggero *et al.*, 1996), and von Klitzing and Kohlrausch (1994) suggested that a decrease in active processing might be responsible for the overshoot effect.

This may be examined using data from Fig. 7. Recall that this shows data collected with no precursor and with a precursor fixed at a spectrum level of 20 dB, for two fixed signal levels. In Fig. 10, the effects of a decrease in active processing are schematized, using data from selected condi-

tions from Fig. 7. In the top row, the effects are shown for a masker with $\Delta f = 0.0$. As above, the signal and the masker both have the same input/output function, assuming that the frequencies near 4 kHz in the noise are masking the signal. The solid line is the input/output function with no precursor, and is again Eq. (3), with $x = 0$. The dashed line represents the more linear input/output function which might be in effect after a precursor, with $x = 1.69$. This was obtained by adjusting x in Eq. (3) to find the best least-squares fit to the average data in Fig. 6 for $\Delta f = 0.0$ and 0.4.[3] In the left column, the input signal level is fixed at 66.7 dB SPL, as in the left column of Fig. 7. As in Fig. 9, the output values are calculated using Eq. (3). The response to the signal with no precursor is shown by the open circle. If the effect of the precursor is to turn down the active process, it would turn down the response to the signal, as shown by the arrow and the filled circle. The masker spectrum level at threshold with no precursor was 9.36 dB, which, assuming the same ERB as above, would produce an input of 35.96 dB SPL on this function. This is shown by the open hourglass. The precursor would reduce the response to the masker more than the signal, as shown by the longer arrow. Therefore, the masker level would need to be increased to achieve the criterion difference in output level between the signal and the masker, as shown by the horizontal arrow. The resulting masker level is shown by the filled hourglass. As in Fig. 9, the difference in output is shown by the difference in matching dotted lines on the ordinate, and should be approximately equal to 10 dB with or without a precursor. In the upper right panel, the effects are schematized for $\Delta f = 0.0$ for a signal level of 80 dB SPL. At this signal level, the response to the signal is only reduced by a few dB, but the response to the masker is still reduced more, and therefore the masker level must still be increased.

The above results show that the increase in masker level when $\Delta f = 0.0$ would be consistent with a decrease in active processing. If this explanation is correct, then it should also explain the effects for the other Δf's. In particular, it should explain why the precursor causes a *decrease* in masker level when $\Delta f = 0.4$. In the bottom row, the effects are schematized for a masker with $\Delta f = 0.4$. In this case, the signal would still have the same nonlinear input/output function, but the masker would have a different input/output function because the closest masker components would be at 2.4 kHz and at 5.6 kHz. Data from Ruggero *et al.* (1997, Fig. 8) show that the response to tones is approximately equal across the whole range of frequencies in the noise band above the signal (relative frequencies of 1.4 to 1.8). For the frequencies in the noise band below the signal, the response to tones is approximately equal across the range from 0.4 to 0.6, or half the noise band (the total lower band extends from 0.2 to 0.6 in relative frequency). Since we do not know which side dominates in masking the signal, for simplicity the masker input value used is the overall level of the masker at threshold from Fig. 7. From the above argument, at most this estimate would be about 6 dB too high, if the lower band dominates in masking the signal. When $\Delta f = 0.4$, the combined bandwidth of the two noise bands is 3.2 kHz, so the overall level is 35.04 dB above the threshold from Fig. 7. The

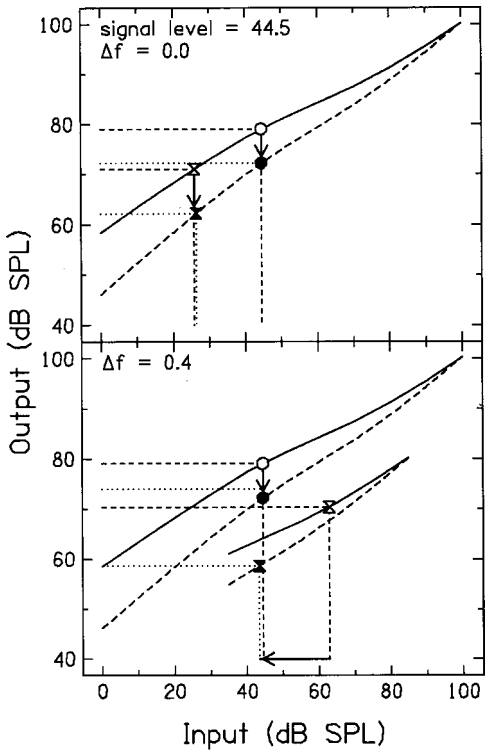


FIG. 11. As in Fig. 10, for data from Fig. 6.

masker function was chosen so that the difference between the signal and masker output values was close to 10 dB for the two signal input levels in Fig. 10, and also for a signal input level of 44.5 dB SPL, which will be discussed later. For simplicity, the input/output functions used were the same as those for the signal, but shifted down 20 dB and to the left by 15 dB.[4] When the signal level is 66.7 dB SPL, in the left panel, the precursor would turn down the response to the signal, as schematized by the open and filled circles. The response to the masker, shown by the hourglasses, would be turned down much less. Because of the decrease in the signal response, the masker would need to be turned down to achieve the same change in output level between the signal and masker. In the right panel, when the signal level is 80 dB SPL, the precursor turns down the response to the signal very little, and has almost no effect on the masker. Therefore the masker would only need to be turned down by a small amount, which is consistent with the data for L3 in Fig. 7. This accounts for the results at the extreme values of Δf.

This model suggests, then, that the precursor turns down active processing, and the effect on masker threshold depends on where the signal and masker are on their respective input/output functions. If the masker is turned down more than the signal, masker threshold will increase. If the signal is turned down more than the masker, masker threshold will decrease. For intermediate values of Δf, the signal and masker would be affected more equally by the precursor, so the change in masker threshold would be smaller.

To show that the model is successful for other precursor levels, in Fig. 11 the results of Fig. 6 are schematized. In this condition, the signal level was 44.5 dB SPL, and the precursor spectrum level was 0 dB. Because the precursor level is lower, it might cause less of a decrease in the active process.

The dashed line was fit so that the signal-to-masker ratio at threshold with a precursor was at the criterion of 10 dB found in Fig. 9 ($x=1.23$). In the top panel, $\Delta f=0.0$. Note that there was very little overshoot in this condition. The precursor turns down active processing, but the slope of the input/output function between the signal and the masker levels changes very little. Therefore, the lack of overshoot does not mean that the active process is not affected, it simply means that the signal and the masker are about equally affected. In the bottom panel, $\Delta f=0.4$, and the masker input levels were calculated as in the previous figure. In this case, the signal is turned down more than the masker, and thus, the masker threshold level decreases.

In terms of the effect of the precursor on tuning, the results for 4 kHz at the lower signal level are very similar to those for 1 kHz for both signal levels. Thus, the results for 4 kHz may be used to explain the results for 1 kHz. The change in signal-to-masker ratio with signal level with no precursor is much smaller for the 1-kHz signal than the 4-kHz one. This suggests that there is less change in the slope of the input/output function with level when there is no precursor present. This would be consistent with less active processing at 1 kHz. It has been shown that the input/output function is more linear in apical turns of the cochlea, corresponding to lower frequencies (Cooper and Yates, 1994; Cooper and Rhode, 1995; Rhode and Cooper, 1996). When $\Delta f=0.0$, the precursor could still decrease active processing, but as long as it turns down the signal and the masker by the same amount, as in Fig. 11, the masker threshold will remain the same, that is, there will be little or no overshoot. When $\Delta f=0.4$, the situation is the same as in Fig. 11. The precursor would turn down the signal, but affect the masker little if at all. Therefore, the threshold masker level would decrease.

As shown in Fig. 3, the average threshold for the 4-kHz signal with no masker was 26.01 dB SPL, and following a precursor of 20 dB was 51.31 dB SPL. These values were approximately the same for the 1-kHz signal. Therefore the difference in results at 1 and 4 kHz is not due to a difference in the threshold shift produced by the precursor. Interestingly, if the quiet threshold for 4 kHz is plotted on the solid line in Fig. 10, and the threshold following the precursor is plotted on the dashed line, the output values are almost equal. Nearly the same output value is found if the threshold for the 4-kHz signal following a precursor of 0 dB (which was 39.31 dB SPL) is plotted on the dashed function in Fig. 11. This suggests that the change in quiet threshold for the signal produced by the precursor (usually called forward masking) could largely be due to the decrease in the active process. This is consistent with data showing changes in forward masking with cochlear hearing loss (Oxenham and Plack, 1997; Nelson and Schroder, 1999).

The fact that the precursor does elevate threshold for the signal means that the data could be interpreted as a case of additivity of masking, although this has usually been studied either when both maskers are simultaneous with the signal, or both are nonsimultaneous with the signal. Consider the data in the upper left panel of Fig. 7, for $\Delta f=0.4$. With no precursor, the masker threshold is 44 dB SPL. With the precursor, the masker threshold is 36 dB SPL. This is true even though the signal is at a level that is above masked threshold with the precursor alone. This type of result, when combining two maskers produces a greater effect than would be expected from just combining their energy, has been called ''excess masking'' in the additivity of masking literature. As shown in the lower-left panel of Fig. 10, the interpretation in the present study would be that the precursor turns down the signal more than the masker. This would be consistent with some recent research on additivity of masking linking excess masking with active processing (Oxenham and Moore, 1995b).

This explanation, that auditory stimulation decreases active processing, is also consistent with psychometric functions for short-duration signals above 2 kHz. Psychometric functions for signals at masker onset are shallower than those for which the signal is delayed from masker onset, when the masker is set at a level such that the signal threshold is between 35 and 75 dB SPL (Wright and Dai, 1994; Oxenham and Moore, 1995a). This would be consistent with the difference in slope at mid input levels between the solid function and the dashed function in Fig. 10. When the masker level is high enough that the signal level is 80 dB or above, psychometric functions are only slightly shallower for signals at masker onset, consistent with the smaller change in slope at this level in Fig. 10 (Oxenham and Moore, 1995a).

The overshoot data, the frequency selectivity data, and the psychometric function data would all be consistent with a decrease in the active process with acoustic stimulation by the precursor. It has been suggested that such a process could be mediated by the efferent system (Schmidt and Zwicker, 1991; Turner and Doherty, 1997). One part of this feedback system, the medial olivocochlear bundle, feeds back to the outer hair cells (Warr and Guinan, 1979; Warr, 1980). Activation of this system by contralateral noise decreases neural firing for tones at CF but not for tones in the tail of the tuning curve (Warren and Liberman, 1989a), suggesting that the active process is turned down. The time course of the onset and the offset of the efferent response (Warren and Liberman, 1989a) is consistent with the time course of overshoot (Bacon and Smith, 1991). The majority of the efferents feeding back to a given ear responds best to sound stimulation in that ear (Liberman, 1988), while a minority will respond to sound stimulation in the contralateral ear. The effects of sound stimulation have been studied mainly by contralateral stimulation, because of course with ipsilateral stimulation it is difficult to separate any possible efferent effects from excitatory effects of the sound. It has been demonstrated that overshoot may be reduced by a broadband contralateral precursor (Turner and Doherty, 1997; Bacon and Liu, 2000). Kidd and Wright (1994) also found a reduction in overshoot for some listeners using a contralateral notched-noise precursor. Although this effect has not been shown with a tonal precursor (Bacon and Healy, 2000), this would be consistent with the fact that the efferent effect is stronger when activated by broadband noise than by tones (Warren and Liberman, 1989b). Because the ipsilateral efferent effect is stronger than the contralateral one, it seems quite possible that reduction in overshoot would be larger with the precursor presented ipsilaterally, as it is. In addition, the fact

that in the present experiment the precursor also broadens tuning would be consistent with an efferent effect. Strickland and Viemeister (1995) also demonstrated a broadening in the psychophysical tuning curve when broadband noise was presented contralaterally at a spectrum level of 30 dB.

Some additional evidence comes from overshoot data measured in humans in whom the efferent bundle has been cut. Zeng *et al.* (2000) found a reduction in overshoot between the surgery ear and the nonsurgery ear in all five listeners tested. In most cases, this was due to a decrease in threshold at masker onset that was consistent with hearing loss in the surgery ear. In two cases in which hearing was normal or near normal in the surgery ear, the decrease in overshoot was due to an increase in threshold for the signal delayed from masker onset, in the surgery ear, which would be consistent with a decrease in efferent activity in this ear. This is in contrast to data of Scharf *et al.* (1997), who found a small decrease in overshoot in two patients that was due to a decrease in threshold for the signal at masker onset.

Overshoot is clearly a complex phenomenon involving peripheral and central mechanisms, and this research does not purport to explain all aspects of it in detail. It does add to a growing body of evidence that overshoot is related to active processing in the cochlea, and specifically, that active processing may actually be turned down as the cochlea is stimulated by sound. It would be logical to ask why other research has shown that tuning seems to increase with stimulation of the cochlea. If the feedback to the cochlea is relatively frequency specific, which appears to be the case (Warren and Liberman, 1989b), then the frequency region of the cochlea which is stimulated is very important. In several previous studies, threshold has been measured as a function of signal delay from the onset of the masker. When the masker frequency is near the signal frequency, feedback mechanisms would tend to affect the masker and signal equally. As the masker frequency departs from the signal frequency, feedback mechanisms might act to turn down active processing in the masker frequency region more than in the signal frequency region. Thus, the masker level would have to be turned up when the signal is delayed from the masker onset, making tuning appear sharper. This may also be the case for long-duration signals. Obviously, other more complex effects such as suppression would also be acting. The approach taken in this paper, therefore, provides a coherent model grounded in physiological and psychophysical evidence that should provide an interpretation not only of overshoot, but also of many other psychoacoustic phenomena.

V. CONCLUSIONS

(1) Growth-of-masking functions are nonlinear for a 4-kHz signal with no precursor, showing a growth steeper than 1 at low levels, and then a decrease at high levels. Growth-of-masking functions for a 4-kHz signal with a precursor, and a 1-kHz signal with or without a precursor, are essentially linear with a slope slightly greater than 1.

(2) Frequency selectivity, measured with a fixed signal level using the notched-noise technique, is highest at 4 kHz with no precursor, somewhat lower at 1 kHz with no precursor, and lowest for a 1- or 4-kHz signal with a precursor.

(3) These results would be consistent with a system in which active processing is stronger at 4 kHz than at 1 kHz, and active processing is decreased by a broadband noise precursor.

ACKNOWLEDGMENTS

I wish to thank Tamara Scott for assistance in data collection, and Beverly Wright for stimulating discussions about the model. I also wish to thank the reviewers and editor for their thorough comments. This research was partially supported by a grant from NIH (No. R03 DC03510).

[1]Preliminary measurements with L2 in which the overlap between the precursor and masker was varied determined that a 5-ms overlap produced the same signal thresholds as a 400-ms masker with the signal delayed 202 ms from masker onset. With zero overlap, thresholds were 6 dB higher with the precursor than with a long delay. The agreement between thresholds with a precursor with a 5-ms overlap, and thresholds with a continuous masker, was confirmed with L3 and a previous set of listeners. Due to time constraints, L3 only had a complete set of data with the continuous masker. Therefore, for the precursor conditions in the overshoot experiment, the data for this listener were actually measured with a continuous masker.

[2]The ERB could also have been estimated from the filter fits in the present experiment. This would have resulted in a change in level of at most 2 dB.

[3]The assumption made for the $\Delta f = 0.0$ data was that the signal-to-masker ratio at threshold was 10 dB, because this was true with no precursor. The assumption made for the $\Delta f = 0.4$ data was that the change in signal threshold with the precursor should be equal to the change in masker threshold; i.e., that the response to the masker was linear with a slope of 1, and the precursor did not affect it.

[4]A linear function with a slope of 1 would fit the data for the two higher masker levels, so the assumptions used in fitting x in Fig. 9 are still valid. The function used to fit the maskers for $\Delta f = 0.4$ is similar to functions measured on the basilar membrane for frequencies above CF (Ruggero *et al.*, 1997).

Bacon, S. P. (**1990**). ''Effect of masker level on overshoot,'' J. Acoust. Soc. Am. **88**, 698–702.

Bacon, S. P., and Healy, E. W. (**2000**). ''Effects of ipsilateral and contralateral precursors on the temporal effect in simultaneous masking with pure tones,'' J. Acoust. Soc. Am. **107**, 589–1597.

Bacon, S. P., and Liu, L. (**2000**). ''Effects of ipsilateral and contralateral precursors on overshoot,'' J. Acoust. Soc. Am. **108**, 1811–1818.

Bacon, S. P., and Moore, B. C. J. (**1986**). ''Temporal effects in simultaneous pure-tone masking: Effects of signal frequency, masker/signal frequency ratio, and masker level,'' Hear. Res. **23**, 257–266.

Bacon, S. P., and Smith, M. A. (**1991**). ''Spectral, intensive, and temporal factors influencing overshoot,'' Q. J. Exp. Psychol. **43A**, 373–399.

Bacon, S. P., and Takahashi, G. A. (**1992**). ''Overshoot in normal-hearing and hearing-impaired subjects,'' J. Acoust. Soc. Am. **91**, 2865–2871.

Bacon, S. P., and Viemeister, N. F. (**1985**). ''Simultaneous masking by gated and continuous sinusoidal maskers,'' J. Acoust. Soc. Am. **78**, 1220–1230.

Beveridge, H. A., and Carlyon, R. P. (**1996**). ''Effects of aspirin on human psychophysical tuning curves in forward and simultaneous masking,'' Hear. Res. **99**, 110–118.

Carlyon, R. P. (**1989**). ''Changes in the masked thresholds of brief tones produced by prior bursts of noise,'' Hear. Res. **41**, 223–235.

Carlyon, R. P., and Butt, M. (**1993**). ''Effects of aspirin on human auditory filters,'' Hear. Res. **66**, 233–244.

Carlyon, R. P., and White, L. J. (**1992**). ''Effect of signal frequency and masker level on the frequency regions responsible for the overshoot effect,'' J. Acoust. Soc. Am. **91**, 1034–1041.

Champlin, C. A., and McFadden, D. (**1989**). ''Reductions in overshoot following intense sound exposures,'' J. Acoust. Soc. Am. **85**, 2005–2011.

Cooper, N. P., and Rhode, W. S. (**1995**). ''Nonlinear mechanics at the apex of the guinea-pig cochlea,'' Hear. Res. **82**, 225–243.
Cooper, N. P., and Yates, G. K. (**1994**). ''Nonlinear input–output functions derived from the responses of guinea-pig cochlear nerve fibers: Variations with characteristic frequency,'' Hear. Res. **78**, 221–234.
Elliott, L. L. (**1965**). ''Changes in the simultaneous masked threshold of brief tones,'' J. Acoust. Soc. Am. **38**, 738–746.
Elliott, L. L. (**1967**). ''Development of auditory narrow-band frequency contours,'' J. Acoust. Soc. Am. **42**, 143–153.
Evans, E. F., Wilson, J. P., and Borerwe, T. A. (**1981**). ''Animal models of tinnitus,'' in *CIBA Foundation Symposium 85, Tinnitus* (Pitman, London), pp. 108–209.
Feth, L. L., Burns, E. M., Kidd, G., Jr., and Mason, C.R. (**1980**). ''Effects of noise exposure on frequency selectivity in normal and hearing-impaired listeners,'' 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. 149–152.
Glasberg, B. R., and Moore, B. C. J. (**1990**). ''Derivation of auditory filter shapes from notched-noise data,'' Hear. Res. **47**, 103–138.
Green, D. M. (**1969**). ''Masking with continuous and pulsed sinusoids,'' J. Acoust. Soc. Am. **46**, 939–946.
Hant, J. J., Strope, B. P., and Alwan, A. A. (**1997**). ''A psychoacoustic model for the noise masking of plosive bursts,'' J. Acoust. Soc. Am. **101**, 2789–2802.
Hicks, M. L., and Bacon, S. P. (**1992**). ''Factors influencing temporal effects with notched-noise maskers,'' Hear. Res. **64**, 123–132.
Hicks, M. L., and Bacon, S. P. (**1999a**). ''Psychophysical measures of auditory nonlinearities as a function of frequency in individuals with normal hearing,'' J. Acoust. Soc. Am. **105**, 326–338.
Hicks, M. L., and Bacon, S. P. (**1999b**). ''Effects of aspirin on psychophysical measures of frequency selectivity, two-tone suppression, and growth of masking,'' J. Acoust. Soc. Am. **106**, 1436–1451.
Kidd, G., Jr., and Wright, B. A. (**1994**). ''Improving the detectability of a brief tone in noise using forward and backward masker fringes: Monotic and dichotic presentations,'' J. Acoust. Soc. Am. **95**, 962–967.
Kimberley, B. P., Nelson, D. A., and Bacon, S. P. (**1989**). ''Temporal overshoot in simultaneous-masked psychophysical tuning curves from normal and hearing-impaired listeners,'' J. Acoust. Soc. Am. **85**, 1660–1665.
Levitt, H. (**1971**). ''Transformed up–down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.
Liberman, M. C. (**1988**). ''Response properties of cochlear efferent neurons: Monaural vs binaural stimulation and the effects of noise,'' J. Neurophysiol. **60**, 1779–1798.
McFadden, D., and Champlin, C. A. (**1990**). ''Reductions in overshoot during aspirin use,'' J. Acoust. Soc. Am. **87**, 2634–2642.
Moore, B. C. J., Peters, R. W., and Glasberg, B. R. (**1996**). ''Detection of decrements and increments in sinusoids at high overall levels,'' J. Acoust. Soc. Am. **99**, 3669–3677.
Moore, B. C. J., Poon, P. W. F., Bacon, S. P., and Glasberg, B. R. (**1987**). ''The temporal course of masking and the auditory filter shape,'' J. Acoust. Soc. Am. **81**, 1873–1880.
Murugasu, E., and Russell, I. J. (**1995**). ''Salicylate ototoxicity: The effects on basilar membrane displacement, cochlear microphonics, and neural responses in the basal turn of the guinea pig cochlea,'' Aud. Neurosci. **1**, 139–150.
Nelson, D. A., and Schroder, A. C. (**1999**). ''Forward masking recovery and peripheral compression in normal-hearing and cochlear-impaired ears,'' J. Acoust. Soc. Am. **106**, 2176–2177(A).
Overson, G. J., Bacon, S. P., and Webb, T. M. (**1996**). ''The effect of level and relative frequency region on the recovery of overshoot,'' J. Acoust. Soc. Am. **99**, 1059–1065.
Oxenham, A. J., and Moore, B. C. J. (**1995a**). ''Additivity of masking in normally hearing and hearing-impaired subjects,'' J. Acoust. Soc. Am. **98**, 1921–1934.
Oxenham, A. J., and Moore, B. C. J. (**1995b**). ''Overshoot and the 'severe departure' from Weber's law,'' J. Acoust. Soc. Am. **97**, 2442–2453.
Oxenham, A. J., and Plack, C. J. (**1997**). ''A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing,'' J. Acoust. Soc. Am. **101**, 3666–3675.
Patterson, R. D., Nimmo-Smith, I., Weber, D. L., and Milroy, R. (**1982**). ''The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold,'' J. Acoust. Soc. Am. **72**, 1788–1803.
Rhode, W. S. (**1971**). ''Observations of the vibrations of the basilar membrane in squirrel monkeys using the Mossbauer technique,'' J. Acoust. Soc. Am. **49**, 1218–1231.
Rhode, W. S., and Cooper, N. P. (**1996**). ''Nonlinear mechanics in the apical turn of the chinchilla cochlea *in vivo*,'' Aud. Neurosci. **3**, 101–121.
Rosen, S., Baker, R. J., and Darling, A. (**1998**). ''Auditory filter nonlinearity at 2 kHz in normal hearing listeners,'' J. Acoust. Soc. Am. **103**, 2539–2550.
Ruggero, M. A., and Rich, N. C. (**1991**). ''Furosemide alters organ of Corti mechanics: evidence for feedback of outer hair cells upon the basilar membrane,'' J. Neurosci. **11**, 1057–1067.
Ruggero, M. A., Rich, N. C., Recio, A., Narayan, S. S., and Robles, L. (**1997**). ''Basilar-membrane responses to tones at the base of the chinchilla cochlea,'' J. Acoust. Soc. Am. **101**, 2151–2163.
Ruggero, M. A., Rich, N. C., Robles, L., and Recio, A. (**1996**). ''The effects of acoustic trauma, other cochlear injury, and death on basilar-membrane responses to sound,'' in *Scientific Basis of Noise-Induced Hearing Loss*, edited by A. Axelsson, H. Borchgrevink, R. P. Hamernik, P.-A. Hellstrom, D. Henderson, and R. Salvi (Thieme Medical, New York), pp. 23–35.
Scharf, B., Magnan, J., and Chays, A. (**1997**). ''On the role of the olivocochlear bundle in hearing: 16 case studies,'' Hear. Res. **103**, 101–122.
Scholl, H. (**1962**). ''Uber die Bildung der Horschwellen und Mithorschwellen von Impulsen,'' Acustica **12**, 91–101.
Schmidt, S., and Zwicker, E. (**1991**). ''The effect of masker spectral asymmetry on overshoot in simultaneous masking,'' J. Acoust. Soc. Am. **89**, 1324–1330.
Smith, R. L., and Zwislocki, J. J. (**1975**). ''Short-term adaptation and incremental responses in single auditory-nerve fibers,'' Biol. Cybern. **17**, 169–182.
Strickland, E. A., and Viemeister, N. F. (**1995**). ''An attempt to find psychophysical evidence for efferent action in humans,'' Abstracts of the Eighteenth Midwinter Research Meeting of the Association for Research in Otolaryngology, 173(A).
Turner, C. W., and Doherty, K. A. (**1997**). ''Temporal masking and the 'active process' in normal and hearing-impaired listeners,'' in *Modeling Sensorineural Hearing Loss*, edited by W. Jesteadt (Erlbaum, Hillsdale, NJ), pp. 387–396.
von Klitzing, R., and Kohlrausch, A. (**1994**). ''Effect of masker level on overshoot in running- and frozen-noise maskers,'' J. Acoust. Soc. Am. **95**, 2192–2201.
Warr, W. B. (**1980**). ''Efferent components of the auditory system,'' Ann. Otol. Rhinol. Laryngol. **89** (Suppl. 74), 114–120.
Warr, W. B., and Guinan, Jr., J. J. (**1979**). ''Efferent innervation of the organ of Corti: Two separate systems,'' Brain Res. **173**, 152–155.
Warren III, E. H., and Liberman, M. C. (**1989a**). ''Effects of contralateral sound on auditory-nerve responses. I. Contributions of cochlear efferents,'' Hear. Res. **37**, 89–104.
Warren III, E. H., and Liberman, M. C. (**1989b**). ''Effects of contralateral sound on auditory-nerve responses. II. Dependence on stimulus variables,'' Hear. Res. **37**, 105–121.
Wightman, F. L., McGee, 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. 295–306.
Wright, B. A. (**1989**). ''Auditory-filter shapes at three test frequencies with a burst and a fringed notched-noise masker,'' J. Acoust. Soc. Am. **86**, S121.
Wright, B. A., and Dai, H. (**1994**). ''Detection of unexpected tones in gated and continuous maskers,'' J. Acoust. Soc. Am. **95**, 939–948.
Zeng, F.-G., Martino, K. M., Linthicum, F. H., and Soli, S. D. (**2000**). ''Auditory perception in vestibular neurectomy subjects,'' Hear. Res. **142**, 102–112.
Zwicker, E. (**1965**). ''Temporal effects in simultaneous masking by white-noise bursts,'' J. Acoust. Soc. Am. **37**, 653–663.

The lower limit of melodic pitch

Daniel Pressnitzer,[a] Roy D. Patterson, and Katrin Krumbholz
Centre for the Neural Basis of Hearing, Department of Physiology, Downing Street, Cambridge CB2 3EG, United Kingdom

(Received 10 February 2000; revised 7 August 2000; accepted 6 February 2001)

An objective melody task was used to determine the lower limit of melodic pitch (LLMP) for harmonic complex tones. The LLMP was defined operationally as the repetition rate below which listeners could no longer recognize that one of the notes in a four-note, chromatic melody had changed by a semitone. In the first experiment, the stimuli were broadband tones with all their components in cosine phase, and the LLMP was found to be around 30 Hz. In the second experiment, the tones were filtered into bands about 1 kHz in width to determine the influence of frequency region on the LLMP. The results showed that whenever there was energy present below 800 Hz, the LLMP was still around 30 Hz. When the energy was limited to higher-frequency regions, however, the LLMP increased progressively, up to 270 Hz when the energy was restricted to the region above 3.2 kHz. In the third experiment, the phase relationship between spectral components was altered to determine whether the shape of the waveform affects the LLMP. When the envelope peak factor was reduced using the Schroeder phase relationship, the LLMP was not affected. When a secondary peak was introduced into the envelope of the stimuli by alternating the phase of successive components between two fixed values, there was a substantial reduction in the LLMP, for stimuli containing low-frequency energy. A computational auditory model that extracts pitch information with autocorrelation can reproduce all of the observed effects, provided the contribution of longer time intervals is progressively reduced by a linear weighting function that limits the mechanism to time intervals of less than about 33 ms. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1359797]

PACS numbers: 43.66.Ba, 43.66.Hg, 43.66.Nm [DWG]

I. INTRODUCTION

A periodic click train produces a strong pitch when the click repetition rate is 100 Hz; however, when the rate is 10 Hz or less, there is no pitch and the individual clicks are heard as separate events. In the transition region, as the pitch percept fades away, the periodicity can still be detected as roughness, pulsation, or flutter (Guttman and Julesz, 1963; Terhardt, 1970; Warren and Bashford, 1981). Pitch differs from the other percepts inasmuch as it alone can convey information about musical intervals and thus, melodies (Plomp, 1976; Moore and Rosen, 1979; Dowling and Harwood, 1986; Houtsma, 1995; Griffiths *et al.*, 1998). The purpose of this study is to delineate the lower boundary of the region where pitch will support melodic patterns similar to those used in Western music.

In an influential study, Ritsma (1962) investigated the existence region of pitch for three-component harmonic complexes, specifically sinusoidally amplitude-modulated (SAM) tones. His data suggest that the lowest repetition rate[1] that produces a pitch is around 40 Hz for a carrier frequency of 150 Hz. In Ritsma's experiment, as the carrier frequency increases, the lower limit of pitch rises to a value of 350 Hz for a carrier frequency of 4.7 kHz. A replication of the study by Moore (1973) confirmed the basic findings. Both studies employed subjective judgments about the presence or absence of a pitch cue.

Ritsma (1971) and Ritsma and Hoekstra (1974) introduced an objective method to investigate the existence region of pitch. They measured rate discrimination threshold (RDT) across the lower boundary of the existence region defined by Ritsma (1962). They reported small RDTs for sounds that were inside the pitch region and large RDTs for sounds outside the pitch region. They concluded that the transition from a small to a large RDT revealed the limit of pitch (Ritsma and Hoekstra, 1974). Recently, in a companion paper, the correspondence between rate discrimination and the lower limit of pitch was re-examined with newer techniques (Krumbholz *et al.*, 2000). The results broadly confirmed the correspondence reported previously. Houtsma and Smurszynski (1990) have questioned the link between RDT and pitch; they showed that harmonic sounds in the region where the RDT is large, nevertheless support musical interval recognition (albeit with reduced accuracy). They suggest that the increase in RDT reflects the transition between the regions of resolved and unresolved harmonics (Shackleton and Carlyon, 1994; Carlyon and Shackleton, 1994) rather than the boundary of pitch perception.

Goldstein (2000) has reviewed evidence showing that the pitch of harmonic complex tones may, in fact, have different components or modes. Schouten (1940) introduced the term ''residue'' to characterize the pitch sensation produced by unresolved frequency components. de Boer (1976) proposed to generalize the use of the term residue to unresolved and resolved components, as the latter seemed to dominate the former in pitch perception (Plomp, 1967). Other terms to describe the pitch sensation corresponding to the missing fundamental of harmonic complex tones include ''low

[a]Author to whom correspondence should be addressed; IRCAM-CNRS, 1 place Stravinsky, 75004 Paris, France; electronic mail: pressnit@ircam.fr

pitch'' (Smoorenburg, 1970) or ''periodicity pitch'' (Terhardt, 1970). Guttman and Pruzansky (1962) suggested that in the case of complex harmonic tones, we should further distinguish between ''pitch'' as described in the American National Standard (the sensation that enables us to order notes on a scale from low to high), and ''musical pitch.'' By ''musical pitch'' they meant a sensation that can be used to convey musical values like diatonic intervals. They reported lower limits of 19 Hz using a subjective criterion like that of Ritsma (1962), and 60 Hz using an objective, octave-matching task. Unfortunately, their octave-matching task requires judgments that are difficult for listeners that are not musically trained. Patterson *et al.* (1983) introduced a melody-change task that involves pitch in a musical context but is much easier to perform. They used the technique to investigate the duration that complex tones need to support pitch.

In this paper, the melody-change task is adapted to determine the lower limit of melodic pitch (LLMP). The term ''melodic'' is introduced to emphasize that the experimental task provides an operational definition of pitch. Links between pitch, musical pitch, and melodic pitch will be discussed in the latter sections of the paper. In the LLMP task, listeners are required to detect a semitone change in a four-note random melody based on the chromatic scale. The range is restricted to 4 semitones; this enables the production of a sufficient number of random melodies while focusing on a limited range of repetition rates. The LLMP task has several advantages. First, it has face validity; melodies are the most fundamental elements of Western music and the semitone is the basic pitch interval of the Western chromatic scale. Second, the task is easy to perform; both musical and nonmusical listeners can perform the task whenever the notes produce a clear pitch. Parncutt and Cohen (1995) have shown that with a semitone change and an eight-note melody task, listeners reach asymptotic performance irrespective of musical education. The same is not the case for interval recognition and labelling tasks (Guttman and Pruzansky, 1962; Houtsma and Goldstein, 1972). Finally, the fact that the melodies are chosen at random minimizes the potential to use contour and knowledge-based cues that can play a part in the recognition of familiar melodies (Dowling and Fujitani, 1971; Patterson *et al.*, 1983).

II. EXPERIMENT I: THE LLMP FOR A BROADBAND, HARMONIC COMPLEX TONE

A. Rationale

The aim of the first experiment was to measure the LLMP for click trains; that is, the lowest repetition rate that enabled listeners to perform a melody task using broadband, harmonic complex tones. Click trains produce strong pitch percepts when the rate is as low as 100 Hz, the pitch of a low male voice. Below this, as the rate decreases, the pitch becomes weaker and eventually disappears to give way to the perception of isolated clicks. At this point, the melody task should become impossible because it requires the perception of pitch. By using a click train, the repetition rate can be lowered continuously while presenting energy across a large frequency range. This avoids confounding factors associated with the steep rise of the audiogram at low frequencies, which are problematic when studying pitch with low-frequency sinusoids. Also, the pitch strength of the three-component tones used by Ritsma (1962) is rather weak. Adding spectral components makes the task easier to perform (Patterson, 1973).

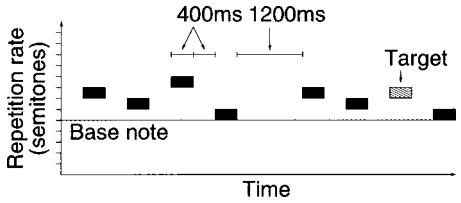


FIG. 1. Schematic of the melody task used in all experiments.

B. Method

1. Stimuli

The repetition rate of the harmonic complex tones was varied from 16 to 512 Hz in semitone steps (6%). Components that fell in the range 10 Hz–10 kHz were included. The components all started in cosine phase and so the sounds were essentially broadband click trains. Each tone was 400 ms long, and included 5-ms, squared-cosine on and off ramps. The silent interval between the tones within a melody was also 400 ms long. The stimuli were generated off-line by additive synthesis in the time domain. The overall presentation level of the broadband harmonic complex was 55 dB SPL.

The stimuli were generated with a 25-kHz sampling rate and presented using a TDT system II. The sound files containing the stimuli were stored on a PC disk. They were played back through a DD1 16-bit digital-to-analog converter, an FT-6 anti-aliasing filter with a 10-kHz cutoff, a PA4 attenuator, and a HB6 headphone buffer. The stimuli were presented diotically through a set of AKG K-240-DF headphones. The experiments took place in a double-walled, sound-insulated booth.

2. Procedure and listeners

A 4-alternative forced-choice (4AFC) task was used (Fig. 1). Each trial began with a short melody of four notes. The melody was characterized by the repetition rate, R_{rep}, of its base note. Given the base note, the melody was produced by drawing four values of R_{rep} randomly, with replacement, from the four semitones above the base note. This means that the melody could contain any note with a repetition rate equal to that of the base note, or 6%, 12%, 18%, or 24% higher than that of the base note. No other constraints were placed on the melody. In particular, any number of repetitions of the same note could occur randomly. In musical terms, the melodies were drawn from the chromatic scale rather than from a diatonic scale.

After a 1200-ms pause, the original melody was then repeated, but with one note changed plus or minus one semitone at random. The position of this target note was chosen at random. This means that the change could occur on the

TABLE I. Results for experiment I. The stimuli were broadband, harmonic complex tones. For individual subjects, the mean and standard deviation are based on the last six reversals of the adaptive runs.

	L1	L2	L3	Mean
Mean LLMP (Hz)	33.5	32.4	29.2	31.7
Standard deviation (Hz)	4.2	2.4	2.5	3.6

lowest or highest note of the melody and that there could be as many as five different values of R_{rep} in the two melodies. The change in repetition rate of the target note was the only difference between the two presentations of the melody. The listeners' task was to indicate the position of the target note by pressing one of four buttons on a response box. No time limit was imposed on listeners to give their response. Visual feedback was provided during a 1-s pause before proceeding to the next trial. If listeners were inattentive for a trial, they had the option of repeating the trial once with the same base note, but with a different melody and a new, random position of the target note.

A 3-down, 1-up adaptive threshold technique was used to track the LLMP (Levitt, 1971). Twelve reversals were measured. After three successive correct identifications of the target note, the repetition rate of the base note was lowered by 4 semitones for the first four reversals and 2 semitones for the last eight reversals. After each mistake, the repetition rate of the base note was increased by 3 semitones for the first four reversals and 1 semitone for the last eight reversals. The last six reversals were averaged to produce the threshold estimate for that run. Theoretically, this adaptive method converges during the last two-thirds of the track (which was the part that was analyzed) toward a probability for correct of $(1/3)^{1/3}$, i.e., the 69% correct point of the psychometric function. The starting base note for the adaptive run was 150 Hz. One complete practice run followed by two experimental runs were performed by each listener.

Three listeners aged 24 to 35 participated in the experiment. Listeners 1 and 2 had moderate musical training, listener 3 had no musical training whatsoever. Listener 1 was the first author. Listener 3 was paid for her participation. All had normal hearing thresholds ($<$15 dB HL) at standard audiometric frequencies.

C. Results and discussion

Results for the three listeners are shown in Table I. The LLMP for broadband harmonic tones is found to be between 30 Hz and 35 Hz.

These values are specific to the criterion used to define the LLMP, namely the 69% correct point when comparing two chromatic, four-note melodies. It is possible that listeners might have been able to use different cues on different trials. When the alteration of the target note changed the contour of the melody, the comparison could have been facilitated (Dowling and Fujitani, 1971). Primacy and recency effects in memory could also facilitate the task when the target note was at the beginning or the end of the melody (Crowder and Morton, 1969; McFarland and Cacace, 1992). Finally, the target note was sometimes well above the base-note used to define threshold. Nevertheless, since a 3-down, 1-up adaptive procedure was used, it is likely that these randomly occurring cues could not be used consistently, and that the only reliable strategy for the listeners was to store the pitch of the four notes in memory whenever possible.

It is also the case that the LLMP values compare well with previous values reported in the literature, even when obtained with very different experimental procedures. The LLMP is slightly lower than the lowest value reported by Ritsma (1962) for subjective perception of residue pitch (40 Hz). This is perhaps because broadband sounds were used instead of three-component complexes. Note also that if the middle note of the melody is chosen to define the LLMP, this small discrepancy vanishes. Guttman and Pruzansky (1962) found that listeners reported a pitch sensation for click trains with repetition rates as low as 19 Hz. However, when the same listeners performed an objective, octave-matching task, the limit was found to be 60 Hz. The LLMP value from the current study falls in between these two values. When analyzing their data with a fixed accuracy criterion (a semitone), Guttman and Pruzansky (1962) found that the limit was between 38 and 45 Hz, which is even closer to the value from the current study.

III. EXPERIMENT II: THE EFFECT OF FREQUENCY REGION

A. Rationale

The lower limit of the existence region of pitch increases with frequency region for SAM tones (Ritsma, 1962). Similarly, rate discrimination performance deteriorates when stimuli are limited to higher-frequency regions (Houtsma and Smurzynski, 1990; Krumbholz *et al.*, 2000). Accordingly, the influence of frequency region on the LLMP was investigated by bandpass filtering the harmonic complex tones of experiment I. Low-frequency regions were included since they are representative of the human voice, and many musical instruments include low-frequency energy.

B. Method

1. Stimuli

The bandpass filtering process was identical to Krumbholz *et al.* (2000). The filter had a nonattenuated section of constant width (600 Hz). The lower edge of this section is referred to as F_c, the filter cutoff. Five values of F_c were investigated: 200, 400, 800, 1600, and 3200 Hz. A frequency region with $F_c = 6400$ Hz was initially included but pilot data showed that the melody task was completely impossible with this high cutoff frequency. On both sides of the nonattenuated section, the filter had linear quarter-cosine skirts to minimize the effects of edge tones and to reduce the possibility of tracking individual harmonics. The lower skirt extended over 200 Hz and the upper skirt over 1 kHz. The repetition rate, R_{rep}, of the harmonic complex varied from 16 to 512 Hz in semitone steps. The stimuli were generated off-line by additive synthesis in the time domain. The filter was applied to the components during the additive synthesis. The overall presentation level of the harmonic complex was 55 dB SPL.

Continuous low-pass filtered pink noise was added to the stimuli before playback to mask any distortion products that might otherwise have been audible below F_c (Goldstein, 1967). The unfiltered pink noise had an overall level of 30 dB SL. It was filtered by an FIR filter with 158 taps, implemented by a TDT PF1. It was a brickwall-filter designed in the frequency domain with a cutoff at $F_c - 200$ Hz. The continuous pink noise was played from a DAT tape through a TDT PF1 filter and a PA4 attenuator. The noise and harmonic tones were then mixed by a SM3 summer.

2. Procedure and listeners

Thresholds were determined in separate adaptive runs for each filter condition. Within each filter condition, the procedure was the same as that of experiment I. The apparatus was also the same. The order of filter conditions was varied across listeners, and all runs for a given condition were done in the same session. Two of the listeners from experiment I (L1 and L2) took part in experiment II. A new listener, L4, with no musical training whatsoever also participated and was paid for her participation. She was 23 years old and had normal-hearing thresholds at standard audiometric frequencies.

C. Results and discussion

The results are shown in Fig. 2 (solid lines). The patterns are consistent across listeners and so the discussion is limited to the average data. There is a strong effect of frequency region on the LLMP. Listeners are extremely good at the melody task in the two lower-frequency regions; the adaptive procedure converges to a threshold around 35 Hz for both the 200- and 400-Hz filter conditions. In fact, listeners are just as good with the current band-limited stimuli as they were with the broadband stimuli in the previous experiment; for comparison the star symbol on the left-hand side of the graph presents the average threshold from that experiment. The LLMP then increases as F_c increases to the point where a repetition rate greater than 270 Hz is necessary to hear melodies in the highest filter condition, where F_c is 3200 Hz. The influence of frequency region on the LLMP region is similar to that reported by Ritsma (1962). The current results demonstrate that listeners can actually use the pitch cue to perform a melody task right down to the lower boundary of the existence region as defined by Ritsma (1962) and Moore (1973).

It is to be noted that listeners still reported hearing a pitchlike sensation at threshold in the higher filter conditions, which was not the case for the low filter conditions. There are pitch experiments performed with bandpass or highpass filtered click trains where the combination of repetition rate and frequency region would mean the pitch was below the LLMP as measured in the current study. Carlyon (1997) reported some informal melody recognition experiments with a 3900–5300 Hz passband and a base rate of 200 Hz. Melody recognition was possible, even though the pitch was below the current LLMP. Recognizing a familiar melody, however, is a task facilitated by higher cognitive processes, where a few cues might be enough to extract sufficient information about the contour to do the task. The task in the current experiments was designed to minimize such cues. Houtsma and Smurzynski (1990) and Kaernbach and Bering (2000) reported that musical interval recognition was still possible for severely high-passed click trains, although with reduced accuracy. The current task involves a more stringent criterion as it requires semitone accuracy in interval recognition, as is the case for the Western musical scale. These observations suggest that the 30-Hz limit obtained in the broadband condition of experiment I is an absolute lower limit for melodic pitch. This value limits performance in the lower filter conditions ($F_c = 200$ or 400 Hz). For higher filter conditions, a pitchlike percept may be perceived by listeners between this absolute limit and the LLMP, but it is not sufficiently precise to support threshold performance on the melody task.

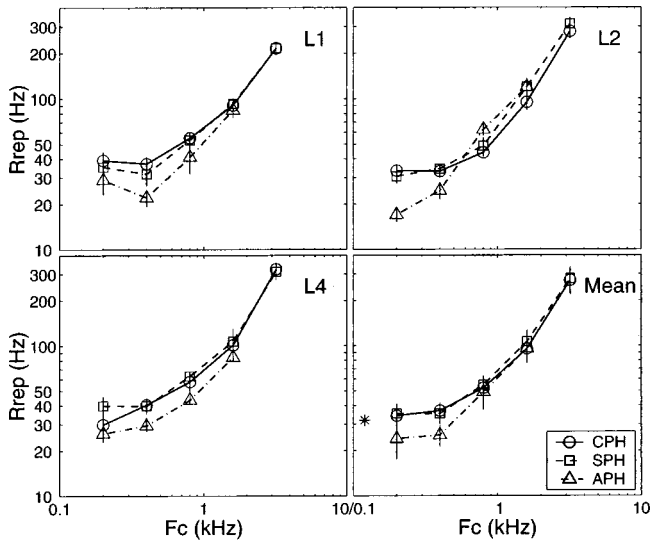


FIG. 2. Results for experiments II and III. The lowest repetition rate, R_{rep}, of the base note for which the melody task could be performed is plotted as a function of filter cutoff, F_c. The parameter is phase condition (CPH for experiment II, SPH and APH for experiment III). The star symbol on the bottom right panel is the mean result for broadband clicks (experiment I).

IV. EXPERIMENT III: THE EFFECT OF PHASE

A. Rationale

Performance in pitch-discrimination tasks is usually better when the stimuli include spectrally resolved components (Houtsma and Smurzynski, 1990; Shackleton and Carlyon, 1994; Carlyon and Shackleton, 1994). The question then arises as to whether the LLMP might reflect the transition from resolved to unresolved components in the internal representation of the sound. To test this hypothesis, performance was compared for stimuli having the same amplitude spectra but different phase spectra. When the components are resolved, differences in phase have little or no influence on performance in pitch tasks; whereas when the components are not resolved phase differences can affect performance (Patterson, 1987; Houtsma and Smurzynski, 1990; Carlyon and Shackleton, 1994). The manipulation of phase also enables evaluation of the effect of wave shape on the LLMP.

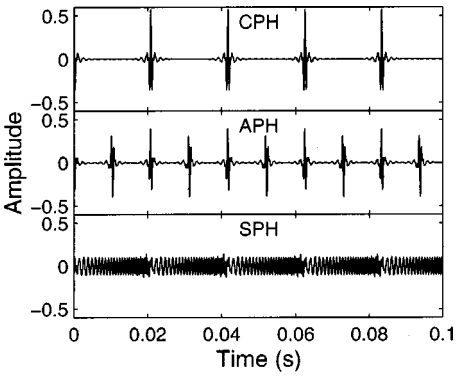


FIG. 3. Waveforms of the stimuli for experiments II (CPH) and III (APH and SPH).

B. Method

1. Stimuli

Three phase relationships were used. In the first condition, all components were in cosine phase (CPH) as in experiment II. This configuration produces the largest possible peak factor for a harmonic complex. The repetition rate is clearly visible in the stimulus waveform (Fig. 3, top panel). In the second condition, every other component was shifted $\pi/2$ radians. This alternating phase (APH; Patterson, 1987) produces a stimulus with a periodic Hilbert envelope at double the repetition rate, $2R_{\text{rep}}$, although the repetition rate of the fine-structure is still R_{rep} (Fig. 3, middle panel). The last phase configuration was derived using the formula proposed by Schroeder (1970). This reduces the amplitude peak factor of the waveform markedly (Fig. 3, bottom panel). The sign of the phase in the Schroeder formula does not change the waveform peak factor. It does change the direction of the chirp in the waveform fine structure. An upward-chirping tone was chosen because evidence exists to indicate that this condition reduces the peak factor of the internal representation of the signal after auditory filtering (Smith *et al.*, 1986). This last condition will be referred to as Schroeder phase (SPH).

2. Procedure and listeners

The thresholds for each combination of filter condition and phase were obtained in separate adaptive runs. The procedure, apparatus and listeners were those of experiment II. The CPH conditions were not repeated; the results were taken directly from experiment II. All thresholds for the other two phase conditions were measured in an identical manner as for the CPH thresholds; the order was varied across subjects.

C. Results

The results are shown in Fig. 2. The CPH and SPH conditions produce very similar results. A Wilcoxon rank-sum test was applied to the raw data for all filter conditions and there were no significant differences ($p>0.2$). This appears to be at variance with the results of Houtsma and Smurzynski (1990) who reported differences in rate discrimination threshold for CPH and SPH harmonic complexes. However, the combinations of F_c and R_{rep} associated with the LLMP in the current experiment are different from the combinations where Houtsma and Smurzynski found reliable differences between the RDTs of CPH and SPH tones—200 Hz at a cutoff frequency of 3.2 kHz or higher. These parameters would produce a stimulus that falls below the pitch region revealed by the current study (270 Hz when F_c is 3.2 kHz). Another difference is that Houtsma and Smurzynski used wideband masking noise that might have interferred with the perception of the stimulus. Lowpass noise was used in the current experiment to avoid interference. Finally, this small discrepancy could reflect a difference between the RDT and LLMP tasks.

In the two lower filter conditions, performance is better for APH stimuli than for CPH or SPH stimuli, and the difference is significant ($p<0.01$, Wilcoxon rank-sum test on the raw data). As F_c increases, the LLMP increases faster for the APH stimuli than for the CPH and SPH stimuli, and no threshold could be measured with the adaptive procedure for the highest filter condition. Perceptually, for a given repetition rate, the pitch increases one octave for unresolved APH tones (Patterson, 1987; Carlyon and Shackleton, 1994). No octave shift has been reported for SPH tones.

V. SPECTRAL RESOLVABILITY

Two analyses were performed to determine whether the increase in the LLMP with frequency region reflected the transition from resolved to unresolved components. The hypothesis is that a clear pitch is required to do the melody task, and it can only be produced by resolved harmonics. One definition of resolvability is that the transition occurs at a constant harmonic number (Plomp, 1964). The precise value of this number varies from 6 to 12 between authors and according to the experimental task. Nevertheless, it should be a fixed value. The number of the lowest harmonic in the stimulus associated with each LLMP value was computed for all combinations of F_c and phase. The results are presented in Fig. 4 (upper panel), which shows that the LLMP does not correspond to a fixed harmonic number. Moreover, the lowest harmonic is the 14th or 15th in some cases.

Shackleton and Carlyon (1994) have proposed modifying the ''constant harmonic number'' rule to introduce a ''constant number of components per auditory filter'' rule. They suggest that the transition region between resolved and unresolved complexes occurs when there are 2 to 3.25 components within the 10-dB bandwidth of the auditory filter, as defined by Glasberg and Moore (1990). The number of components in the auditory filter centered on F_c was computed for stimuli at the LLMP, for each experimental condition. The results are presented in Fig. 4 (lower panel) which shows that number of components per filter at the LLMP varies with F_c, and that most of the observed values occur in the ''unresolved'' region as defined by Shackleton and Carlyon (1994).

In summary, these two analyses, which involve relatively large estimates of the upper limit of spectral resolution, nevertheless indicate (1) that the LLMP does not correspond to the loss of spectral resolution for either criterion, and (2) the LLMP is typically associated with stimuli having no resolved components. Further support for these conclu-

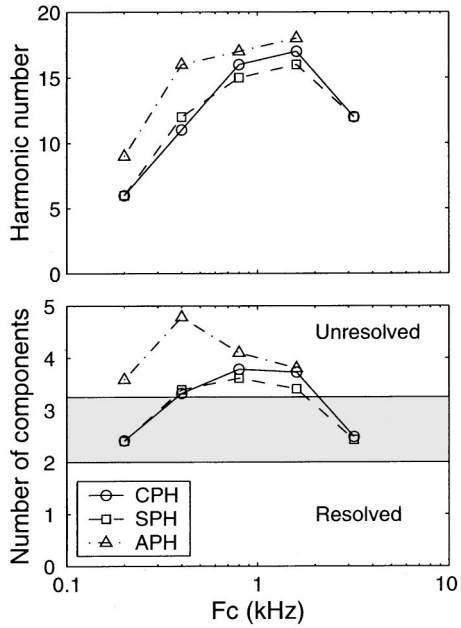


FIG. 4. Harmonic number (upper panel) and number of components between the 10-dB-down points of the auditory filter centered on F_c (lower panel), at threshold, for experiments II and III.

sions is provided by the phase effect in experiment III; the phase effect should only occur if the complexes are unresolved[2] (Patterson, 1987; Carlyon and Shackleton, 1994).

VI. SPECTRAL AND TEMPORAL MODELS OF THE LLMP

From the above discussion, it is clear that simple spectral models of pitch perception that only predict a pitch for perceptually resolved harmonics would fail to predict the LLMP. There exist, however, more sophisticated spectral models where harmonics can be represented above the psychophysical limit of resolution. In the Central Spectrum model (Goldstein, 1973; Srulovicz and Goldstein, 1983), the spectral representation is derived from time intervals in simulated auditory nerve fibres. The resolution of the model is improved by the use of a matched filter on the interval distribution for each fibre. Up to 15 components can be present in the central spectrum (Goldstein *et al.*, 1978; Srulovicz and Goldstein, 1983). The SPINET model (Cohen *et al.*, 1995) uses a on-center, off-surround mechanism that enhances spectral contrast. Both of these models could maintain some sprectral representation at rates corresponding to the LLMP. They would also predict a decrease of performance in high-frequency regions because of the reduction in the precision of the spectral components (Goldstein, 1973; Cohen *et al.*, 1995).

Temporal models where pitch is associated with the dominant periodicities in a range of frequency channels (Schouten *et al.*, 1962) do not immediately explain the LLMP. The temporal precision of the envelope at the output of auditory filters improves in high-frequency channels, but the LLMP nevertheless increases for high filter conditions. Moore (1973, 1997) proposed a theoretical model in which the time intervals in a given channel with center frequency, CF, were restricted to those between 0.5/CF and 15/CF ms. This CF-dependent limit would cause the LLMP to increase in high-frequency regions; however, 15/CF leads to very long intervals for low CFs (75 ms for a 200-Hz channel). The predicted LLMP value would be above the observed one. In the next section, we develop a modified autocorrelation model (Licklider, 1951; Meddis and Hewitt, 1991a) with a CF-independent limit on time intervals, that is able to reproduce the experimental LLMP results. Thus the present data do not preclude either modeling approach. Rather, they indicate that a residue mechanism, be it temporal or spectral, can convey melodies for notes as low as those that can be played on the lowest octave of the piano keyboard.

VII. AN AUTOCORRELATION MODEL OF THE LLMP

A. The autocorrelation model of pitch perception

Licklider (1951) produced the first computational model of pitch perception based on time intervals within auditory frequency channels. The incoming signal is bandpass filtered to simulate cochlear frequency selectivity; then, in each channel, a running autocorrelation function (ACF) is calculated to reveal any periodicity. The output is an array of ACFs and so the dimensions are autocorrelation lag versus filter center frequency, at a given moment in time. This representation is typically referred to as an autocorrelogram (ACG) and it has been used to explain pitch perception. There are distinct limitations to the autocorrelation approach, as noted by Kaernbach and Demany (1998). In this subsection, the Meddis and Hewitt (1991a) implementation of the autocorrelation model is described as a simple means to quantify the time-interval information present in auditory channels.

Meddis and Hewitt (1991a) introduced two additional stages to the autocorrelation model to enable quantitative predictions of pitch perception. First, they averaged the ACFs of all frequency channels to form a summary autocorrelogram (SACG). This emphasizes the time intervals common to a range of frequency channels. The dominant peak in the SACG specifies the period of the predicted pitch. Second, in order to model pitch discrimination performance, they defined a decision statistic equal to the Euclidean distance between the SACGs of pairs of sounds, referred to as d^2. The SACG and the d^2 statistic have been successful in accounting for a number of pitch phenomena (Meddis and Hewitt, 1991a, b; Meddis and O'Mard, 1997).

It is useful for the purposes of the following discussion to point out some of the properties of the autocorrelation calculation at the heart of the model. Licklider proposed performing a running autocorrelation of the signal, s, with an exponential time window [Eq. (1)]:

$$\mathrm{ACF}(t,\tau)=\int_0^{+\infty} s(t-T)s(t-\tau-T)e^{(-T/\Omega)}\,dT. \qquad (1)$$

The time constant, Ω, determines the decay rate of the exponential window and so the time over which the ACF is averaged. Licklider (1951) suggested a value of 2.5 ms for Ω. The ACF fluctuates over time when this parameter is

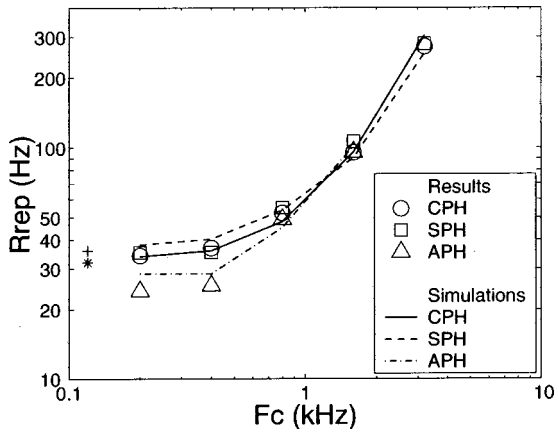


FIG. 5. Results and simulations for experiments I, II, and III. The mean experimental data are represented as unconnected symbols, the model simulation as curves without symbols. The star and cross symbols on the left-hand side of the figure are, respectively, the results and the simulation for the broadband condition.

short (Wiegrebe *et al.*, 1998). To reduce this variability when comparing the SACG of two stimuli, Meddis and Hewitt (1991a) suggested calculating the SACG at times corresponding to a multiple of the stimulus period. With this constraint and an Ω value of 2.5 ms, Meddis and Hewitt were able to explain a wide range of pitch phenomena with the SACG model. Alternatively, Ω can be set to a longer value to produce a better estimate of the long-term unbiased ACF. When explaining the perception of vowels, Meddis and Hewitt (1992) used a value of 25 ms. The time constant Ω is relevant to the integration of pitch over time. For stationary sounds, a relatively long value is usually satisfactory to stabilize the SACG. This parameter is not, however, directly involved in the determination of the LLMP.

B. Representation of the LLMP in the autocorrelation model

With regard to the LLMP, the important parameter is the maximum lag, henceforth denoted τ_m, for which the individual ACFs are computed. The model cannot explain the pitch of stimuli whose period is greater than this value since there would be no peak in the SACG to associate with the pitch. The role of τ_m has not been studied explicitly. Licklider (1951) pointed out the need to specify the maximum lag, and in the absence of experimental data, proposed a value of 33 ms (30 Hz). Similarly, Yost *et al.* (1996) noted that this parameter is related to the lowest audible pitch, and set the value of 35 ms. In general, however, τ_m seems to be set just long enough to ensure that the peaks produced by the stimulus of interest appear in the SACG (Meddis and Hewitt, 1991a, 1992; Meddis and O'Mard, 1997; Yost *et al.*, 1996). This has led to values of τ_m ranging from 10 to 35 ms, without any direct estimation of the parameter.

The problem with the current model is that there is a discontinuity at the boundary. If the maximum lag is set to $\tau_m = 33$ ms, for instance, a 35-Hz tone (29-ms period) produces some activity in the SACG whereas a 25-Hz tone (40-ms period) does not. As a consequence a large d^2 would be obtained predicting excellent discrimination for this semitone change, which is clearly at variance with experimental data (Ritsma and Hoekstra, 1974; Houtsma and Smurzynski, 1990; Krumbholz *et al.*, 2000).

A straightforward way to avoid this problem is to apply a weighting function to the SACG. The weighting chosen here decreases linearly from one at a lag of 0 ms to zero at a lag of 33 ms. This eliminates the discontinuity by making long time intervals disappear gradually from the SACG. Intuitively, the weighting function reduces pitch strength for long lags. It should be noted that the same result is obtained if, instead of a running ACF, a biased, long-term ACF is calculated with 33-ms, unwindowed, portions of the signal. This alternative implementation would depart significantly from the traditional structure of the autocorrelation model, so for the purposes of the current paper we focus on the established models of Licklider (1951) and Meddis and Hewitt (1991a).

C. Simulation of the experimental results

The first stages of the model were identical to Meddis and O'Mard (1997). There were 60 frequency channels regularly distributed on an ERB scale between 100 and 8000 Hz, each with a gammatone filter and a hair-cell simulator. The individual ACFs of the ACG were computed on the hair-cell outputs. The ACF was computed at the end of a complete period of the stimulus as in Meddis and O'Mard (1997). The time constant of the ACF, Ω, was increased from 2.5 to 15 ms. The linear weighting function was applied to the summary ACF terminating at $\tau_m = 33$ ms. The weighted SACG is also referred to as the SACG, for convenience.

LLMP values were produced with this model as follows: SACGs were calculated for all stimuli from the experiments with R_{rep} values between 16 and 340 Hz. Then, for each combination of filter and phase condition, d^2 was computed between the SACGs of stimuli with repetition rates separated by 6%. A threshold value, d^2_{thres} was fixed and the lowest note of the pair of notes whose d^2 just exceeded d^2_{thres} was taken as the estimate of threshold for that combination of stimulus filter and phase. It was verified that the d^2 increased monotonically for notes above threshold. A complete set of LLMP values was produced for the fixed d^2_{thres} and then, the value of d^2_{thres} was varied to find the set of LLMP values that matched the observed values. d^2_{thres} was the only parameter varied in the fitting process. The low-pass noise was not included in the simulations because the model does not produce distortion products. As it is a deterministic model, the addition of the random noise would have needlessly complicated the computation. The results are presented in Fig. 5.

The model reproduces most of the important features of the experimental data (Fig. 5). For the broadband condition, the simulated LLMP is 35 Hz, similar to that derived from experiment I. There is little difference between this condition and the CPH condition for $F_c = 200$ Hz, and the LLMP increases rapidly with increasing frequency region for CPH stimuli (experiment II). The model also exhibits the effect observed in experiment III, where the APH stimuli produce lower LLMP values than the CPH stimuli in the lowest filter conditions ($F_c = 200$ or 400 Hz). In the highest filter condi-

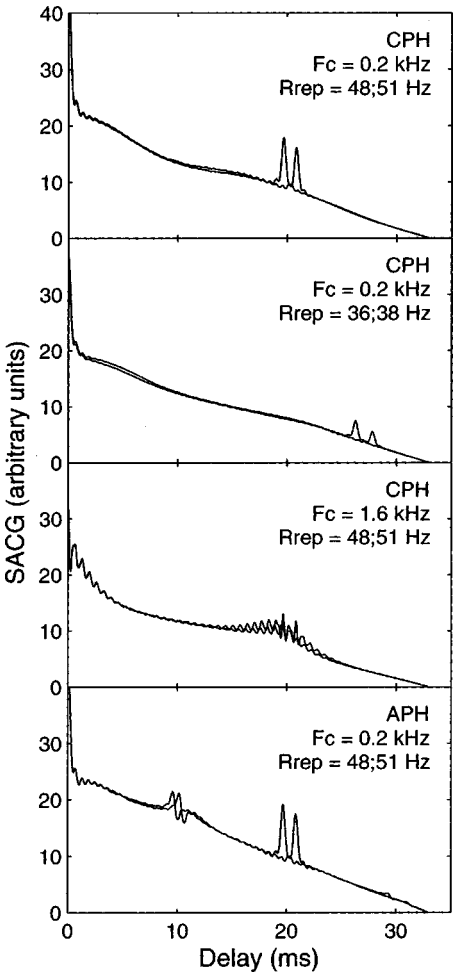


FIG. 6. Pairs of SACGs for stimuli differing by a semitone. See inserts for stimulus parameters.

tion ($F_c = 3200$ Hz), d^2 never reached the threshold criterion and so there is no predicted LLMP value. This is also the condition where the listeners could not perform the task. There is a slight discrepancy between the model values and the experimental data in the low filter conditions ($F_c = 200$ or 400 Hz); the LLMP for the CPH and APH conditions are greater than the experimental ones.

D. Discussion of the simulation

Having imposed a weighting function on the SACG that reduces it to zero at 33 ms, it is not surprising to find that the lowest LLMP values produced by the model are just greater than 30 Hz for CPH. The intriguing finding is that the LLMP values increase with frequency region at the same rate as observed in the data, and that the effect of APH is also reproduced. To understand how this arises, consider the SACGs presented in Fig. 6. Each panel presents two superimposed SACGs for notes separated by 1 semitone—the difference that distinguishes the pair of melodies in a given trial of the experiments and which was used to calculate the d^2 values.

In the upper panel, the stimulus filter has a cutoff of 200 Hz and the notes are well above the LLMP (48 and 51 Hz). The SACGs exhibit clear peaks at the delays of the two notes. In this case, d^2 is large and the model predicts good performance as is observed. As the pitch is lowered toward the LLMP at this filter cutoff (next panel, 36 and 38 Hz), the peaks in the SACG shrink, producing ever smaller d^2 values until eventually d^2 falls below threshold.[3] This is the direct result of the introduction of the weighting function.

Now consider what happens when the filter cutoff is increased to 1.6 kHz. The third panel presents the SACGs for the same notes as in the upper panel (48 and 51 Hz) but with the higher filter cutoff. The weighting function is the same but the peaks are shorter and more spread out. As a result, d^2 is reduced to the point where it falls below threshold and this condition is correctly predicted to be below the LLMP in this filter condition. The broadening of the peaks when energy is restricted to a band of high-frequency channels was noted by Meddis and O'Mard (1997). They argued that the effect arises from the loss of harmonic resolution at high harmonic numbers. However, as Carlyon (1998) noted, it is more an effect of frequency region than of resolvability. We suspect that there are several factors at work here. The first is the loss of phase locking that occurs at high frequencies and restricts the encoding of temporal fine structure. The second factor has to do with the averaging of ACFs across channels. This enhances activity at the stimulus period because there is activity at this lag in all channels. Activity associated with the center frequency of the channel varies with channel and cancels out in the cross-channel averaging. In experiments II and III, the width of the stimulus filter was fixed. As a result, in the low-frequency conditions, there were more active channels which leads to more summation of the common stimulus period and more cancellation of auditory filter activity than in the high-frequency conditions where the stimuli excite fewer filters. The spreading of the peaks is due to this lack of cancellation of auditory filter ringing combined with the loss of phase locking.

Finally, consider the SACGs of the APH stimuli when the cutoff of the stimulus filter is 200 Hz (bottom panel). The phase shift produces smaller peaks mid-way through the period of the stimulus which results in secondary peaks in the SACG mid-way between the main peaks. As the pitch is lowered and the period passes 33 ms, the listeners can switch from the main to the secondary peaks and so perform the melody task for periods that are nominally below the LLMP. This interpretation is compatible with previous explanations of the perception of APH sounds by autocorrelation models (Meddis and Hewitt, 1991b). When the filter cutoff is increased, the secondary peaks fade into the floor activity and so this cue cannot be used in high-frequency regions.

E. Limitations of the d^2 statistic

The model as presented in this paper is not invariant with respect to the bandwidth of the stimuli. The upper and middle panels of Fig. 7 show the output of the model in response to stimuli with a high filter cutoff ($F_c = 3.2$ kHz). In the upper panel, bandwidth is that used in experiments II and III. In the middle panel, the bandwidth has been enlarged to be proportional to auditory filter width (Glasberg and Moore, 1990). The proportional bandwidth equates the num-

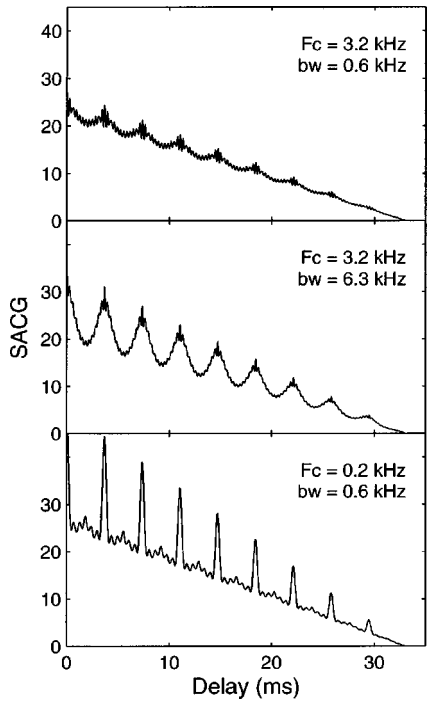


FIG. 7. SACGs for constant or proportional bandwidth stimuli. Phase condition for all stimuli is CPH, repetition rate is $R_{rep}=271$ Hz. The top, middle, and bottom panels illustrate the cases of constant bandwidth, proportional bandwidth, and constant/proportional bandwidth, respectively.

ber of auditory channels activated in the low and high filter conditions ($F_c=200$ Hz and 3.2 kHz). The increased bandwidth leads to more activation in the SACG and larger peaks, which in turn produces larger values of d^2. As a consequence, lower values of the LLMP are predicted compared to the LLMP obtained with constant bandwidth stimuli.

In contrast to this prediction, Krumbholz *et al.* (2000) have shown that the RDTs obtained with a fixed passband were similar to those obtained by other experimenters using proportional bandwidth or high-pass stimuli. This suggests that the LLMP is not much affected by bandwidth. Figure 7 thus exhibits a limitation of the autocorrelation model in its current implementation, mainly because of the nonauditory aspect of the d^2 statistic. The peaks produced by the proportional bandwidth stimulus in the high filter condition are much broader than those produced in the $F_c=200$-Hz condition (bottom panel); it is possible that a more sophisticated statistic based on peak-picking, and taking into account the width of the peaks, could still account for the increase of LLMP in high-frequency regions. Such a modification is beyond the scope of the current paper. The autocorrelation model has served the purpose of demonstrating that the information present in time intervals within auditory channels can in principle explain the LLMP results, and this is sufficient to indicate that models based more closely on the physiology of temporal processing in the auditory system are worth investigating.

VIII. THE LLMP AND RATE DISCRIMINATION THRESHOLD

In a companion paper (Krumbholz *et al.*, 2000), the RDT for many of the stimuli used in the current study were measured—for CPH and APH stimuli in all frequency regions. It is the case that RDT rises rapidly around the LLMP. Krumbholz *et al.* (2000) found that the LLMP actually corresponds to a RDT of about 2.5%. This is substantially less than the 6% rate change that was available in the melody task. In fact, for most listeners, RDT did not exceed 5% for rates as low as 16 Hz in all frequency regions. The listeners would have been able to perform a semitone discrimination task with rates much lower than the LLMP. Thus the LLMP is not the point where the target note and its semitone neighbors become indiscriminable.

There is an obvious difference between the melody task and the task used to measure RDTs. In the RDT case, listeners compared a single pair of sounds. In contrast, the melody task involves eight different sounds. A simple model of the listeners' strategy would assume that they stored all the sounds, or some attributes of the sounds, in memory and then performed pairwise comparisons to make their judgments. In this case, the LLMP task is similar to four, parallel, RDT tasks, and signal detection theory predicts a reduction in discriminability by a factor of $\sqrt{4}$, as measured by d' (Macmillan and Creelman, 1991). Plack and Carlyon (1995) have shown that d' is proportional to the rate difference for RDT tasks with harmonic complex tones. Thus the fact that the threshold for the melody task with 6% between notes corresponds to a RDT of about 3% is consistent with the decision model based on the storage and retrieval of the sequence of sounds.

This leads to the argument that the cues used to perform the LLMP task were, indeed, pitch cues. The eight rates have to be stored over 4 s. Each note is perceived as a distinct sound event because it is 400 ms long, but the task clearly includes a memory component on a relatively long time scale. McFarland and Cacace (1992) have studied the perception of binary tone patterns similar in duration to the melodies used in the current experiments. The tone patterns were constructed by randomly alternating between two values of a given parameter, which was frequency, amplitude or duration. The maximum number of stimuli in the pattern that could be reliably memorized was determined by means of an adaptive procedure. To make comparison across parameters realistic, the difference between the two parameter values in the sequences was always a constant number of *jnd*'s for that parameter. McFarland and Cacace (1992) found that patterns based on alternation in frequency were retained longer than patterns based on alternation of amplitude or duration (4.4 s for frequency, as opposed to 1.7 s for amplitude or duration).

The superior performance with pitch sequences can be explained by the results of Semal and Demany (1991, 1993) who have shown that there is a pitch-specific memory that cannot be used for loudness. Clément *et al.* (1999) found that the accuracy of the pitch trace is maintained longer than that for loudness. Moore and Rosen (1979) had previously observed that no melody could be recognized when pitch intervals were replaced by loudness intervals, even when listeners were selecting from a closed set of familiar melodies. Interestingly, Semal and Demany (1993) also demonstrated that the timbre cues associated with repetition rates below 30 Hz,

and hence below the LLMP, could not be stored in pitch memory.

These findings point to the following interpretation of the LLMP and its link to RDTs. For the higher repetition rates, pitch cues mediate both the melody task and RDT ($<1\%$). Listeners can store the pitch cues in memory and use them in the melody task until criterion performance is reached. As repetition rate decreases and the pitch cues become less reliable, listeners switch to other cues such as roughness or pulse rate to perform rate discrimination, and this explains the sudden rise in RDT to around 5%. The melody task requires a memory that operates for longer than a standard 2AFC rate discrimination trial, and thus pitch is the sole cue that works in the melody task.

IX. MELODIC AND MUSICAL PITCH

It is debatable whether the pitch cue used in the LLMP task can truly be called ''musical.'' This is why the more restricted term of melodic pitch was used. To establish whether pitch retains a musical quality down to the LLMP, further measurements involving interval recognition would be needed (Houtsma and Goldstein, 1972). It is likely that such measurements would be difficult to obtain, however, because of the need to control for (1) a restricted range and thus a limited number of intervals, (2) the ability of listeners to learn how to label arbitrarily a small set of rate differences with musical names. Randomly transposing the melodies seems a more appropriate approach to resolve this issue (de Boer and Houtsma, personnal communication). Also, the 69% correct value chosen as threshold might prove insufficient to convey melodies reliably in a musical context (Goldstein, 2000). Measurement of the rates that support performance closer to 100% correct is not practical with an adaptive procedure; they would have to be inferred from psychometric functions. In spite of these restrictions concerning the musical relevance of the LLMP and its relation to more established definitions of musical pitch, it is to be noted that the 30 Hz value corresponds to the lowest note available on the piano keyboard (A0, 27.5 Hz).

X. CONCLUSIONS

A melody-change task was used to measure the lower limit of melodic pitch which was found to be around 30 Hz, provided there was energy in the stimulus below 800 Hz. The value of 30 Hz corresponds roughly to the lowest note on the piano keyboard (27.5 Hz), but it is an octave above the 16-Hz pipe on large organs.

When the stimuli were bandpass filtered, the LLMP was found to increase rapidly with frequency in the region above 800 Hz. This effect is consistent with the results of Ritsma (1962) and Moore (1973) despite the differences in stimuli and experimental task. Above 30 Hz and below the LLMP, listeners still experience a weak pitch sensation but it is not sufficiently well defined to convey melodies with semitone accuracy.

The LLMP does not correspond to the loss of spectral resolution for individual harmonics, as usually defined; unresolved harmonics can support melodic pitch.

The data can be simulated by a modified autocorrelation model where a limit of about 33 ms is imposed on the time intervals that the pitch mechanism can accommodate.

ACKNOWLEDGMENTS

The authors would like to thank C. Kaernbach, C. Plack, and J. L. Goldstein for helpful comments on earlier versions of the manuscript. We also wish to thank A. de Cheveigné, L. Demany, and R. P. Carlyon for their much appreciated input. Supported by the U.K. Medical Research Council (G9703469). The first author is supported by the French Center National de la Recherche Scientifique.

[1]Throughout this paper, stimuli are described in terms of their waveform repetition rates, in Hertz. This is a convenient way to describe harmonic sounds that may or may not contain energy at their fundamental frequency.

[2]The APH effect occurs in lower-frequency regions, which seems to be the opposite of what is normally observed. Note, however, that the important factor for an octave shift is harmonic resolvability, not frequency region. The LLMP occurs at very low rates in low-frequency regions where the components are unresolved.

[3]In the modified model we assume that d^2 represents only the pitch cues that can be used for discrimination between the two notes. For long periods (>33 ms), the SACG would be empty and the model would not predict any discriminability. In this case, listeners would still be able to discriminate between sounds, but by using cues that are not pitch cues, and therefore, which would not appear in the current model.

de Boer, E. (**1976**). ''On the ''residue'' and auditory pitch perception,'' in *Handbook of Sensory Physiology*, edited by W. D. Keidel and W. D. Neff (Springer Verlag, Berlin).

Carlyon, R. P. (**1997**). ''The effect of two temporal cues on pitch judgment,'' J. Acoust. Soc. Am. **102**, 1097–1105.

Carlyon, R. P. (**1998**). ''Comments on 'A unitary model of pitch perception','' J. Acoust. Soc. Am. **104**, 1118–1121.

Carlyon, R. P., and Shackleton, T. M. (**1994**). ''Comparing the fundamental frequencies of resolved and unresolved harmonics: Evidence for two pitch mechanisms?,'' J. Acoust. Soc. Am. **95**, 3541–3554.

Clément, S., Demany, L., and Semal, C. (**1999**). ''Memory for pitch versus memory for loudness,'' J. Acoust. Soc. Am. **106**, 2805–2811.

Cohen, M. A., Grossberg, S., and Wyse, L. L. (**1995**). ''A spectral network model of pitch perception,'' J. Acoust. Soc. Am. **98**, 862–879.

Crowder, R. G., and Morton, J. (**1969**). ''Precategorical acoustic storage,'' Percept. Psychophys. **5**, 365–373.

Dowling, W. J., and Fujitani, D. S. (**1971**). ''Contour, interval, and pitch recognition in memory for melodies,'' J. Acoust. Soc. Am. **49**, 524–531.

Dowling, W. J., and Harwood, D. L. (**1986**). *Music Cognition* (Academic, London).

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. (**1967**). ''Auditory nonlinearity,'' J. Acoust. Soc. Am. **41**, 676–689.

Goldstein, J. L. (**1973**). ''An optimum processor theory for the central formation of the pitch of complex tones,'' J. Acoust. Soc. Am. **54**, 1496–1515.

Goldstein, J. L. (**2000**). ''Pitch perception,'' in *Encyclopedia of Psychology*, edited by A. E. Kazdin, Vol. 6 (American Psychological Association & Oxford University Press), pp. 201–210.

Goldstein, J. L., Gerson, A., Srulovicz, P., and Furst, M. (**1978**). ''Verification of the optimal probabilistic basis of aural processing in pitch of complex tones,'' J. Acoust. Soc. Am. **63**, 486–497.

Griffiths, T. D., Büchel, C., Frackowiak, R. S. J., and Patterson, R. D. (**1998**). ''Analysis of temporal structure in sound by the human brain,'' Nat. Neurosci. **1**, 422–427.

Guttman, N., and Julesz, B. (**1963**). ''Lower limits of auditory periodicity analysis,'' J. Acoust. Soc. Am. **35**, 610.

Guttman, N., and Pruzansky, S. (**1962**). ''Lower limits of pitch and musical pitch,'' J. Speech Hear. Res. **5**, 207–214.

Houtsma, A. J. M. (**1995**). ''Pitch perception,'' in *Hearing, Handbook of Perception and Cognition*, 2nd ed., edited by B. C. J. Moore (Academic, San Diego), pp. 267–295.
Houtsma, A. J. M., and Goldstein, J. L. (**1972**). ''The central origin of the pitch of complex tones: Evidence from musical interval recognition,'' J. Acoust. Soc. Am. **51**, 520–529.
Houtsma, A. J. M., and Smurzynski, S. (**1990**). ''Pitch identification and discrimination for complex tones with many harmonics,'' J. Acoust. Soc. Am. **87**, 304–310.
Kaernbach, C., and Bering, C. (**2000**). ''Ist temporale tonhöhe musikalisch?,'' in Proceedings of DAGA 2000, Oldenburg (in press).
Kaernbach, C., and Demany, L. (**1998**). ''Psychophysical evidence against the autocorrelation theory of auditory temporal processing,'' J. Acoust. Soc. Am. **104**, 2298–2306.
Krumbholz, K., Patterson, R. D., and Pressnitzer, D. (**2000**). ''The lower limit of pitch as determined by rate discrimination,'' J. Acoust. Soc. Am. **108**, 1170–1180.
Levitt, H. (**1971**). ''Transformed up–down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.
Licklider, J. C. R. (**1951**). ''A duplex theory of pitch perception,'' Experientia **7**, 128–133.
Macmillan, N. A., and Creelman, C. D. (**1991**). *Detection Theory: A User's Guide* (Cambridge University Press, Cambridge, UK).
McFarland, D. J., and Cacace, A. T. (**1992**). ''Aspects of short-term acoustic recognition memory: Modality and serial position effects,'' Audiology **31**, 342–352.
Meddis, R., and Hewitt, M. J. (**1991a**). ''Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification,'' J. Acoust. Soc. Am. **89**, 2866–2882.
Meddis, R., and Hewitt, M. J. (**1991b**). ''Virtual pitch and phase sensitivity of a computer model of the auditory periphery. II: Phase sensitivity,'' J. Acoust. Soc. Am. **89**, 2883–2894.
Meddis, R., and Hewitt, M. J. (**1992**). ''Modeling the identification of concurrent vowels with different fundamental frequencies,'' J. Acoust. Soc. Am. **91**, 233–245.
Meddis, R., and O'Mard, L. (**1997**). ''A unitary model of pitch perception,'' J. Acoust. Soc. Am. **102**, 1811–1820.
Moore, B. C. J. (**1973**). ''Some experiments relating to the perception of complex tones,'' Quart. J. Exp. Psychol. **25**, 451–475.
Moore, B. C. J. (**1997**). *An Introduction to the Psychology of Hearing*, 4th ed. (Academic, London).
Moore, B. C. J., and Rosen, S. M. (**1979**). ''Tune recognition with reduced pitch and interval information,'' Q. J. Exp. Psychol. **31**, 229–240.
Parncutt, R., and Cohen, A. J. (**1995**). ''Identification of microtonal melodies: Effects of scale-step size, serial order, and training,'' Percept. Psychophys. **57**, 835–846.
Patterson, R. D. (**1973**). ''The effect of the relative phase and number of components on residue pitch,'' J. Acoust. Soc. Am. **53**, 1565–1572.
Patterson, R. D. (**1987**). ''A pulse ribbon model of monaural phase perception,'' J. Acoust. Soc. Am. **82**, 1560–1586.
Patterson, R. D., Peters, R. W., and Milroy, R. (**1983**). ''Threshold duration for melodic pitch,'' *Hearing—Physiological Bases and Psychophysics*, edited by Klinke and Hartmann (Springer, Berlin), pp. 321–326.
Plack, C. J., and Carlyon, R. P. (**1995**). ''Differences in frequency modulation detection and fundamental frequency discrimination between complex tones consisting of resolved and unresolved harmonics,'' J. Acoust. Soc. Am. **98**, 1355–1364.
Plomp, R. (**1964**). ''The ear as a frequency analyzer,'' J. Acoust. Soc. Am. **36**, 1628–1636.
Plomp, R. (**1967**). ''Pitch of complex tones,'' J. Acoust. Soc. Am. **41**, 1526–1533.
Plomp, R. (**1976**). *Aspects of Tone Sensation* (Academic, London).
Ritsma, R. J. (**1962**). ''Existence region of the tonal residue I.,'' J. Acoust. Soc. Am. **34**, 1224–1229.
Ritsma, R. J. (**1971**). ''Psychological correlates of a frequency shift,'' in Proc. 7th ICA (p. 19H16), Budapest.
Ritsma, R. J., and Hoekstra, A. (**1974**). ''Frequency selectivity and the tonal residue,'' in *Psychophysical Models and Physiological Facts in Hearing*, edited E. Zwicker and E. Terhardt (Springer, Berlin).
Schouten, J. F. (**1940**). ''The residue and the mechanism of hearing,'' Proc. K. Ned. Akad. Wet. **43**, 991–999.
Schouten, J. F., Ritsma, R. J., and Cardozo, B. L. (**1962**). ''Pitch of the residue,'' J. Acoust. Soc. Am. **34**, 1418.
Schroeder, M. R. (**1970**). ''Synthesis of low peak-factor signals and binary sequences with low autocorrelation,'' IEEE Trans. Inf. Theory **16**, 85–89.
Semal, C., and Demany, L. (**1991**). ''Dissociation of pitch from timbre in auditory short-term memory,'' J. Acoust. Soc. Am. **89**, 2404–2410.
Semal, C., and Demany, L. (**1993**). ''Further evidence for an autonomous processing of pitch in auditory short-term memory,'' J. Acoust. Soc. Am. **94**, 1315–1322.
Shackleton, T. M., and Carlyon, R. P. (**1994**). ''The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination,'' J. Acoust. Soc. Am. **95**, 3529–3540.
Smith, B. K., Sieben, U. K., Kohlrausch, A., and Schroeder, M. R. (**1986**). ''Phase effects in masking related to dispersion in the inner ear,'' J. Acoust. Soc. Am. **80**, 1631–1637.
Smoorenburg, G. F. (**1970**). ''Pitch perception of two frequency stimuli,'' J. Acoust. Soc. Am. **48**, 924–942.
Srulovicz, P., and Goldstein, J. L. (**1983**). ''A central spectrum model: A synthesis of auditory nerve timing and place cues in monaural communication of frequency spectrum,'' J. Acoust. Soc. Am. **73**, 1266–1276.
Terhardt, E. (**1970**). ''Frequency analysis and periodicity detection in the sensation of roughness and periodicity pitch,'' in *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G. Smoorenburg (A. W. Sijthoff, Leiden), pp. 278–287.
Warren, R. M., and Bashford, J. A. J. (**1981**). ''Perception of acoustic iterance: Pitch and infrapitch,'' Percept. Psychophys. **29**, 395–402.
Wiegrebe, L., Patterson, R. D., Demany, L., and Carlyon, R. P. (**1998**). ''Temporal dynamics of pitch strength in regular interval noises,'' J. Acoust. Soc. Am. **104**, 2307–2313.
Yost, W. A., Patterson, R. D., and Sheft, S. (**1996**). ''A time domain description for the pitch strength of iterated rippled noise,'' J. Acoust. Soc. Am. **99**, 1066–1078.

A measure for predicting audibility discrimination thresholds for spectral envelope distortions in vowel sounds

Preeti Rao
Advanced Center for Research in Electronics, Indian Institute of Technology, Bombay, Powai, Mumbai 400076, India

R. van Dinther, R. Veldhuis, and A. Kohlrausch[a)]
IPO, Center for User–System Interaction, 5600 MB Eindhoven, The Netherlands

(Received 13 April 2000; accepted for publication 18 January 2001)

Both in speech synthesis and in sound coding it is often beneficial to have a measure that predicts whether, and to what extent, two sounds are different. This paper addresses the problem of estimating the perceptual effects of small modifications to the spectral envelope of a harmonic sound. A recently proposed auditory model is investigated that transforms the physical spectrum into a pattern of specific loudness as a function of critical band rate. A distance measure based on the concept of partial loudness is presented, which treats detectability in terms of a partial loudness threshold. This approach is adapted to the problem of estimating discrimination thresholds related to modifications of the spectral envelope of synthetic vowels. Data obtained from subjective listening tests using a representative set of stimuli in a 3IFC adaptive procedure show that the model makes reasonably good predictions of the discrimination threshold. Systematic deviations from the predicted thresholds may be related to individual differences in auditory filter selectivity. The partial loudness measure is compared with previously proposed distance measures such as the Euclidean distance between excitation patterns and between specific loudness applied to the same experimental data. An objective test measure shows that the partial loudness measure and the Euclidean distance of the excitation patterns are equally appropriate as distance measures for predicting audibility thresholds. The Euclidean distance between specific loudness is worse in performance compared with the other two. [DOI: 10.1121/1.1354986]

PACS numbers: 43.66.Ba, 43.66.Cb, 43.71.Es [RVS]

I. INTRODUCTION

Two important problems in sound compression and speech synthesis are the prediction of whether two sounds are perceived as different and how to express supra-threshold quality differences. An objective distance measure for predicting audibility thresholds and supra-threshold quality differences is important in both areas of research. Although it is valuable to have an objective distance measure which can assess the subjective quality of an entire sentence or phrase, and which correlates well with subjective test scores, it is also useful to evolve a measure which can predict the quality of short steady segments. Such a measure can serve as the basis for an overall quality measure, and can be used in an analysis-by-synthesis framework where the difference between the reference sound and the synthesized sound needs to be estimated.

Commonly used basic objective distance measures, such as signal-to-noise ratios and spectral distances, are derived directly from differences in the waveforms or in the power spectra of the reference and test signals (Quackenbush *et al.*, 1988). However, because it is the perception of the distortion that needs to be quantified, it is expected that measures derived from models of the auditory system will provide the most accurate predictions.

This paper addresses the problem of finding a perceptual distance measure that predicts audibility discrimination thresholds of modifications to the spectral envelope of steady vowel-like sounds. Such sounds are completely specified by their power spectra which can be represented as a set of harmonic components at multiples of a specified fundamental frequency. Modifications to the power spectrum occur in the form of magnitude changes of the harmonic components. We do not consider the effect of phase changes in vowel spectra because it is known that phase distortion has a relatively minor effect on the sound quality of complex tones (Plomp, 1976). In the case of sounds with harmonic spectra, the spectral magnitude changes can be viewed as distortions of the spectral envelope of the harmonic components. Sources of this type of distortion are, for example, filtering by a nonuniform gain transfer function and the inaccurate modeling of the spectral envelope, for instance, in linear predictive synthesis. It is of interest to predict whether the modifications give rise to discriminable changes in perceived quality, and if so, to quantify the extent of perceptual degradation.

In the next section we review the past development of auditory distance measures for the distortion of the spectral envelope of vowel sounds. We motivate and propose a new distance measure based on partial loudness for the prediction of the discrimination threshold. In Sec. III, a brief overview is given of the loudness model recently proposed by Moore *et al.* (1997), which forms the basis of the present work. The adaptation of the partial loudness measure to the prediction

[a)]Also affiliated with Philips Research Laboratories, Eindhoven.

of audibility discrimination thresholds for arbitrary modifications of the spectral envelope is discussed. In Sec. IV, measured discrimination thresholds using a 3IFC adaptive procedure with a representative set of stimuli are used to validate the applicability of the model in this context. The experimental results are discussed and possible explanations for deviations from the predicted thresholds in some cases, are provided. Other previously proposed and commonly used auditory distance measures are also evaluated on the same set of experimental data and their performance is compared with that of the partial loudness based distance measure.

II. VOWEL QUALITY DISTANCE MEASURES

Previous work on the problem of prediction of perceptual differences for vowel sounds has been based on modelling, to various degrees, the differences in the internal representations of reference and modified sounds. The auditory system includes the auditory periphery as well as central processing. It is assumed that perceptual discriminability (under optimal listening conditions and using well-trained subjects) depends largely on the resolution properties of the auditory periphery, and should be predictable by any good model of peripheral auditory processing (Gagné and Zurek, 1988).

Such an assumption led to the work of Plomp (1976) in which the spectral levels of the input stimulus power are summed over 1/3-octave bands, approximating the critical bands of the auditory system, to obtain a spectral representation more closely matched to that assumed in auditory processing. The quadratic distance between the reference and test signal representations was used to predict subjective quality differences in a set of steady sounds. In a further refinement, perceptual distance measures based on auditory excitation patterns have been applied to explain a variety of subjective discrimination data by postulating a threshold difference in excitation levels for detectability. Excitation patterns, or the excitation level per critical band, were first proposed by Zwicker and Scharf (1965) as part of the ''power spectrum model'' for auditory processing. These are calculated from the power spectrum as the output of the auditory filters with centers distributed uniformly on a critical band scale. Excitation patterns were used by Gagné and Zurek (1988), who investigated resonance-frequency discrimination of single formant vowels. The difference in the excitation patterns of the reference and modified signals was used to derive a distance measure given by either the single, largest magnitude difference (single-band model) or by the appropriately combined differences across bands (multiband model). A similar approach is followed in Kewley-Port (1991) who reported on detection thresholds for isolated vowels and examined several detection hypotheses of vowel spectra, based on their excitation patterns. Sommers and Kewley-Port (1996) studied the modelling of formant frequency discrimination of female vowels and evaluated an excitation-pattern model for this purpose.

While the excitation pattern represents the distribution of excitation along the basilar membrane, the loudness per critical band (specific loudness) corresponds more closely to the distribution of neural activity. The specific loudness is closely related to the subjective perception of loudness. A perceptual measure based on specific loudness is justified by the fact that the specific loudness versus critical band rate represents the best psychoacoustical equivalent of the power spectrum (Zwicker and Fastl, 1990). Distance measures based on applying various Minkowski metrics to the difference between specific loudnesses have been used to predict subjective distances in vowel quality (Bladon and Lindblom, 1981). This approach has also been followed more recently to explain the variation in formant-frequency discrimination thresholds observed in steady-state vowels (Kewley-Port and Zheng, 1998). The Euclidean distance between the reference and modified signals' specific loudnesses is used as the distance measure. However, the approach of applying a distance measure to specific loudness suffers from two serious shortcomings as far as the prediction of discrimination thresholds is concerned.

(1) It lacks a sound basis for the mathematical form of the distance measure, e.g., Euclidean, area, etc. Such a development is possible only in a purely experimental manner by observing the correlation between the distance measure and subjective data in specific situations.
(2) There is no basis for selecting the numerical value of the threshold level of the distance metric for the prediction of audibility.

Because we are interested in predicting the thresholds of discrimination for wide-ranging modifications to the spectral envelope, it is of importance to have a relatively invariant threshold level for the distance measure, preferably one based on a large and diverse body of psychoacoustical data.

Prediction of the discrimination threshold is a part of the larger problem of quantifying the perceptual effect of a distortion of the signal. That is, treating the difference between the original and modified signals as the signal to be detected, we wish to quantify its audible significance or its perceived loudness. The type of distortion under consideration in this paper involves a spectral gain modification. Since no new frequency components are created, it constitutes a linear distortion. For the purpose of computing auditory distance measures, these can be treated as additive distortion with a power spectrum equal to the difference in the power spectra of the reference and modified signals (Schroeder *et al.*, 1979). We wish, then, to estimate the audibility of this additive distortion which can be viewed as the ''signal'' to be detected in the presence of the background ''noise'' representing the reference signal. The background sound generally reduces the perceived loudness of the signal, an effect known as partial masking. The loudness of the signal in the presence of the background noise, or the partial loudness of the signal, is then a valid basis for an objective distance measure between the original and modified power spectra. To assess the partial loudness it requires the availability of a computational procedure such as the one given by Zwicker's loudness model (Zwicker and Scharf, 1965; Zwicker and Fastl, 1990). Recently a modified version of Zwicker's loudness model incorporating a more analytical formulation, was introduced by Moore *et al.* (1997). This revised model has been shown to account more accurately for various subjective loudness data. An enhancement to the earlier model particularly rel-

evant to our problem is the quantification of subthreshold levels of partial loudness and the consequent outcome of a threshold of audibility in terms of a partial loudness threshold. With such a threshold definition, this model has been used to predict thresholds related to the detection of tones in noise backgrounds as measured in various masking experiments (Moore *et al.*, 1997). In the next section we discuss the implementation of the partial loudness model and its adaptation to the problem of the prediction of discrimination thresholds for arbitrary envelope modifications of steady harmonic complexes.

III. THE PARTIAL LOUDNESS MODEL

The loudness model of Moore *et al.* (1997) is based on the approximate stages of auditory processing representing the conversion of the input power spectrum to the excitation pattern on the basilar membrane and the subsequent transformation to a specific loudness density. In the case of a signal presented with a background sound or masker (henceforth referred to simply as the ''noise''), a partial specific loudness distribution is derived for the signal based on the computed excitation pattern of the signal as well as that of the noise. The overall partial loudness of the signal, in sone, is then given by the total area under the partial specific loudness distribution. While the loudness model is based on analytical formulations, which represent approximately the stages of physiological processing, the exact nature of the formulations and their various parameters have been optimized to fit a large body of psychoacoustical data on masked thresholds and partial loudness judgements for a variety of multitone and noise stimuli. We next review the structure of the stages of the model in some detail.

A. Computing the excitation pattern

The excitation pattern of a sound is calculated as the output of the auditory filters representing the frequency selectivity of hearing at specific center frequencies. Figure 1 shows the stages involved in obtaining the excitation pattern from the input signal power spectrum which is specified by the frequencies and power spectral levels in dB SPL of its components. The first two blocks describe transfer functions from the free field to the eardrum and through the middle ear, respectively. For sounds presented over headphones, the fixed filter modeling the transfer function from the free field to the eardrum is replaced by one with a flat frequency response. In the third stage the excitation pattern of a given sound is calculated from the effective spectrum reaching the cochlea. According to Moore and Glasberg (1987), excitation patterns can be thought of as the distribution of ''excitation'' evoked by a particular sound in the inner ear along a frequency axis. In terms of a filter analogy, the excitation pattern represents the output level of successive auditory filters as a function of their center frequencies. The excitation pattern is generally presented as a function of the ERB rate rather than as a function of frequency. ERB refers to the equivalent rectangular bandwidth of the auditory filter and is a function of the filter center frequency. The ERB rate is a value on the ERB scale, which is closely related to the critical-band scale of the auditory system. On this scale the auditory filters are uniformly spaced with the ERB rate related to the frequency in kHz through a approximately logarithmic relation (Moore *et al.*, 1997).

Auditory filter shapes, experimentally derived from notched-noise experiments, are characterized as rounded exponential (RoEx) filters with parameters that control the filter selectivity (Moore and Glasberg, 1987). The frequency selectivity depends both on the center frequency of the auditory filter and the input stimulus level. With increasing input level the lower slope of the filter becomes shallower. The contribution of each stimulus component to the excitation pattern is calculated with a filter shape particular to that component. The lower slope of a filter is determined by the total stimulus level within the one-ERB band surrounding the stimulus component under consideration (van der Heijden and Kohlrausch, 1994). Thus to calculate the excitation level corresponding to the output of a given auditory filter, the input power spectral components are each weighted depending on their level and distance from the filter center frequency and combined additively as depicted in Fig. 1. This is repeated for all filter center frequencies spaced at intervals of 0.1 ERB in the range of 50 Hz to 15 kHz. We thus obtain the complete excitation pattern as a density, i.e., in dB SPL per ERB.

B. Calculating the partial loudness

The next stage of the model is the transformation from excitation pattern to specific loudness, which is the loudness density in sone per ERB. The specific loudness is obtained from the excitation distribution versus ERB rate by a compressive nonlinearity. The partial specific loudness of a signal in a background noise refers to its reduced perceived loudness and hence depends on the excitation distributions of the signal as well as that of the noise background. The formulas in Moore *et al.* (1997) provide this mapping based on psychoacoustical studies of loudness (Stevens, 1957; Zwicker and Scharf, 1965) as well as several subsequent experimental data on loudness perception and discriminability thresholds. Figure 10 in Moore *et al.* (1997) shows plots of the model output in terms of partial specific loudness (sone per ERB) versus signal excitation level for a range of noise excitation levels. The center frequency influences the computations by way of the level of the threshold in quiet which is assumed to vary with frequency in the model. From an examination of this figure, several features become evident. (1) The partial specific loudness is related to the signal excitation by a compressive nonlinearity that increases in strength with increasing noise excitation levels. This arises from the increased levels of masking at higher noise levels. (2) At levels of signal excitation well above the noise excitation, the partial specific loudness curves for the various noise levels converge and approach the specific loudness for the signal in quiet. (3) For a given noise excitation level, as the signal excitation approaches its masked threshold, the partial specific loudness rapidly attains low values and continues to decrease in value with decreasing signal excitation level.

For a signal presented in a background noise, the calculation of partial specific loudness requires the computation of

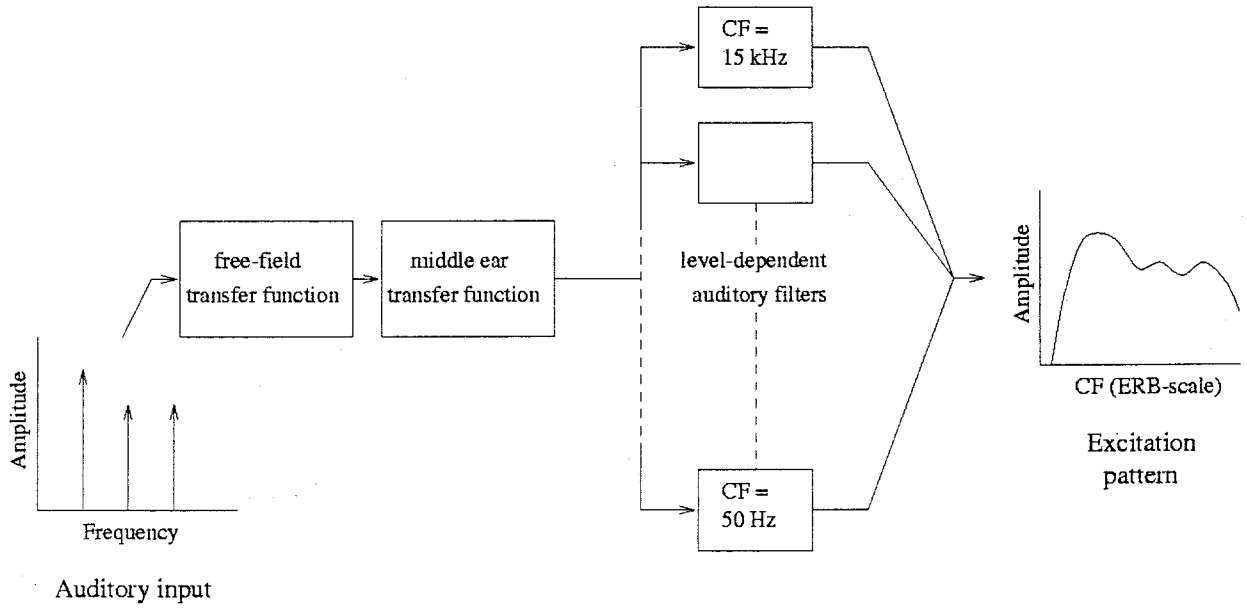


FIG. 1. Schematic diagram for calculating excitation patterns from the power spectrum of a sound represented by the frequencies and amplitudes of its harmonic components.

three excitation patterns. First, an excitation pattern is calculated for the total sound, that is, the signal plus the background noise. The auditory filter shape parameters obtained in the course of this computation are stored. Then, using these parameters, two further excitation patterns are calculated: one for the background noise and one for the signal. The partial specific loudness of the signal is next calculated using the formulas relating to the functions of Fig. 10 in Moore *et al.* (1997), at each ERB rate location as a function of the corresponding excitation levels of the signal and the noise as well as the threshold in quiet at that frequency location. The overall loudness of the given signal, in sone, is assumed to be the area under the specific loudness density. According to the model, the absolute or masked threshold of a sound corresponds to the level at which its partial loudness is 0.003 sone. Hence the model predicts, using the same transformation, both the subjective loudness and the discrimination threshold. The overall partial loudness as computed by the model is therefore a suitable candidate for quantifying the audible significance of the signal.

We see that to predict the discrimination threshold, the model integrates the specific loudness contributions across the entire ERB-rate range and as such can be considered a ''multiband'' model. The model has been used successfully to predict threshold data from a number of previous experiments on multicomponent complex tones in noise by assuming the threshold in overall partial loudness to be at levels between 0.003 sone and 0.008 sone (Moore *et al.*, 1997).

C. Partial loudness of arbitrary spectral-envelope distortions

The partial loudness measure can be applied to the problem of discriminating modifications of the spectral envelope of a steady sound in the following way. The reference sound is intended to take the role of the background noise and the modified sound that of signal plus background noise. We assume that the linear spectral distortion can be treated as additive noise with a power spectrum given by the difference between the reference and modified power spectra. The distortion is then the signal to be detected and its partial loudness can be calculated as described earlier. The amount of distortion at which the partial loudness attains the value of 0.003 sone is taken as the discrimination threshold. There is a problem, however, in that such a procedure would be suitable only when the modification can be considered as a positive additive distortion of the power spectrum. Because we wish to study arbitrary changes of the spectral envelope, we need to incorporate the treatment of cases in which the spectral level may actually decrease, at least for some spectral components. Figure 2(a) shows an example of such a case. The spectral envelope of the vowel ''a'' is subjected to an decrease in spectral tilt by means of highpass filtering. It can be seen that the low-frequency components are attenuated while the higher-frequency components are amplified. Here we must compute the partial loudness of two distinct types of distortion, one being a positive change in spectral level and the other a negative change.

Our approach to the problem of computing the partial loudness of an arbitrary distortion of the spectral envelope is illustrated by Fig. 3. We first compute separately the excitation patterns of the reference and modified signals. Then based on the channel-wise comparison of these two excitation patterns, we redefine the signal and noise excitation patterns to be used in the partial loudness model as follows. Let E_1 be the excitation pattern of the reference sound and E_2 that of the modified sound. The excitation pattern of the background noise is then defined as $\min(E_1,E_2)$, that of the total sound as $\max(E_1,E_2)$ and that of the signal as $|E_1 - E_2|$. Negative changes are treated in the same way as positive ones, therefore only the absolute value of the difference

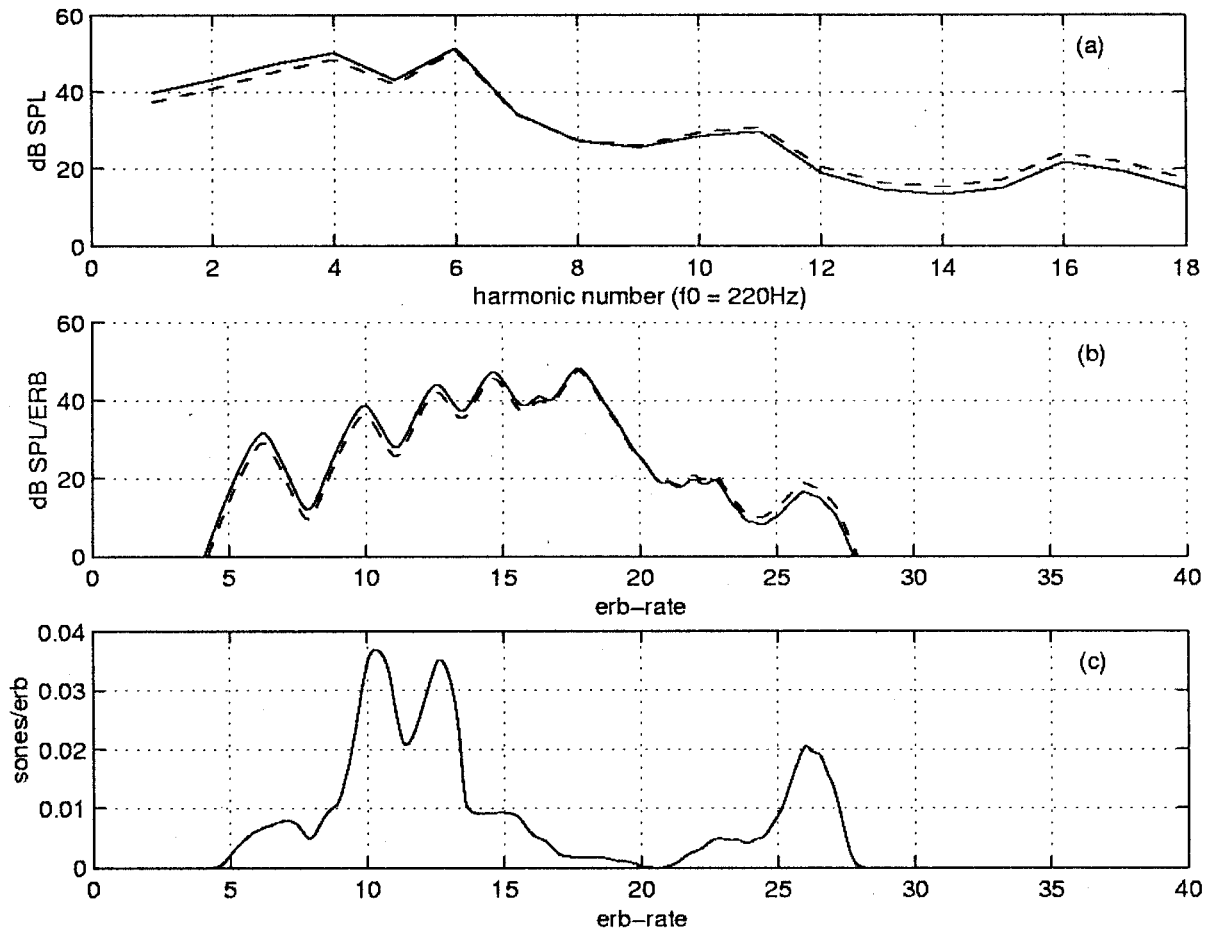


FIG. 2. An example of an arbitrary spectral-envelope modification. (a) The reference (solid line) and modified (dashed line) spectral envelopes of the simulated vowel /a/ with a fundamental frequency of 220 Hz. The modified sound is obtained by applying a single-pole, highpass filter to the reference sound. (b) The excitation patterns of the reference and modified signals. (c) The partial specific loudness distribution.

is of interest. The excitation patterns for the reference and modified signals of Fig. 2(a) are shown in Fig. 2(b). By applying these excitation patterns in the computation of the partial specific loudness of the distortion, we get the distribution shown in Fig. 2(c). The overall partial loudness is obtained by integrating the resulting (always greater-than-zero) values of partial specific loudness. It is the partial loudness measured in this way that we adopt as a measure for the perceptual distance between the sound with excitation pattern E_1 and the sound with excitation pattern E_2. Furthermore, we use a measure that is symmetric, i.e., when the reference sound and the modified sound are exchanged we obtain the same numerical value for the partial loudness of the difference. To what extent this distance measure is capable of predicting audibility thresholds in the context of spectral envelope distortions is investigated by means of the subjective experiment described in the next section.

IV. EXPERIMENT

A. Aim

The aim of the experiment is to validate whether partial loudness, computed according to the model presented earlier, can be used to predict audibility discrimination thresholds for arbitrary modifications of the spectral envelope of steady harmonic complexes. We also will compare our results with two alternative distance metrics, namely the Euclidean distance between excitation patterns, further denoted as the excitation pattern distance, and the Euclidean distance between specific loudnesses, further denoted as the specific loudness

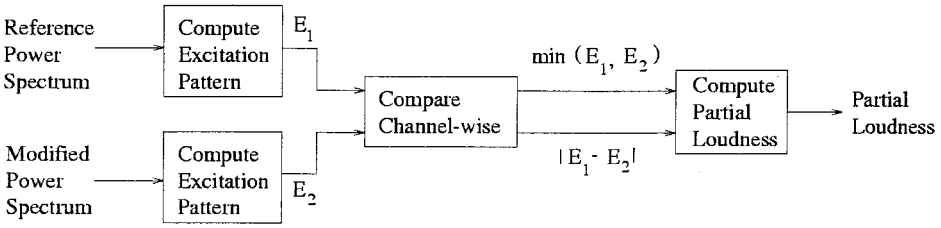


FIG. 3. Schematic diagram illustrating the computation of partial loudness from the excitation distributions of the reference and modified signals.

TABLE I. The 12 experimental conditions with a description of the corresponding spectral modifications.

	Conditions and corresponding modifications			
	Stimulus descriptions			
Condition	$F0$ (Hz)	Phase	Vowel	Modifications
1	220	random	a	harmonic 6, positive
2	220	random	a	harmonics 10–11, positive
3	220	random	a	harmonics 12–15, positive
4	220	random	a	harmonics 12–15, positive and harmonic 6, negative
5	220	random	a	modification of spectral tilt, low pass filter
6	220	random	a	modification of spectral tilt, high pass filter
7	220	random	i	harmonics 1–2, negative
8	220	random	i	harmonics 4–8, positive
9	220	random	i	harmonic 12 positive
10	220	random	i	harmonics 4–8, positive and harmonics 1–2, negative
11	110	random	a	modification of spectral tilt, low pass filter
12	110	regular	a	modification of spectral tilt, low pass filter

distance, applied to the same experimental data. A requirement for a distance measure that is useful for a broad class of speech and musical sounds, is that it should be capable of predicting audibility thresholds for a large variety of spectral envelope modifications. Therefore we chose a representative set of modification conditions for the experiment, distributed over the spectra of two simulated steady vowels, /a/ and /i/. A good measure is expected to produce the same threshold values for distinct conditions, at least for each individual subject. A relative variation, quantifying the range of spread across conditions and defined as the standard deviation of the measured thresholds for the various conditions divided by their mean will, therefore, be used as an indication of the quality of the measures.

B. Stimuli

The reference sound spectra were derived from the amplitude spectra of the vowels synthesised by the cascade combination of an LF model glottal source and a formant filter based on the linear prediction coefficients (LPC) (Fant *et al.*, 1985). A constant overall level of about 55 dB SPL was maintained. The set of modifications was chosen in a way to encompass distinct types of gain changes of the spectral envelope. Table I gives an overview of all the modifications. Specifically, we considered localized spectral amplitude changes at the formant peaks and in the valleys, and also combinations of these changes, both in opposite and in equal directions. The amplitudes of the harmonics were modified by multiplication with a factor close to 1. When more than one harmonic was modified, each harmonic was multiplied by the same factor. For example, the condition 2 of Table I corresponds to a scaling of the harmonics 10 and 11 of the harmonic spectrum by a factor greater than 1. We also investigated modifications that are relatively broadband, or have more spectral spread, by varying the overall spectral tilt. This was achieved by either lowpass filtering to increase the spectral tilt, or by highpass filtering to reduce it. The filter parameters were adjusted so that the overall loudness of the sound was not changed significantly. Figure 4 depicts the set of stimuli and modifications by indicating which harmonic components are affected in each of the conditions.

It is generally accepted that amplitude changes in the spectrum of harmonic sounds are more detectable than phase changes. It was found that for complex tones with a fundamental frequency beyond 150 Hz the maximal effect of phase on timbre is smaller than the effect of changing the slope of the amplitude pattern by 2 dB/oct (Plomp and Steeneken, 1969). Therefore a fundamental frequency of 220 Hz for the vowel-like spectra was used. We applied random phase for the stimuli to maintain an equal distribution of energy within each pitch period. To investigate the influence of changing the fundamental frequency, one of the spectral envelope modifications was repeated at a fundamental frequency of 110 Hz. At this lower fundamental, however, there could exist phase effects, which could lead to temporal cues. Therefore adding a condition with a phase derived from the glottal-pulse model tested the influences of these effects.

In each of the conditions, the spectral amplitudes were modified in small steps corresponding to the calculated partial loudness of the distortion as given by the model. The reference and modified sounds were generated as the sum of harmonics with the specified amplitudes and random phases. Only in the final condition (number 12), were the actual phases provided by the vowel synthesizer applied. The duration of the stimuli was 300 ms with raised cosine ramps of 25 ms at the beginning and end of the signal. For each condition we measured the value of the partial loudness at which the subject was just able to discriminate between the reference sound and the modified sound.

C. Method

Four subjects (JB, JG, PR, and RD) participated in the experiments. The subjects' ages and sexes are presented in Table II. All were young adults with normal hearing and no reported history of hearing impairment. In addition, measurements in Sec. V C showed normal absolute thresholds for all subjects at a frequency of 1 kHz. The stimuli were presented binaurally over headphones at a level of approximately 55 dB SPL to subjects seated in a sound-proof booth. A 3-interval forced-choice adaptive procedure (Levitt, 1971) was used to obtain the thresholds. In this procedure each trial consisted of three stimuli, two stimuli representing the refer-

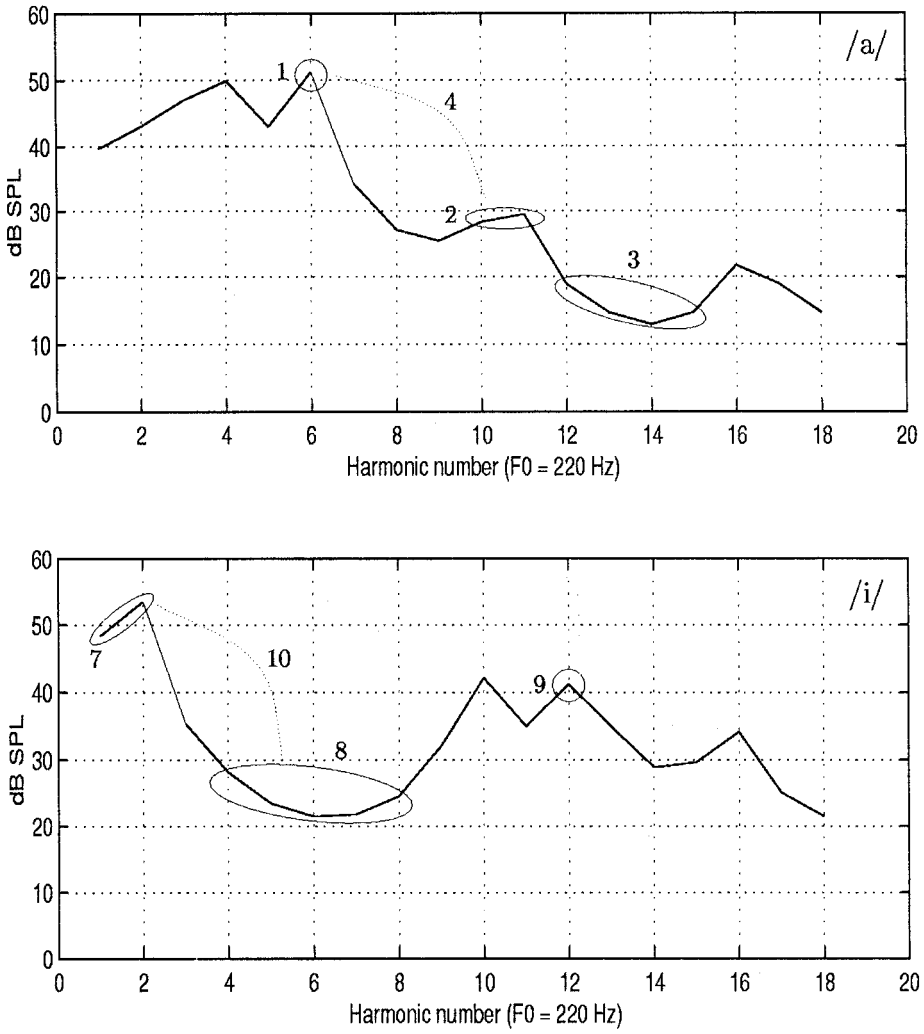


FIG. 4. The two panels show the reference spectra of the stimuli used in the subjective experiment. Top panel: Vowel /a/. Bottom panel: Vowel /i/, both with a fundamental frequency of 220 Hz. The encircled points indicate the modifications of harmonic amplitudes and the corresponding condition numbers. Not in the diagrams are the following four modifications of Table I: condition numbers 5, 6, 11, and 12.

ence sound and one the modified sound. The pause before one trial was 300 ms and the interstimulus interval was 400 ms. The assignment of the odd stimulus to one of the three intervals was randomized. The subject's task was to indicate the odd interval. Immediately after each response, feedback was given indicating whether the response was correct or incorrect. After two correct responses the amount of spectral modification was reduced by one step. After one incorrect response it was increased by one step. The spectral modifications of a stimulus were divided into 20 steps reaching from about 1 sone to 0.001 sone of partial loudness for the modification. A run began with a modification of the spectrum that produced an easily discriminable change. A test run was completed after 12 up–down reversals. A single-run estimate of the partial loudness at threshold was obtained by taking the median of the steps at the last eight reversals of the run. In this way the 70.7% correct detection threshold is measured. For each experimental condition, a final estimate of the partial loudness at threshold for each subject was based on the median of five single-run estimates taken over a period of several days.

TABLE II. Characteristics of the subjects.

	Characteristics of subjects			
	JG	JB	PR	RD
Sex	male	male	female	male
Age	24	30	38	26

V. RESULTS AND DISCUSSION

A. Partial loudness

In Fig. 5 the medians of the partial loudness levels at threshold are given for each subject, where the data of JB, JG, PR, and RD are indicated with a circle, triangle, cross, and star, respectively. The interquartile ranges are indicated with bars.

An examination of the data shows that for most of the conditions, the subjects' thresholds are between 0.003 sone and 0.02 sone, which is close to the range of 0.003 to 0.008 sone used by Moore *et al.* (1997) to predict detection thresholds for simple psychoacoustic stimuli. We particularly note that the thresholds for the widely differing spectral modifications namely localized perturbations (conditions 1, 2, 7, 9) and spread perturbations (conditions 4, 5, 6, 10, 11, 12) fall within the same narrow range. Spectral modifications described by a combination of positive and negative changes appear to be adequately treated by the proposed procedure. Changing the fundamental frequency while maintaining the

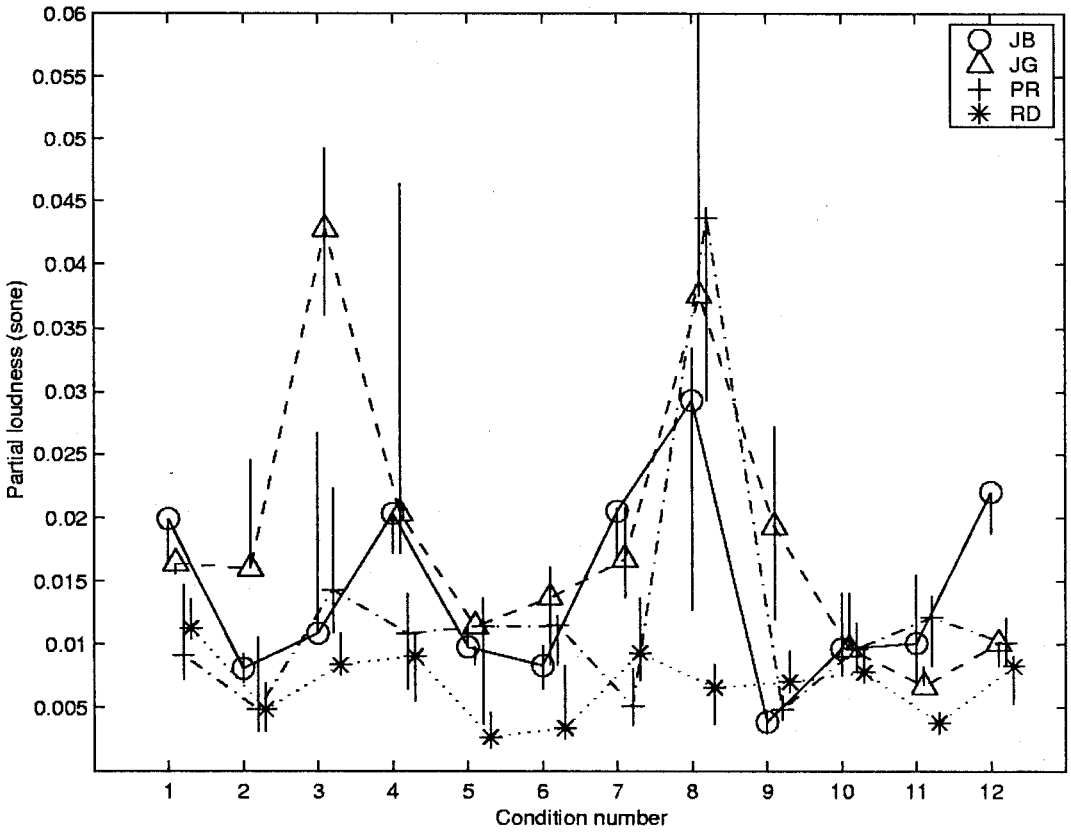


FIG. 5. Experimentally obtained thresholds plotted in terms of partial loudness. The bars indicate the interquartile ranges.

same spectral envelope (conditions 5 and 11) does not impact the accuracy of the predictions. The conditions 3 and 8 corresponding to modifications localized at the valleys of the spectral envelope are clear exceptions, however. For condition 8, three subjects show thresholds that are distinctly higher than thresholds measured for the other conditions while for condition 3, one subject shows high thresholds.

The partial loudness model of Moore *et al.* (1997) is based on the average of results of a large number of experiments involving listeners with normal hearing. The parameters of individual subjects, however, may vary from these average values. The greatest variability is expected in the selectivity of the auditory filters. Hence the predictions of the model cannot be expected to be accurate for all individual listeners. Later in this section we will attempt to correlate the large differences in the threshold levels with possible individual differences in auditory frequency selectivity. First we examine the performance of the Euclidean distance-based metrics on the same data.

B. Comparison with the excitation pattern and specific loudness distances

As discussed in Sec. II, the Euclidean distances between excitation patterns and between specific loudness have both been widely applied in the prediction of vowel quality differences. In contrast to the partial loudness measure, these measures are based on a direct comparison of the internal representations of the reference and test sounds.

The auditory model of Moore *et al.* (1997) considered in this paper was used in the computation of the excitation patterns and the specific loudness for the Euclidean metrics. For each subject and condition, the excitation patterns and the specific loudness of the reference sound and the modified sound corresponding to the just discriminable condition were computed. Figures 6 and 7 show the discrimination thresholds versus condition numbers for the Euclidean distances in the excitation patterns and the specific loudness, respectively.

Next, we compared the measures' performances. A requirement for a measure is that the distance values obtained at threshold for a large variety of spectral modifications are approximately constant for each individual subject. To compare the measures we therefore used the standard deviation of the measured thresholds for the various conditions divided by their mean, which is referred to as the relative variation. Figures 6 and 7 revealed that both the excitation pattern distance and the specific loudness distance display a range of overall variability of discrimination thresholds that is smaller than that of the partial loudness measure. The excitation pattern distance has a slightly smaller range than the specific loudness distance and shows a smaller variability across conditions. This would indicate that the partial loudness measure performs worse than the other two measures. Curves shown in Fig. 8 and Fig. 10 of Moore *et al.* (1997) indicate that both specific and partial loudness show an increased sensitivity with respect to excitation level when approaching the threshold of detectability. In a fair comparison of the measures'

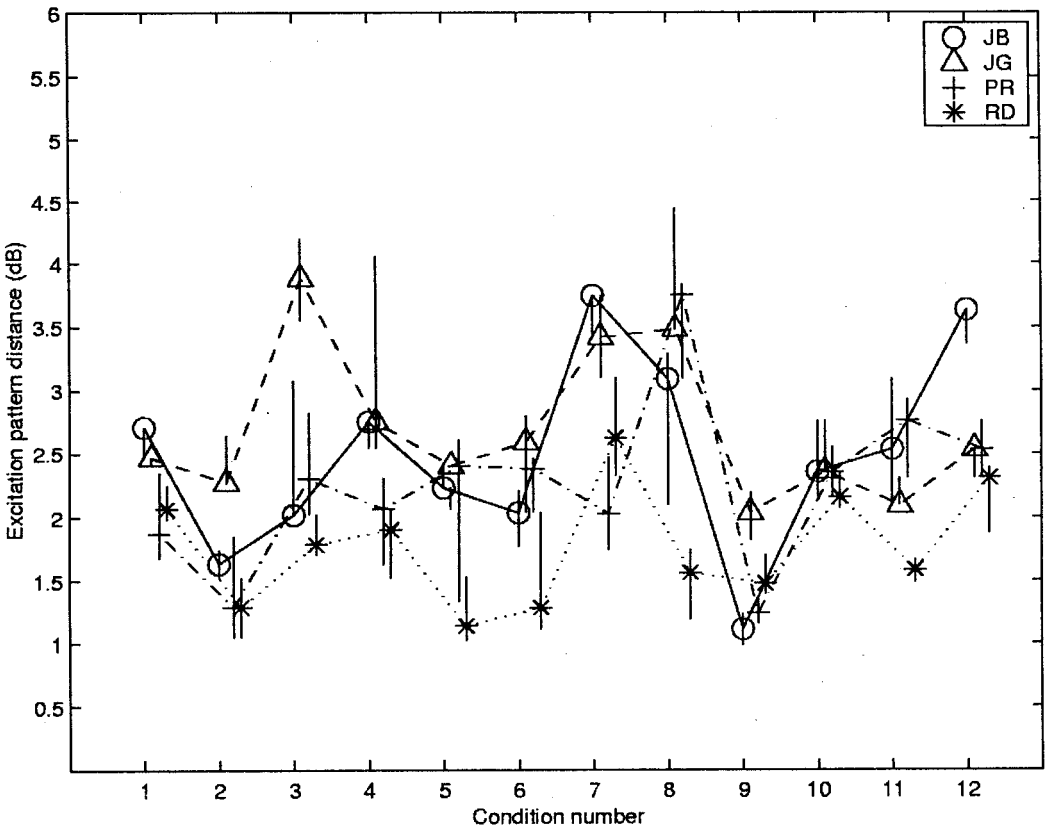


FIG. 6. Experimentally obtained thresholds plotted in terms of excitation pattern distance. The bars indicate the interquartile ranges.

performances, this difference in sensitivity must be compensated for. To objectively compare the relative variations of the three distinct distance measures, we carried out a normalization which takes into account the different sensitivities of the three measures near threshold. The sensitivities of the partial loudness and the specific loudness at threshold were normalized to match the sensitivity of the excitation patterns as follows. Plots of the log of partial loudness distance and the log of specific loudness distance as functions of the log of excitation pattern distance derived from the stimuli that were used, showed bundles of nearly parallel lines. This implies that there is a nearly constant proportional relation between relative variations in the excitation pattern distance and the other two measures. We could, therefore, use the means of slopes of theses curves at the various threshold points as estimates for two normalization factors ρ_{sl} and ρ_{pl}, by which the relative variations in the specific loudness distance and the partial loudness measure, respectively, were divided. The means and standard deviations of the normalization factors are plotted in Table III. Table IV presents the relative variations for the four subjects and the three distance measures before and after normalization. In the unnormalized case, the excitation pattern distance always has the lowest relative variation and it depends on the subject whether the partial loudness measure or the specific loudness distance performs second best. If we regard the mean over the subjects, presented in the last row of the table, the partial loudness measure comes out last and the excitation pattern distance first. In the normalized case, both the partial loudness measure and the excitation pattern distance come out best for two subjects. If we regard the mean over the subjects the partial loudness measure and the excitation pattern distance share the first position.

An error analysis or a presentation of confidence intervals for the results presented in Table IV would be in its place. However, such an analysis turns out to be analytically difficult. Therefore, we performed an error analysis by simulation. First of all, we assumed that the obtained data points had additive Gaussian errors with zero mean. The averages of the interquartile ranges across conditions were used to estimate the standard deviations of these errors for each subject and each distance measure. Then for each subject and each distance measure, the normalized relative variations were computed in 1000 simulation runs in which independent Gaussian errors were added to the computed threshold values. Figure 8 shows the resulting distributions of the relative variations, under the assumption that these distributions are also Gaussian. The distributions of the relative variations can be used to compute for each subject the probabilities that one distance measure performs better than another. These probabilities are presented in Table V. The notations P{PL>EPD}, P{SLD>EPD}, and P{PL>SLD} denote the probabilities that the partial loudness measure performs better than the excitation pattern distance, the specific loudness distance performs better than the excitation pattern distance and the partial loudness measure performs better than the specific loudness distance, respectively. The same probabilities have also been derived directly from the simulation data,

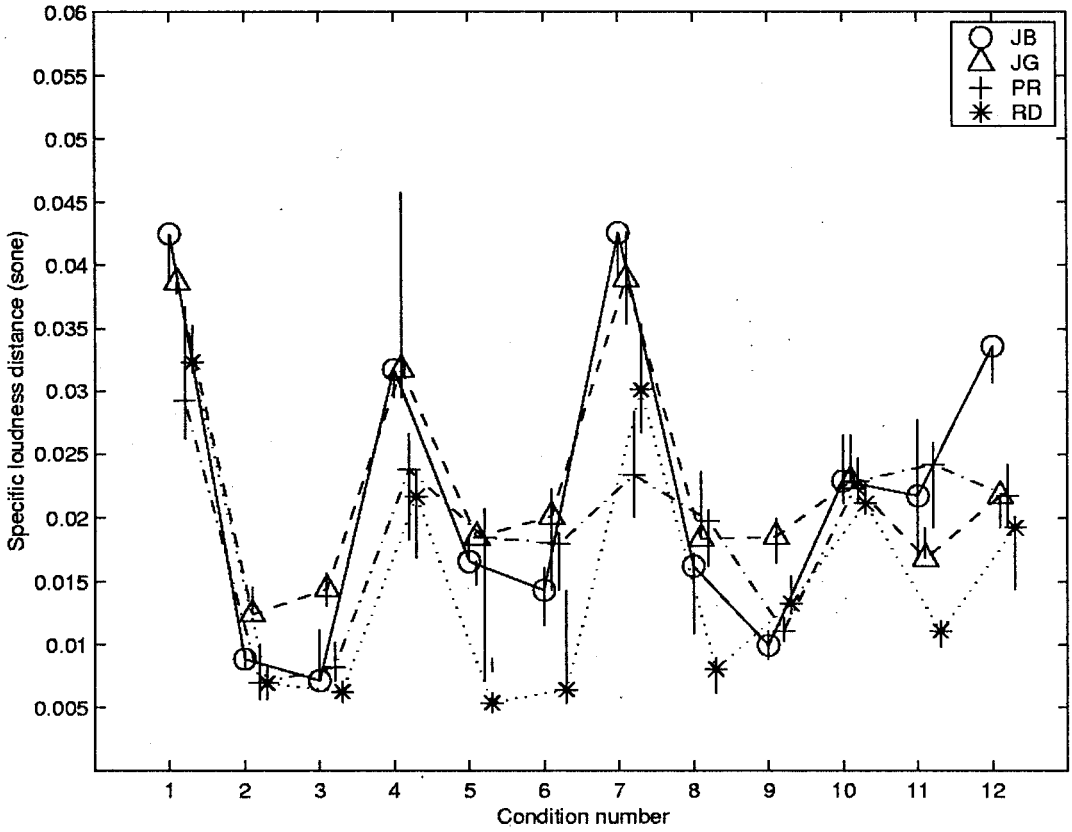


FIG. 7. Experimentally obtained thresholds plotted in terms of specific loudness distance. The bars indicate the interquartile ranges.

without the assumption of Gaussian distributions for the relative variations, but the results only differed in the first decimal of the percentages. These simulations confirm the results based on the data, namely that partial loudness measure and the excitation pattern distance perform equally well and clearly outperform the specific loudness distance.

C. Investigating the variations in partial loudness thresholds

We assumed that the explanation for the particularly high spread in partial loudness threshold values across subjects for the conditions 3 and 8 might be found in individual differences in auditory frequency selectivity. To support such an assumption, we investigated the effect of varying the auditory model filter parameters on the partial loudness levels at threshold, as well as looked for a basis on which any specific alteration of the model's auditory filter parameters may be justified. With this in mind, we picked condition 8 for further investigation.

A high value of calculated partial loudness at threshold implies that the modification is more difficult to detect than predicted by the model. A salient characteristic of condition 8 is that it involves the detection of a signal at a center frequency that is higher than that of the dominant masker. In such a situation it is natural to attribute the difference in detectability to a difference in the upward spread of masking. To follow this possible explanation an additional experiment was carried out to measure the upward spread of masking. We used a masker frequency of 440 Hz and a target frequency of 1000 Hz to create a similar situation as in condition 8. The masker levels were 70, 60 and 55 dB SPL. Table VI shows the medians of the masked thresholds of four sessions. The data in Table VI show a great variability in upward spread of masking and are in line with the assumption that the subjects' differences for condition 8 are due to differences in upward spread of masking. Subject PR shows

TABLE III. Means of normalization factors for the relative variations and their standard deviation.

	Mean	Standard deviation
ρ_{sl}	1.12	0.23
ρ_{pl}	2.19	0.44

TABLE IV. Relative variations of the unnormalized and normalized measures, excitation pattern distance (EPD), specific loudness distance (SLD), and partial loudness measure (PL).

	Relative variations					
	Unnormalized			Normalized		
	PL	SLD	EPD	PL	SLD	EPD
JB	0.53	0.56	0.31	0.24	0.50	0.31
JG	0.60	0.39	0.22	0.27	0.35	0.22
PR	0.84	0.36	0.29	0.38	0.32	0.29
RD	0.39	0.63	0.26	0.18	0.56	0.26
Mean	0.59	0.49	0.27	0.27	0.43	0.27

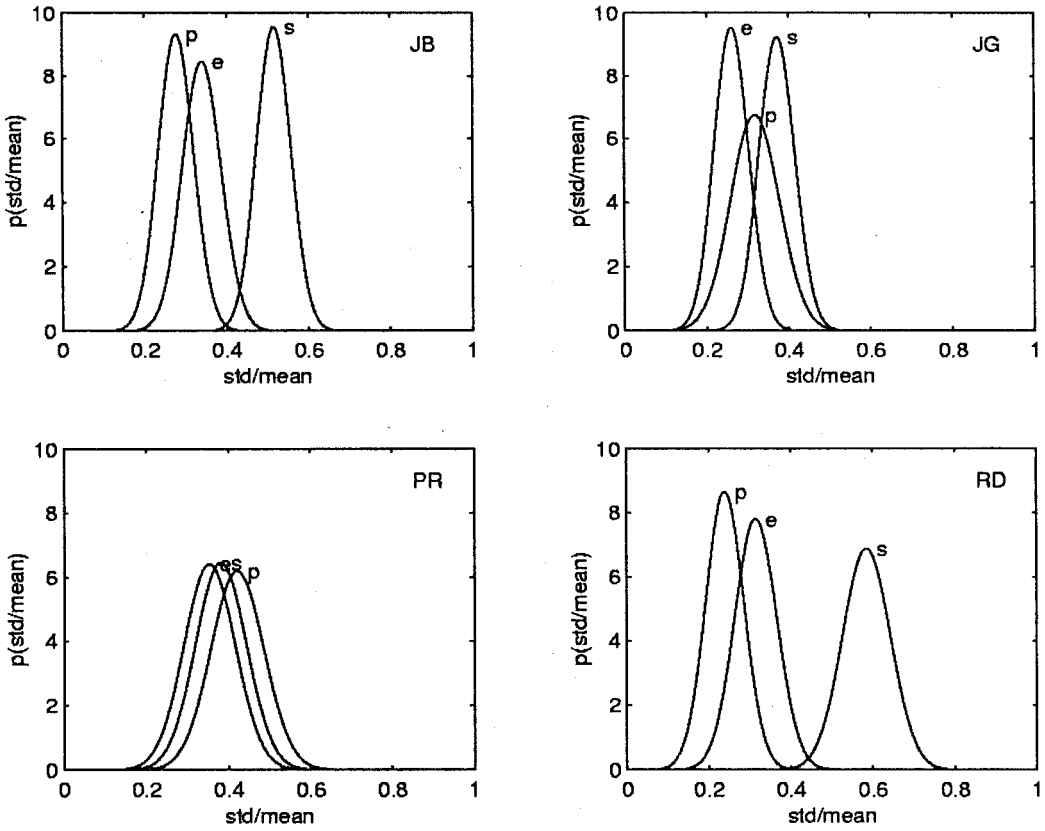


FIG. 8. Distributions of the relative variation for each subject, obtained by simulation.

high masked thresholds accompanied by a high score at condition 8, whereas subject RD shows low masked thresholds accompanied with a low score at condition 8. The subjects JB and JG have masked thresholds between those of PR and RD.

Individual differences in the upward spread of masking can be incorporated in the loudness model by modifying the auditory filter parameters. Decreasing the lower slope of the RoEx(p) filter by decreasing p is the most effective way to increase the predicted upward spread of masking. The filter slope influences the bandwidth (ERB) of the filter however. An alternative way to model the increased spread of masking is by introducing a small, non-zero value of ''r'' in the Roex(p,r) approximation of filter shape (Moore and Glasberg, 1987). The effect of this is to add a low-level skirt to the filter gain function while leaving its passband (upto 30 dB below the filter tip) essentially unchanged. The parameter ''r'' is thought to be related to absolute threshold effects which may vary among individuals (Moore, 1987). Both these approaches were considered separately by computing the parameter, ''*p*'' or ''*r*,'' required to fit the masked threshold data of Table VI for the subject PR, and then applying these modified parameters to calculate the partial loudness value at threshold for each condition. The results are shown in Fig. 9. We see that the value of partial loudness at threshold decreases for the conditions 3 and 8. Although the threshold levels for the other conditions too are affected to some extent, the threshold levels for modifications at the spectral valleys (including condition 2 in which the masker is primarily below the signal frequency) are clearly more sensitive to filter parameter changes. The conditions 1 and 7 can be characterized as being complementary to the spectral valley conditions and show the expected increase in the predicted threshold level with the increased upward spread of masking. So we see that while the modified parameters explain the high threshold of condition 8, they adversely impact the predictions for conditions 1, 2, and 3. However, it must be kept in mind that the modified parameter settings

TABLE V. Probabilities (in percentages) of relative performances of the measures per subject obtained by simulation.

	P{PL>EPD}	P{SLD>EPD}	P{PL>SLD}
JB	85	0	100
JG	21	3	78
PR	24	49	34
RD	86	0	100

TABLE VI. Median masked thresholds in dB SPL for three different masker levels in the upward-spread-of-masking experiment.

	Masked threshold			
	Subject			
Masker level	JG	JB	PR	RD
70	24.50	20.75	32.00	12.25
60	9.25	6.25	21.50	4.25
55	5.75	2.50	17.75	2.00

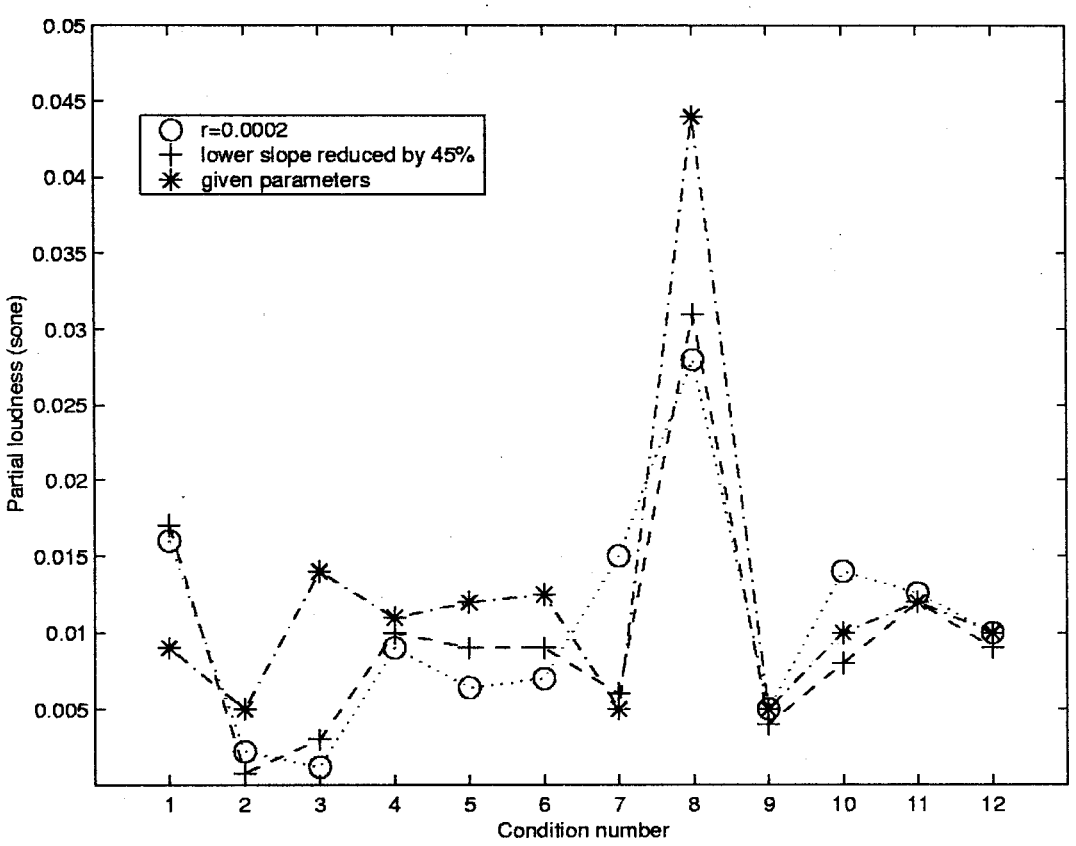


FIG. 9. Variation in partial loudness threshold levels with auditory filter selectivity computed for the experimental data of subject PR. The auditory frequency selectivity, modelled by the RoEx(p,r) function, was manipulated by changing the parameters p and r independently.

were derived from a masking situation applicable to the conditions 7, 8, and 10 and therefore may not be completely relevant for the other conditions.

The sensitivities of the Euclidean distance-based metrics to auditory filter parameter changes were also examined. It was found that the excitation pattern distance is also sensitive to the parameter changes but to a significantly lesser extent than the partial loudness measure. The specific loudness distance on the other hand is relatively insensitive to changes in filter parameter settings. These facts can also be seen in Figs. 6 and 7 where we observe less variation among subjects at any given condition in the specific loudness metric as compared to the excitation pattern metric. On the other hand, the specific loudness distance at threshold appears to be more dependent on the actual nature of the spectral modification.

The relative variations of the measures with the changed parameters were computed for subject PR. In Table VII the relative variations and performances of the measures are presented for the case of the lower slope reduced by 45%, the case of $r=0.0002$ and the original case. For the case of the lower slope reduced by 45% the normalization factors were recomputed as $\rho_{sl}=1.10$ and $\rho_{pl}=2.19$. For the case $r=0.0002$ the factors were recomputed as $\rho_{sl}=1.01$ and $\rho_{pl}=2.35$. The results in Table VII show that the performance of the partial loudness measure has improved slightly after adjusting the filter parameters. It can also be noted that the performance of the partial loudness measure has become better than the other two measures and that the results of subject PR have become more in line with the results of the other subjects.

VI. CONCLUSIONS

The partial loudness measure computed from the auditory model of Moore *et al.* (1997) was proposed and adapted for the problem of predicting perceptual differences caused by spectral envelope modifications of steady sounds. The partial loudness measure is based on a spectral model and does not take into consideration phase effects. The effectiveness of this measure for the prediction of discrimination thresholds of spectral envelope modifications in simulated vowel sounds was studied by means of subjective experiments. Our results indicate that the assumptions of the model

TABLE VII. Relative variations of the normalized measures and probabilities of the relative performances of the measures obtained by simulation for subject PR with adapted parameters of the auditory filter. Case 1 represents the reduced lower slope by 45%. Case 2 represents the parameter change $r=0.0002$.

	Relative variations			Relative performances of measures		
Case	PL	SLD	EPD	P{PL>EPD}	P{SLD>EPD}	P{PL>SLD}
1	0.35	0.44	0.36	69	47	87
2	0.31	0.44	0.39	82	50	85
Original	0.38	0.32	0.29	24	49	34

are justified and that the experimentally determined thresholds are reasonably close to the predicted values. Our results provide a range for the discrimination thresholds applicable to realistic data such as steady vowels. A score of 0.01 sone is a good estimate, although there appear easily variations of a factor of 0.5 to 2. Previously proposed vowel quality distance measures were also evaluated on the same experimental data. A relative variation, quantifying the range of spread across conditions, was defined in order to compare the measures. At a first glance, the Euclidean distance between excitation patterns gives a narrower range of spread in discrimination thresholds compared to the partial loudness measure. However, it was argued that the greater variability of the partial loudness measure and of the specific loudness distance is due to the sensitivity of these measures at threshold. Once normalized to match the sensitivity of the excitation pattern distance at threshold, the variability of both the partial loudness measure and the specific loudness distance is reduced and the performances of partial loudness measure and excitation pattern distance are similar and clearly better than that of the specific loudness distance.

The occasionally very large deviations from the predicted value were found to be related to individual differences in the upward spread of masking. It was found that the computed partial loudness measure is sensitive to changes in the auditory model filter parameters. This sensitivity is most pronounced for modifications localized at spectral envelope valleys. An attempt was made to model individual differences as measured in a masked threshold experiment and to link the results with the experimentally measured partial loudness thresholds for one subject. In this way we were able to bring the results of this subject more in line with the results of the other subjects. This effort illustrates an approach to explaining individual variation in behavioral results, which is potentially useful in the development of robust tools for use in clinical settings. It is worthwhile to investigate in how far careful tuning of filter parameters for subjects can improve the results in terms of the relative variation.

Summarizing our results on the prediction of the discrimination thresholds for spectral envelope modifications of vowel sounds, we see that the partial loudness measure as well as the excitation pattern distance are equally appropriate measures for predicting audibility discrimination thresholds.

Moore investigated extensively supra-threshold differences for sounds in noise and evaluated the predictability of partial loudness on this in particular. It will be of interest to extend the present work to evaluate the performance of partial loudness measure in the prediction of supra-threshold differences in vowel quality.

ACKNOWLEDGMENTS

The work described in this paper was carried out while the first author was visiting IPO. The authors would like to thank Steven van de Par and Jeroen Breebaart for their help with setting up the experiments.

Bladon, R. A. W., and Lindblom, B. (**1981**). ''Modeling the judgement of vowel quality differences,'' J. Acoust. Soc. Am. **69**, 1414–1422.

Fant, G., Liljencrants, J., and Lin, Q. (**1985**). ''A four-parameter model of glottal flow,'' Speech Transmission Laboratory Quarterly Progress Report 4/85, KTH, 1–3.

Gagné, J. P., and Zurek, P. M. (**1988**). ''Resonance-frequency discrimination,'' J. Acoust. Soc. Am. **83**, 2293–2299.

Kewley-Port, D. (**1991**). ''Detection thresholds for isolated vowels,'' J. Acoust. Soc. Am. **89**, 820–829.

Kewley-Port, D., and Zheng, Y. (**1998**). ''Auditory models of formant frequency discrimination for isolated vowels,'' J. Acoust. Soc. Am. **103**, 1654–1666.

Levitt, H. (**1971**). ''Transformed up-down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 971–995.

Moore, B. C. J. (**1987**). ''Distribution of auditory filter bandwidths at 2 kHz in young normal listeners,'' J. Acoust. Soc. Am. **81**, 1633–1635.

Moore, B. C. J., and Glasberg, B. R. (**1987**). ''Formulae describing frequency selectivity as a function of frequency and level and their use in calculating excitation patterns,'' Hear. Res. **28**, 209–225.

Moore, B. C. J., Glasberg, B. R., and Baer, T. (**1997**). ''A model for the prediction of thresholds, loudness and partial loudness,'' J. Audio Eng. Soc. **45**, 224–240.

Plomp, R., and Steeneken, H. J. M. (**1969**). ''Effect of phase on the timbre of complex tones,'' J. Acoust. Soc. Am. **46**, 409–421.

Plomp, R. (**1976**). *Aspects of Tone Sensation* (Academic, London).

Quackenbush, S. R., Barnwell, T. P., and Clements, M. A. (**1988**). *Objective Measures of Speech Quality* (Prentice Hall, Englewood Cliffs, New Jersey).

Schroeder, M. R., Atal, B. S., and Hall, J. L. (**1979**). *Objective Measure of Certain Speech Signal Degradations Based on Masking Properties of Human Auditory Perception. Frontiers of Speech Communication Research* (Academic, New York).

Sommers, M. S., and Kewley-Port, D. (**1996**). ''Modeling formant frequency discrimination of female vowels,'' J. Acoust. Soc. Am. **99**, 3770–3781.

Stevens, S. S. (**1957**). ''On the psychophysical law,'' Psychol. Rev. **64**, 153–181.

van der Heijden, M., and Kohlrausch, A. (**1994**). ''Using an excitation-pattern model to predict auditory masking,'' Hear. Res. **80**, 38–52.

Zwicker, E., and Scharf, B. (**1965**). ''A model of loudness summation,'' Psychol. Rev. **72**, 3–26.

Zwicker, E., and Fastl, H. (**1990**). *Psychoacoustics—Facts and Models* (Springer, Berlin).

Revisiting relations between loudness and intensity discrimination

William S. Hellman[a)]
Department of Physics and Hearing Research Center, Boston University, Boston, Massachusetts 02215

Rhona P. Hellman
Auditory Perception Laboratory, Department of Psychology and Institute for Hearing, Speech, and Language, Northeastern University, Boston, Massachusetts 02115

(Received 8 November 1999; accepted for publication 27 February 2001)

A comparison is made between the variation of ΔL_{jnd} with L (loudness), based on the beat-detection data of Riesz at 1 kHz [Phys. Rev. **31**, 867–875 (1928)], and analogous relations obtained from a cross section of studies. Data analysis shows that only beat detection exhibits the degree of level-dependent variation in slope relating log (ΔL_{jnd}) to log (L) described in a recent paper by Allen and Neely [J. Acoust. Soc. Am. **102**, 3628–3646 (1997)]. Moreover, the slope variation determined from beat detection is not dependent on the detailed shape of the loudness function. The results imply that Allen and Neely's strong conclusions about the dependence of ΔL_{jnd} on L are too tightly coupled to Riesz's methodology to be generally applicable. © *2001 Acoustical Society of America.*
[DOI: 10.1121/1.1366373]

PACS numbers: 43.66.Ba, 43.66.Cb, 43.66.Fe [RVS]

I. INTRODUCTION

Recently, there has been renewed interest in models connecting loudness and intensity discrimination and their implications (e.g., Allen and Neely, 1997; Buus *et al.*, 1997; Hellman and Hellman, 1990; Schlauch *et al.*, 1995; Zwislocki and Jordan, 1986). Based on the suggestion of McGill and Goldberg (1968a, b) that ΔL is proportional to $L^{1/2}$, Hellman and Hellman (1990) showed that the overall form of the loudness function obtained by Hellman and Zwislocki (1963) can be predicted from the detailed shapes of intensity-discrimination functions measured with different experimental paradigms in a variety of studies. Using the same general approach, Allen and Neely (1997) calculated the change in loudness, ΔL_{jnd} due to a just-noticeable change in intensity ΔI_{jnd} for a 1-kHz tone and for broadband noise. The calculations were made with specific pairs of loudness and intensity-discrimination functions by means of the relation

$$\Delta L_{\text{jnd}} = L(I + \Delta I_{\text{jnd}}) - L(I), \tag{1}$$

where $L(I)$ is the loudness function that is associated with the intensity-discrimination measurement paradigm. Within the model, ΔL_{jnd} is defined as the change in loudness for a given loudness function resulting from a just-noticeable change in intensity. By combining the loudness function of Fletcher and Munson (1933) with the intensity-discrimination data of Riesz (1928) for a 1-kHz tone, Allen and Neely (1997) found that ΔL_{jnd} can be represented as a power function of L in the form

$$\Delta L_{\text{jnd}} = cL^{1/p}, \tag{2}$$

where $1/p$ is an intensity-dependent exponent. According to their calculations, $p=2$ below a sensation level (SL) of 20 dB, increases to 3 between 20 and 60 dB SL, and decreases to 1 as the level increases to about 80 dB SL and above. These values of p correspond to slope values of 0.5, 0.33, and 1.0 in log–log coordinates. This result differs from the Poisson-like relation, $\Delta_{\text{jnd}} = cL^{0.5}$ used by Hellman and Hellman (1990). In light of these findings, we reexamine the loudness-intensity-discrimination relation for a 1-kHz tone, the most widely studied psychoacoustical stimulus.

According to Riesz (1928), intensity-discrimination measures were obtained by the method of beat detection for a ΔF of 3 Hz. This rate of fluctuation was chosen because it evoked the smallest discrimination values. The loudness function used by Allen and Neely (1997) was obtained from calculated values reported in Table III by Fletcher and Munson (1933) in their seminal article. Fletcher and Munson's tabled values were based on three sets of loudness-matching data: (1) binaural-to-monaural loudness matches, (2) loudness matches between a single tone and a two-tone complex, and (3) loudness matches between a single tone and a ten-tone harmonically spaced complex. A smooth function was fitted to the three data sets. However, in the 1930s, no measurements of the loudness relation between a multi-tone complex and a single tone were available. Thus, the extent of the effects of mutual masking among the components in the tonal complexes, especially for the ten-tone complex, on loudness magnitude and the rate of loudness growth at moderate to high levels could only be estimated. Therefore, above a loudness level of about 40 phons the loudness values in Table III are an approximation.

Another concern is the data of Riesz (1928); they do not appear to be representative of intensity-discrimination data obtained in more recent studies (e.g., Florentine *et al.*, 1987; Viemeister and Bacon, 1988; Zwislocki and Jordan, 1986). A question of some relevance is whether the results of Allen and Neely are due to the slight deviations from linearity above 40 phons in the initial loudness function of Fletcher and Munson (1933), to the intensity-discrimination data of

a)Electronic mail: hellman@buphyc.bu.edu

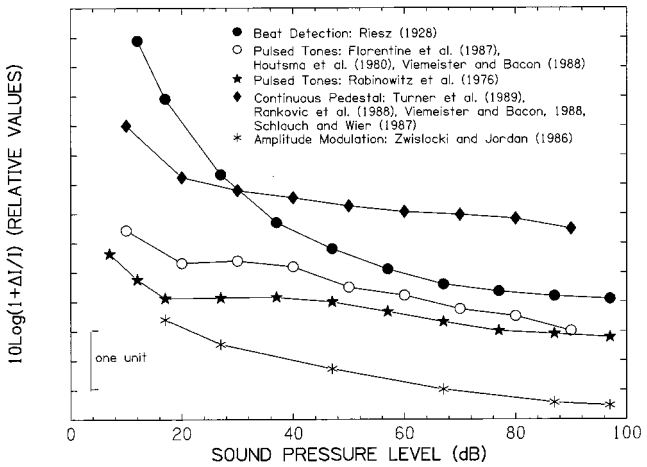


FIG. 1. Comparison between beat-discrimination data of Riesz (1928) and intensity-discrimination data obtained by pulsed-sinusoid, continuous pedestal, and amplitude-modulation procedures from a cross section of studies. The symbols denote the results of each study or group of studies.

Riesz (1928), or to both. In order to provide some clarification of these important issues, it was deemed necessary to evaluate and analyze a broader range of data. To this end, we extend the analysis of Allen and Neely (1997) by determining the dependence of ΔL_{jnd} on L from intensity-discrimination measures for beat detection as well as for pulsed-sinusoid, continuous-pedestal, and amplitude-modulation data.[1] All produce results significantly different from those obtained from Riesz's data.

II. PROCEDURE

The independent variable for representing the dependence of loudness on level is taken to be the decibel $y = 10 \log (I/I_0)$, where I_0 is 10^{-16} W/cm^2. The ΔL_{jnd}'s were calculated from $10 \log (L)$. It then follows that $\Delta y = 10 \log (1+\Delta I/I)$. Thus, if $10 \log (L)$ is adequately represented either graphically or analytically as a function of y, then the value of ΔL_{jnd} corresponding to a Δl_{jnd} can be obtained by computing $L(y+DL(y))-L(y)$, where $DL(y) = \Delta y_{\text{jnd}} = 10 \log (1+\Delta I_{\text{jnd}}/I)$.

Figure 1 compares Weber fractions plotted as $10 \log (1+\Delta I/I)$ from the beat-detection measures of Riesz (1928) to intensity-discrimination data obtained from a variety of psychophysical procedures and studies. To make this comparison, the measures of Riesz were converted from sensation level to sound pressure level by adding 7 dB, the monaural threshold value published by Fletcher (1953, Table 20) for a 1-kHz tone, to the reported group sensation levels.[2] The curves are shifted vertically with respect to each other to clarify the presentation.

According to Fig. 1, the shape of the intensity-discrimination function obtained from the method of beat detection differs distinctly from the shapes obtained by all other methods. This disparity was previously noted by Jesteadt *et al.* (1977) for pulsed tones and by Zwislocki and Jordan (1986) for amplitude modulation. More recently, Wojtczak and Viemeister (1999) found that Riesz's results differ from both continuous-pedestal and amplitude-modulation data.

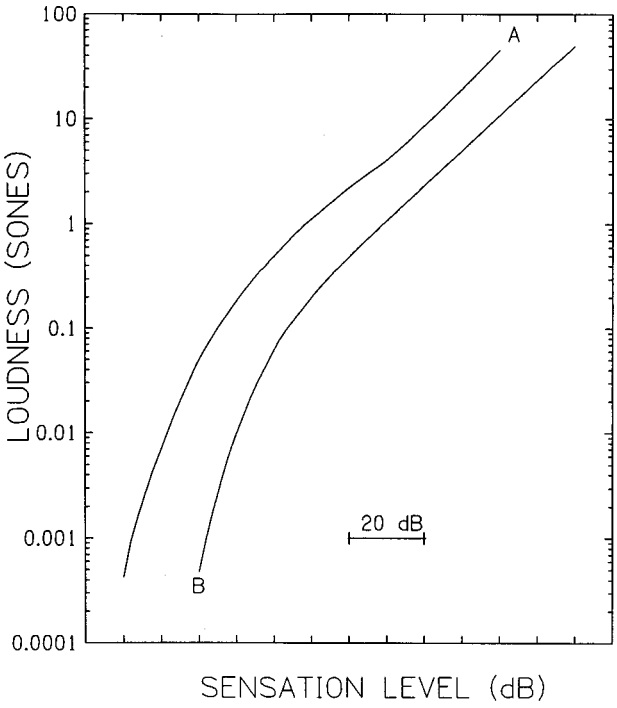


FIG. 2. Relation between loudness in sones and sensation level in dB determined for a 1-kHz tone. Curve A shows the initial loudness function in Table III by Fletcher and Munson (1933); curve B shows the modified loudness function derived by Fletcher (1935, 1940) from loudness matches and also, from fractionation data. For clarity, curve B is displaced by 20 dB along the abscissa from curve A.

The next step in our analysis was to relate the intensity-discrimination functions in Fig. 1 to the 1-kHz loudness functions obtained by Fletcher (Fletcher and Munson, 1933; Fletcher, 1935, 1940). These functions are plotted in Fig. 2 as curves A and B. All loudness values are given in standard sones units.[3] Fletcher's (1935, 1940) modified loudness function in curve B is based on both the loudness-matching data obtained at Bell Laboratories in the 1930s and on the halving and doubling data available at that time. The inclusion of the additional loudness measures filled gaps in the matching data above 40 phons, enabling Fletcher to make a stronger case for the overall shape of the loudness function between 40 and 80 phons.

III. RESULTS

Figure 3 shows the variation of $\log (\Delta L_{\text{jnd}})$ with $\log (L)$ calculated from the Weber fractions in Fig. 1. The insets display the variation of $\log (\Delta L_{\text{jnd}}/L)$ with $\log (L)$. Panel (A) presents the results determined from Fletcher and Munson's (1933) initial loudness function, and panel (B) presents those determined from Fletcher's (1935, 1940) modified function. The dashed lines are the linear fits; the solid lines connect the points. As in Fig. 1, the curves have been shifted vertically with respect to each other to clarify the presentation. The symbols refer to the same studies as in Fig. 1.

A. The variation of ΔL with L

Only the curves obtained from Riesz's (1928) data exhibit the degree of level-dependent variation in slope described by Allen and Neely (1997). In panel (A), the slope of

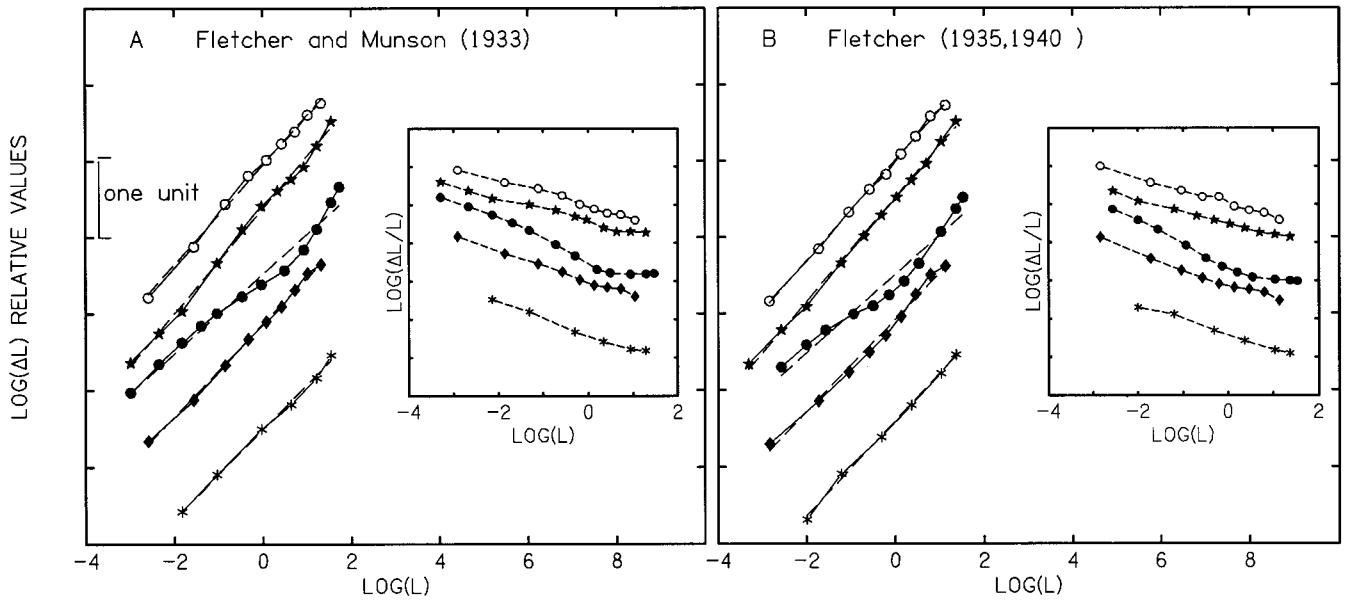


FIG. 3. The variation of log (ΔL_{jnd}) with log (L) calculated from the intensity-discrimination data in Fig. 1. Results derived from Fletcher and Munson's (1933) initial loudness function are presented in panel (A) and those derived from Fletcher's (1935, 1940) modified loudness function are presented in panel (B). The insets show the relation between log ($\Delta L_{jnd}/L$) and log (L) for each loudness function. The symbols represent the same studies as in Fig. 1.

the curve obtained from Riesz's values is 0.56 from 0 to 20 dB SL, 0.32 from 20 to 50 dB SL, and 0.51 from 50 to 65 dB SL. Above 65 dB SL it increases to 1.0. This slope variation is in good agreement with the calculation of Allen and Neely (1997). Likewise, the features that characterize the calculations of log (ΔL_{jnd}) in panel (A) are preserved in the calculations in panel (B). Here, the calculated slope values are 0.49 from 0 to 20 dB SL, 0.30 from 20 to 50 dB SL, 0.59 from 50 to 65 dB SL, and 0.90 above 65 dB SL. By comparison, in both panels (A) and (B) the curves obtained from the other data sets are well approximated by a linear function.

Table I shows that the standard deviation from a linear fit to Riesz's (1928) data is 0.117 derived from Fletcher and Munson's (1933) initial loudness function and 0.125 derived from Fletcher's (1935, 1940) modified loudness function.[4] In contrast, the standard deviations of the linear fits determined for the other paradigms are much smaller; they range from 0.040 to 0.060 for Fletcher and Munson's initial function, and from 0.029 to 0.071 for Fletcher's modified function. The corresponding mean slope values range from 0.589 to 0.696 for Fletcher and Munson's initial function, and from 0.603 to 0.690 for Fletcher's modified function.

For further comparison with the results of Allen and Neely (1997), a similar analysis was performed using Miller's (1947) loudness-matching and intensity-discrimination data for broadband noise together with Fletcher and Munson's (1933) initial loudness function. The calculations produced a log (ΔL_{jnd}) vs log (L) function with a standard deviation from a linear fit of 0.067 in accord with those from the various pure-tone paradigms, excluding Riesz's (1928) data.

B. The approach to Ekman's law ($\Delta L/L = \text{const}$)

Another way to depict the relations in Fig. 3 is to plot log ($\Delta L_{jnd}/L$) as a function of log (L). The insets display the results. Replotted in this format, a region with zero slope corresponds to $\Delta L/L = \text{const}$ in accord with Ekman's law (1956, 1959). The inset in panel (A) reveals that the curves derived from the data of Riesz (1928) and from the data of Rabinowitz *et al.* (1976) approach Ekman's law when loudness is above about 3 sones. The three other curves do not exhibit this behavior nor do any of the curves shown in the inset in panel (B), with the exception of the one generated

TABLE I. Mean slopes and standard deviations of linear fits to derived ΔL vs L functions at 1 kHz from a variety of intensity-discrimination paradigms.

	Fletcher and Munson (1933)		Fletcher (1935, 1940)	
Study	Mean	σ	Mean	σ
Beat detection Riesz (1928)	0.500	0.117	0.483	0.125
Pulsed sinusoids Florentine *et al.* (1987) Houtsma *et al.* (1980) Viemeister and Bacon (1988)	0.658	0.049	0.656	0.029
Pulsed sinusoids Rabinowitz *et al.* (1976)	0.696	0.056	0.690	0.044
Continuous pedestal Turner *et al.* (1989) Rankovic *et al.* (1988) Viemeister and Bacon (1988) Schlauch and Wier (1987)	0.610	0.040	0.603	0.071
Amplitude modulation Zwislocki and Jordan (1986)	0.589	0.060	0.621	0.039
Average modulo Riesz	0.638	0.051	0.642	0.046

from Riesz's data. Overall, these data do not support Allen and Neely's (1997) contention that Ekman's law is satisfied at high levels as a general rule.

IV. DISCUSSION

Taken together, our results make it evident that the slope variation generated by the beat-detection data of Riesz (1928) is quite robust despite small differences in shape between Fletcher's (Fletcher and Munson, 1933; Fletcher, 1935, 1940) two loudness functions. Excluding Riesz's data, the mean slope of the linear fits determined from the nine studies considered in this article is 0.64. The associated standard deviations are 0.051 from Fletcher and Munson's (1933) initial loudness function and 0.046 from Fletcher's (1935, 1940) modified function. The mean slope of 0.64 reveals that the relation between ΔL_{jnd} and L is close to but not quite Poisson. This result means that the original Poisson assumption used in the McGill and Goldberg (1968a, b) model relating loudness to intensity discrimination, and in its generalization by Hellman and Hellman (1990), should be modified. Within such models, an increase in slope from 0.50 to a value between 0.60 and 0.70 would yield an improved calculation of the intensity-discrimination functions derived from the loudness functions of Fletcher and Munson (1933), Fletcher (1935, 1940), and Hellman and Zwislocki (1963).

The results also cast doubt on Allen and Neely's assertion that the increase in slope to a value of 1.0 of the $\log(\Delta L_{\mathrm{jnd}})$ vs $\log(L)$ relation arises from a ''near miss'' to Stevens's power law. Our analysis reveals that the slope variation between $\log(\Delta L_{\mathrm{jnd}})$ and $\log(L)$ predicted from Riesz's data is preserved even when the high-level segment of the loudness function in a log–log plot is approximated by a power function. As noted in the Introduction, the slight deviations from linearity observed above about 40 phons in Fletcher and Munson's (1933) initial loudness function are mainly ascribed to the effects of mutual masking among the components in a multi-tone complex on the growth of loudness at moderate-to-high levels (e.g., Scharf, 1964, 1978). According to Fig. 3 these deviations in shape of the loudness function have no effect on the relation between $\log(\Delta L_{\mathrm{jnd}})$ and $\log(L)$. This result holds irrespective of the experimental paradigm used to measure intensity discrimination. It also holds for loudness measures obtained with pulsed tones (e.g., Hellman and Zwislocki, 1963).[5] The implication is clear. The near-miss to Weber's law cannot be accounted for by the moderate-to-high level deviations of the loudness function from power-law behavior as Allen and Neely contend. A more likely basis for the near-miss relation is the excitation-pattern model proposed by Florentine and Buus (1981).

Finally, the explanation for the discrepancy between the beat-detection data of Riesz (1928) and the results of all other intensity-discrimination paradigms remains elusive. Especially noteworthy is the finding that the $\log(\Delta L_{\mathrm{jnd}}/L)$ vs $\log(L)$ functions derived from amplitude-modulation and the beat-detection method of Riesz differ in overall shape. By comparison, amplitude-modulation and continuous-pedestal methods give rise to similar functions (Johnson *et al.*, 1993; Turner *et al.*, 1989; Wojtczak and Viemeister, 1999). The contrast between the results of beat-detection and amplitude-modulation paradigms is of interest because the temporal pattern of the beating signals is similar to the pattern for amplitude modulation. However, the spectral energy distribution for the two experimental paradigms is different. This difference does not appear to affect the perceptual response at the detection threshold for low modulation rates (Kohlrausch *et al.*, 2000), but it does not preclude the possibility that the perceptual response is altered at suprathreshold levels. For example, starting from the same threshold value the loudness of a 1-kHz tone in noise is independent of noise bandwidth at low sound pressure levels, but it depends strongly on noise bandwidth at moderate-to-high levels (Hellman, 1970). Additional intensity-discrimination measurements using the beat-detection method are needed to determine what aspects of the stimulus configuration produce the observed subtle, but important, perceptual differences.[6] Until the perceptual effects of beating tones are better understood, specific predictions of the relation between ΔL_{jnd} and L may not be justified.

ACKNOWLEDGMENTS

This work was supported in part by NEDO, Japan. We thank Torsten Dau for constructive comments about amplitude modulation.

[1]Although beat detection can be classified as amplitude modulation, by amplitude modulation here we mean that the physical spectrum contains the carrier and both sidebands.

[2]Threshold values were not available in all studies. Thus, for uniformity, sound pressure level was used as the independent variable.

[3]Fletcher and Munson's LU units were arbitrary. According to Fletcher (1935, p. 6), to construct a loudness function ''a value of N=1 was taken arbitrarily for a pressure 1 (in Bars) that is at the threshold of hearing.'' He found that ''for such a choice, a pressure ratio of 100 corresponded very closely to a loudness N of 1000. For this reason, the pair of values p=100 and N=1000 was taken as the starting point (Fletcher, 1935, p. 7). These statements do not indicate that Fletcher's LU scale is based on spike counts, nor do they suggest that the LU scale proposed by Fletcher and Munson is a more natural scale than the sone scale as Allen and Neely (1997) contend. In contrast, many experiments indicate that a 40 dB SL tone presented binaurally is closely coupled to a loudness of 1 in accord with the standard sone function (e.g., Hellman and Zwislocki, 1961; Stevens, 1975; Zwislocki and Goodman, 1980). Therefore, conventional sone units are used in Fig. 2.

[4]The standard deviations are included in Table I to quantify the visually larger deviations from linearity observed in Riesz's (1928) data compared to the results from the other intensity-discrimination paradigms. These measures do not imply that Riesz's data are noisier or have greater uncertainty than the other data. However, they do make clear that Riesz's data uniquely do not support a linear approximation for the $\log(\Delta L_{\mathrm{jnd}})$ vs $\log(L)$ functions.

[5]The analysis using the loudness function of Hellman and Zwislocki (1963), as in a previous paper (Hellman and Hellman, 1990), produces virtually the same results.

[6]Our calculations show that the alternative formula proposed by Wojtczak and Viemeister (1999) does not produce a significant change in the shape of the $\log(\Delta L_{\mathrm{jnd}})$ vs $\log(L)$ function for Riesz's (1928) data. The slope and standard deviation from a linear fit are 0.581 and 0.129, respectively, in good agreement with the results in Table I.

Allen, J. B., and Neely, S. T. (**1997**). ''Modeling the relation between the intensity just-noticeable difference and loudness for pure tones and wide-band noise,'' J. Acoust. Soc. Am. **102**, 3628–3646.

Buus, S., Florentine, M., and Poulsen, T. (**1997**). ''Temporal integration of loudness, loudness discrimination, and the form of the loudness function,'' J. Acoust. Soc. Am. **101**, 669–680.
Ekman, G. (**1956**). ''Discriminal sensitivity on the subjective continuum,'' Acta Psychol. **12**, 233–243.
Ekman, G. (**1959**). ''Weber's law and related functions,'' J. Psychol. **47**, 343–352.
Fletcher, H. (**1935**). ''Newer concepts of the pitch, loudness, and timbre of musical tones,'' J. Franklin Inst. **220**, 405–429.
Fletcher, H. (**1940**). ''Auditory patterns,'' Rev. Mod. Phys. **12**, 47–65.
Fletcher, H. (**1953**). *Speech and Hearing in Communication* (Van Nostrand, Princeton, NJ).
Fletcher, H., and Munson, W. A. (**1933**). ''Loudness, its definition, measurement and calculation,'' J. Acoust. Soc. Am. **5**, 82–108.
Florentine, M., and Buus, S. (**1981**). ''An excitation-pattern model for intensity discrimination,'' J. Acoust. Soc. Am. **70**, 1646–1654.
Florentine, M., Buus, S., and Mason, C. (**1987**). ''Level discrimination as a function of level for tones from 0.25 to 16 kHz,'' J. Acoust. Soc. Am. **81**, 1528–1541.
Hellman, R. P. (**1970**). ''Effect of noise bandwidth on the loudness of a 1000-Hz tone,'' J. Acoust. Soc. Am. **48**, 500–504.
Hellman, W. S., and Hellman, R. P. (**1990**). ''Intensity discrimination as the driving force for loudness. Application to pure tones in quiet,'' J. Acoust. Soc. Am. **87**, 1255–1265.
Hellman, R. P., and Zwislocki, J. J. (**1961**). ''Some factors affecting the estimation of loudness,'' J. Acoust. Soc. Am. **33**, 687–694.
Hellman, R. P., and Zwislocki, J. J. (**1963**). ''Monaural loudness function at 1000 cps and interaural summation,'' J. Acoust. Soc. Am. **35**, 856–865.
Houtsma, A. J. M., Durlach, N. I., and Braida, L. D. (**1980**). ''Intensity perception XI. Experimental results on the relation of intensity resolution to loudness matching,'' J. Acoust. Soc. Am. **68**, 807–813.
Jesteadt, W., Wier, C. G., and Green, D. M. (**1977**). ''Intensity discrimination as a function of frequency and level,'' J. Acoust. Soc. Am. **61**, 169–177.
Johnson, J. H., Turner, C. W., Zwislocki, J. J., and Margolis, R. H. (**1993**). ''Just noticeable differences for intensity and their relation to loudness,'' J. Acoust. Soc. Am. **93**, 983–991.
Kohlrausch, A., Fassel, R., and Dau, T. (**2000**). ''The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers,'' J. Acoust. Soc. Am. **108**, 723–734.
McGill, W. J., and Goldberg, J. P. (**1968a**). ''A study of the near-miss involving Weber's law and pure-tone intensity discrimination,'' Percept. Psychophys. **4**, 105–109.
McGill, W. J., and Goldberg, J. P. (**1968b**). ''Pure-tone intensity discrimination and energy detection,'' J. Acoust. Soc. Am. **44**, 576–581.
Miller, G. A. (**1947**). ''Sensitivity to changes in the intensity of white noise and its relation to masking and loudness,'' J. Acoust. Soc. Am. **19**, 609–619.
Rabinowitz, W. M., Lim, J. S., Braida, L. D., and Durlach, N. I. (**1976**). ''Intensity perception VI. Summary of recent data on deviations from Weber's law for 1000-Hz tone pulses,'' J. Acoust. Soc. Am. **59**, 1506–1509.
Rankovic, C. M., Viemeister, N. F., Fantini, D. A., Cheeseman, M. F., and Uchiyama, C. L. (**1988**). ''The relation between loudness and intensity difference limens for tones in quiet and in noise backgrounds,'' J. Acoust. Soc. Am. **84**, 150–155.
Riesz, R. R. (**1928**). ''Differential intensity sensitivity of the ear for pure tones,'' Phys. Rev. **31**, 867–875.
Scharf, B. (**1964**). ''Partial Masking,'' Acustica **14**, 16–23.
Scharf, B. (**1978**). ''Loudness,'' in *Handbook of Perception. Vol. IV*, edited by E. Carterette and M. P. Friedman (Academic, New York).
Schlauch, R. S., Harvey, S., and Lanthier, N. (**1995**). ''Intensity resolution and loudness in broadband noise,'' J. Acoust. Soc. Am. **98**, 1895–1902.
Schlauch, R. S., and Wier, C. C. (**1987**). ''A method for relating loudness-matching and intensity-discrimination data,'' J. Speech Hear. Res. **30**, 13–20.
Stevens, S. S. (**1975**). *Psychophysics*, edited by G. Stevens (Wiley, New York).
Turner, C. W., Zwislocki, J. J., and Filion, P. R. (**1989**). ''Intensity discrimination determined with two paradigms in normal and hearing-impaired subjects,'' J. Acoust. Soc. Am. **86**, 109–115.
Viemeister, N. F., and Bacon, S. P. (**1988**). ''Intensity discrimination, increment detection, and magnitude estimation for 1-kHz tones,'' J. Acoust. Soc. Am. **84**, 172–178.
Wojtczak, M., and Viemeister, N. F. (**1999**). ''Intensity discrimination and detection of amplitude modulation,'' J. Acoust. Soc. Am. **106**, 1917–1924.
Zwislocki, J. J., and Goodman, D. A. (**1980**). ''Absolute scaling of sensory magnitudes: A validation,'' Percept. Psychophys. **28**, 28–38.
Zwislocki, J. J., and Jordan, H. N. (**1986**). ''On the relations of intensity jnd's to loudness and neural noise,'' J. Acoust. Soc. Am. **79**, 772–780.

Infants' sensitivity to broadband noise

Lynne A. Werner and Kumiko Boike
Department of Speech and Hearing Sciences, University of Washington, 1417 NE 42nd Street, Seattle, Washington 98015-6246

(Received 12 October 1999; accepted for publication 20 February 2001)

Infants have higher pure-tone thresholds than adults. One explanation is that infants do not adopt the frequency-selective listening strategy that adults use when detecting tones. In contrast to other models of infants' immature sensitivity, the listening strategy account predicts that infants will be more sensitive to broadband sounds, relative to adults. Infants 7–9 months old were tested in two experiments to examine their sensitivity to broadband noise. Unmasked and masked thresholds for a 1000-Hz tone and for broadband noise were estimated adaptively for infants and adults using an observer-based behavioral procedure. The difference between infants and adults in unmasked threshold were 14 and 7 dB for tones and noise, respectively. The difference between infants and adults in masked threshold were 10 and 5 dB for tones and noise, respectively. Psychometric functions for detection of broadband noise were also obtained from some infants and adults. Infants' psychometric functions were similar to those obtained in tone detection with shallower slopes and lower upper asymptotes than adults'. This suggests that the relative improvement in infants' threshold for broadband noise is not due to greater attentiveness to the noise. A model of infants' sound detection invoking inattentiveness, listening strategy, and an unspecified source of internal noise may account for the characteristics of the infant psychometric function. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1365112]

PACS numbers: 43.66.Cb, 43.66.Dc [RVS]

I. INTRODUCTION

Infants' detection thresholds for tones, in quiet or in noise, are higher than those of adults (reviewed by Werner and Gray, 1998). Infants' pure-tone thresholds at 1000 Hz are about 15 dB higher than adults' in quiet and about 8 dB higher than adults' in noise. The age difference for thresholds in quiet is partially accounted for by immaturity of the middle ear (Keefe *et al.*, 1993). The remaining 8-dB age difference in sensitivity has been explained in various ways.

Inattentiveness has been offered as an explanation for many age differences in performance on psychophysical tasks, including the infant–adult detection threshold difference. Inattentiveness is commonly modeled as guessing on some proportion of trials. The guessing rate that would be required to account for infants' detection thresholds, however, is much greater than infants' responses to clearly audible tones would suggest. For example, Bargones, Werner, and Marean (1995) found that infants detected apparently audible tones about 85% of the time, consistent with a guessing rate of 30%. Analyses by several investigators have suggested that a guessing rate of about 50% would be necessary to account for the observed age difference in threshold (Allen and Wightman, 1994; Bargones *et al.*, 1995; Schneider and Trehub, 1992; Viemeister and Schlauch, 1992; Werner, 1992; Wightman and Allen, 1992). Thus, inattentiveness alone does not appear to be a sufficient explanation for the observed immaturity of infants' pure-tone detection thresholds, although it is likely to make some contribution.

Schneider, Trehub, and colleagues have argued that the most reasonable explanation for infants' immature detection thresholds is immaturity of the nonlinear ''transfer function mapping sound pressure onto neural activity,'' resulting in slower growth of neural activity with increasing sound level in the infant system than in the adult system (Schneider and Trehub, 1992; Schneider *et al.*, 1989, p. 41). Thus, a higher signal-to-noise ratio would be required to produce a given level of neural activity in the infant auditory nervous system than in the adult auditory system. Under the assumption that detection by infants and adults requires the same increase in neural activity, infants' detection thresholds would be higher. It might be expected that immaturity of the relationship between sound intensity and primary neural activity would be reflected in the characteristics of auditory evoked potentials, and there is some evidence that evoked potentials mature along a time course similar to that of masked thresholds (Durieux-Smith *et al.*, 1985; Eggermont, 1985; Fabiani *et al.*, 1979; Hall, 1992). However, there is no direct evidence linking immaturities of evoked potentials to those of sensitivity after 3 months of age.

Bargones *et al.* (1995) pointed out that the characteristics of the infant psychometric function for pure-tone detection and the observed age differences in threshold could be accounted for by immaturity of listening strategies. There is a wealth of evidence indicating that adults' listening strategy in pure-tone detection is to monitor the output of the auditory filter centered at the frequency of the expected signal (e.g., Green and Swets, 1988; Hafter and Schlauch, 1992; Scharf, 1987). Infants, however, may not monitor the output of the optimal auditory filter as adults do. In fact, using the probe-signal method, Bargones and Werner (1994) demonstrated that infants listen less selectively in the frequency domain than adults do. If infants monitor the output of an auditory filter centered at a frequency well removed from that of the

expected tone, their psychometric function would be predicted to be shallower than that of adults, as observed by Bargones *et al.* (Hübner, 1993). In addition, the expected age difference in threshold would be on the order of 7 dB (Dai, Scharf, and Buus, 1991), close to the observed age difference in masked threshold at 1000 Hz. We refer to this listening strategy as the ''wandering listening band.''

The observed age differences in pure-tone sensitivity as well as the findings of Bargones and Werner (1994) could also be explained by infants' monitoring a broad range of frequencies in pure-tone detection, rather than a narrow frequency range that is not centered on the frequency of the expected tone. Hübner (1993) has shown that such a ''broadband'' listening strategy will produce a steeper psychometric function than monitoring the output of a single auditory filter centered on the frequency of the expected tone. Because the psychometric functions of infants and children are typically shallower than those of adults, the broadband listening strategy has not been discussed in any detail as a potential explanation for age-related differences in sensitivity (e.g., Allen and Wightman, 1994; Bargones *et al.*, 1995). If infants are broadband listeners, then additional factors must be invoked to account for the slope and upper asymptote of their psychometric function. We return to this point at the end of the paper.

These alternative explanations for infants' detection of frequency-specific stimuli make different predictions concerning infants' detection of broadband stimuli. The growth-of-neural-response explanation predicts that infants' detection of broadband noise, for example, would be similar to their detection of pure tones with about the same threshold relative to adults. The listening strategy account predicts that infants will be relatively better at detecting broadband noise, because broadband noise contains energy at all frequencies, making selective listening more or less irrelevant.[1] If infants listen in a more mature, broadband way when detecting broadband sounds, a threshold improvement on the order of 7 dB would be predicted relative to pure-tone thresholds (Dai *et al.*, 1991). A change in inattentiveness could lead to relatively better infant thresholds for broadband noise, but only if infants are more attentive to noise than to tones. If infants are perfectly attentive to noise, threshold for broadband noise would improve at most by 2 or 3 dB relative to pure-tone thresholds (see analyses by Viemeister and Schlauch, 1992; Schneider and Trehub, 1992; Werner, 1992; and Wightman and Allen, 1992).

Several investigators have noted that infants appear to be more responsive to broadband than to narrow-band sounds (e.g., Eisenberg, 1976; Gerber and Mencher, 1979; Turkewitz, Birch, and Cooper, 1972). Different investigators have published thresholds for sounds of different bandwidths, but no consistent interactions between age and bandwidth are evident in cross-study threshold comparisons. Berg and her colleagues examined the effects of bandwidth on infants' detection threshold (Berg, 1991; Berg and Boswell, 1995). Berg obtained unmasked thresholds for 1/3-octave filtered clicks and 300-ms noise bursts centered at 4000 Hz, and for unfiltered clicks and 300-ms broadband noise bursts to examine the effects of bandwidth on temporal summation in infants' detection in quiet. The difference between infants' and adults' thresholds was smaller for the broadband noise burst than for the 1/3-octave filtered noise burst, but the opposite appeared to be true for the clicks. Berg's thresholds for ''long-duration'' 4000-Hz tones and 4000-Hz, octave-band noise detected in quiet and in noise are consistent with infants' being more sensitive to the broader bandwidth sound, but the tone and noise band were also of different durations. Berg and Boswell conducted a similar study, but with 500-Hz tones and octave-band noises. In this case, no interaction between bandwidth and age effects on sensitivity was observed. In sum, there is no compelling evidence that infants' sensitivity is dependent on bandwidth, but the issue has not been directly addressed.

The purpose of the present study was to compare infants' and adults' detection of a 1000-Hz tone and a broadband noise under unmasked and masked conditions. We chose 1000 Hz, because we have previously studied infants' detection at that frequency extensively (e.g., Bargones and Werner, 1994; Bargones *et al.*, 1995; Werner and Bargones, 1991; Werner and Marean, 1991). In experiment 1, thresholds for unmasked and masked tones and noise were estimated adaptively for the two age groups. Experiment 2 was a preliminary examination of the infant psychometric function for detection of unmasked noise.

II. EXPERIMENT 1: THRESHOLDS FOR MASKED AND UNMASKED TONES AND BROADBAND NOISE

A. Method

1. Subjects

Seventy-three 7–9-month-old infants and 40 18–30-year-old adults provided thresholds. The average age of the infants was 33.0 weeks (s.d.=4.8 weeks), and the average age of the adults was 22.1 years (s.d.=2.8 years) at the first test session. All subjects were free of risk factors for hearing loss, assessed by parental- or self-report, had had no more than two episodes of otitis media, had not been under treatment for otitis media within the prior week, and were healthy on the day of testing. All subjects passed screening tympanometry at each test session, with a peak admittance of at least 0.2 mmhos at a pressure between −200 and 50 daPa. No subject had participated in other psychoacoustical experiments or had more than 2 years of musical training. The data of 64 additional infants who participated in the study were excluded from analysis; 21 infants did not reach training criterion, 8 did not provide enough data to estimate a threshold, 35 completed testing but were excluded because of high false-alarm rate, low probe trial response rate, or high variability of reversals in the adaptive run. The data of 3 adults were excluded, 1 due to equipment failure, 1 due to failure to meet training criteria, and 1 due to highly variable reversals.

2. Stimuli and procedure

The stimuli were a 1000-Hz tone and a broadband noise. The tone was digitally produced, attenuated, filtered, and amplified. The noise was produced by a noise generator, attenuated, low-pass filtered (48 dB/oct) at 6000 Hz, and amplified. Signal duration was 500 ms, with 16 ms rise–fall. When

used, the same noise was presented continuously as a masker at a level of 20 dB N_0. All stimuli were presented via an Etymotic ER1 insert earphone in a foam tip, trimmed to fit the ear canal as necessary. Stimuli were calibrated in a Zwislocki coupler. Each subject was tested in one of four detection tasks: unmasked tone detection, unmasked noise detection, masked tone detection, or masked noise detection. In the masked noise detection task, the signal was produced by increasing the level of the background noise for 500 ms. The stimuli and experiment were controlled by computer.

Infants were tested using an observer-based procedure (Werner, 1995). This is a go/no-go procedure in which an observer decides on each trial whether a signal was presented to the infant based solely on the infants' behavior. The infant was tested while sitting on the parent's lap in the test booth. An assistant in the test booth kept the infant entertained and attentive by manipulating silent toys on a table in front of the infant. The parent and the assistant wore headsets and listened to masking sounds to make certain that they could not hear any of the sounds presented to the infant. The observer was seated outside the booth and watched through a window or on a video monitor. When the infant was quiet and alert, the observer began a trial. Signal and no-signal trials were presented with equal probability. Four tone or noise bursts occurred, at 1-s intervals, on signal trials. On no-signal trials, either no sound was presented or the background noise continued unchanged for 4 s. The observer's response had to be recorded during the trial. The observer had no prior knowledge of trial type, but received feedback after each trial. The infant was reinforced for responding on signal trials by the activation of a mechanical toy.

Each session began with a training phase, in which the signal was presented at levels expected to be clearly audible to the infants. The infant/observer team was required to achieve 80% correct on both signal and no-signal trials to meet training criterion. The number of trials to criterion averaged 20.3 (s.d.=11.2) and did not vary systematically with condition. Thresholds were then determined adaptively using a one-up, two-down algorithm. Tone level started at 45 dB SPL in quiet and at 65 dB SPL in noise. Noise burst spectrum level started at 5 dB SPL in quiet and at 35 dB SPL in noise. Initial step size was 10 dB; step size varied during a run according to PEST rules (Taylor and Creelman, 1967). Only signal trials affected the direction of the adaptive track. Testing continued until at least eight reversals were obtained; the average of the last six reversals was taken as threshold. Thresholds were only accepted if the false-alarm rate during the run was less than 0.4. Infant false-alarm rate averaged 0.25. In every six trials, a ''probe'' trial was included at a randomly determined position. Probe trials were signal trials with the signal level set at the training level. The outcome of probe trials did not affect the adaptive track, but the thresholds were only accepted if the response rate on probe trials was greater than 0.6. Infant probe response rate averaged 0.94. Infant thresholds were based on an average of 40.9 trials (s.d.=8.4) with no differences across conditions.

Adults were tested in as similar a way to the infants as possible. Adults listened alone in the test booth and were instructed to respond when they heard a sound that would make the mechanical toy come on. They were told that their goal was to make the toy come on as much as possible. Training and testing procedures and criteria were the same as those used for the infants. Adults reached training criteria in a minimum number of trials (M=10.0, s.d.=2.6), had a mean false-alarm rate of 0.01 (s.d.=0.05), a universal probe response rate of 1.0, and thresholds based on an average of 44.1 trials (s.d.=7.7).

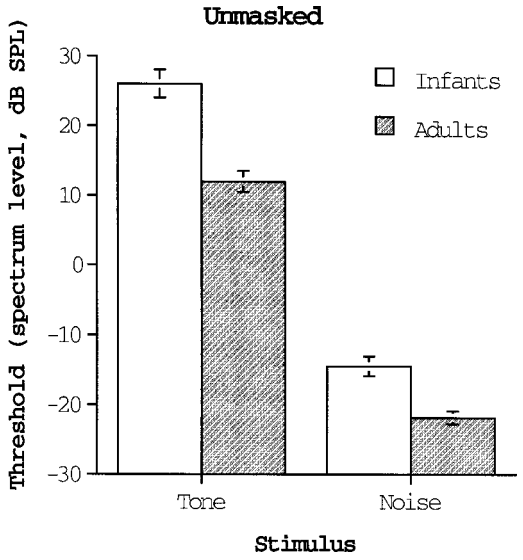


FIG. 1. Average (±1 standard error) unmasked thresholds for a 1000-Hz tone and a broadband noise, for infants and adults, from experiment 1.

The number of sessions required to obtain a threshold was 2–4 for infants (mode=2) and 1 for adults. Infants who reached training criterion but did not produce a threshold in a session, completed only a ''reminder'' phase, 2–3 trials at the original training level, prior to testing in subsequent sessions. All visits were completed within a 2-week period.

B. Results

All thresholds were calculated as the spectrum level of the signal in dB SPL. The mean detection thresholds for unmasked tones and noise are plotted in Fig. 1. Infant thresholds for both tones and noise were higher than those of adults, but it appears that infants are relatively better at detecting noise than they are at detecting tones in a quiet background. Because there were apparent differences in variability across experimental conditions, nonparametric tests of statistical significance are reported, although parametric analysis produced the same results. The Mann-Whitney U test, corrected for ties, indicated that the difference in threshold between infants and adults was highly significant both for tones ($p<0.001$) and for noise ($p<0.001$). More important for the present question, however, is whether the size of the infant–adult difference depends on the stimulus. The adjusted ranks test (Sawilowsky, 1990) was used to test the age×stimulus interaction. The interaction was significant ($p<0.05$). Thus, the infant–adult threshold difference for tones, about 14 dB, is significantly greater than the infant–adult threshold difference for noise, about 7 dB, when detection is tested in quiet.

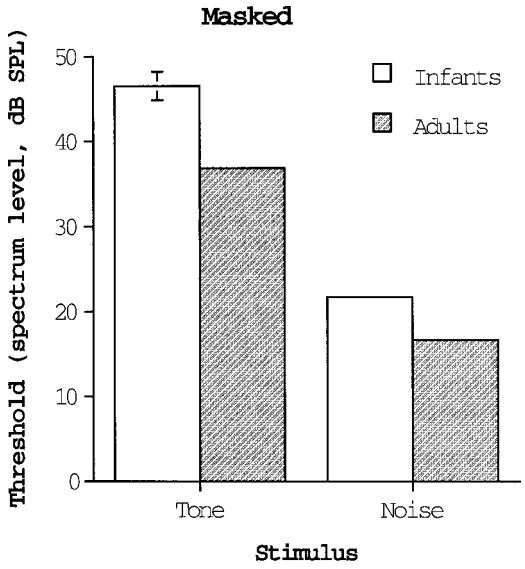


FIG. 2. Average (±1 standard error) masked thresholds for a 1000-Hz tone and a broadband noise, for infants and adults, from experiment 1. The masker was a broadband noise. Some of the error bars are very small; the standard deviation in those conditions was 2–3 dB.

The mean detection thresholds for masked tones and noise are plotted in Fig. 2. Again, infants' thresholds appear to be higher than those of adults in both stimulus conditions, and again, it appears that infants are relatively better at detecting noise than they are at detecting tones. The statistical analyses confirmed that the difference between infants' and adults' thresholds was highly significant in both stimulus conditions ($p<0.001$ for tones and for noise), but the adjusted ranks analysis also indicated a significant age×stimulus interaction ($p<0.01$). Thus, in a noise background, infants' thresholds for detecting noise are relatively more mature than their thresholds for detecting tones. The infant–adult threshold difference for tones was about 10 dB, while that for noise was about 5 dB.

C. Discussion

The 1000-Hz thresholds reported here are comparable to those previously reported for both infants and adults. The thresholds for an unmasked 1000-Hz tone in quiet are within 5 dB of those reported by Bargones *et al.* (1995) for adults and for infants of the same age, tested using the same transducer, and by Nozza and Wilson (1984) for 6-month-old infants and for adults, tested using a TDH-49 earphone calibrated in a 6-cc coupler. Perhaps more importantly, the age differences in threshold at 1000 Hz reported in these studies were 16 dB (Bargones *et al.*, 1995) and 14 dB (Nozza and Wilson, 1984), while that observed in the present study was 15 dB. Similarly, signal-to-noise ratio (S_0/N_0) at masked threshold was within 5 dB of that reported for 1000-Hz tones in the Bargones *et al.* and Nozza and Wilson studies, as well as to those reported by several investigators for tones of other frequencies (Berg, 1991; Berg and Boswell, 1995; Nozza and Wilson, 1984). The age difference in masked threshold reported by Bargones *et al.* and by Nozza and Wilson, 8 dB, is close to the 9.5-dB difference found here. Berg reports an age difference of 12 dB at 4000 Hz and Berg and Boswell report an age difference of 10 dB at 4000 Hz.

The thresholds for detecting a masked noise in this study are more or less consistent with previous reports. For adults, the threshold here was about −4 dB S_0/N_0. Green and Sewall (1962) report values for trained observers tested in a two alternative, forced-choice (2AFC) paradigm of −10 to −8 dB. Berg and Boswell's (1998) thresholds for octave-band noise increments for adults tested in an infant test procedure are equivalent to −5 or −6 dB S_0/N_0. For infants, estimates of S_0/N_0 at threshold for broadband sounds range from −2 to 4 dB (Kopyar, 1997; Bull, Eilers, and Oller, 1984; Schneider, Bull, and Trehub, 1988). The current estimate of −0.43 dB falls within this range. Thresholds for a broadband noise in quiet have not been previously estimated for infants,[2] and in fact have not often been reported for adults. Yost and Klein (1979), however, reported a threshold of 5 dB SPL for broadband noise, low-pass filtered at 10 000 Hz and Raab, Osman, and Rich (1963) reported a threshold of 4 dB SPL for broadband noise, low-pass filtered at 7000 Hz for adult subjects. The threshold for detection of noise bursts in quiet here, low-pass filtered at 6000 Hz, was about 16 dB SPL overall level, quite a bit higher than the published reports. Whether this difference results from training, the difference in transducers, the difference in test method, or some other variable is not clear.

Experiment 1 demonstrates that infant's thresholds are closer to those of adults when they are detecting a broadband noise than when they are detecting a pure tone, for both masked and unmasked signals. Models that account for threshold immaturity in terms of primary sensory immaturity, such as reduced growth of neural response with increasing intensity, do not predict greater sensitivity to broadband sound in infants. The present finding is consistent with the idea that infants do not listen in a frequency selective way, or at least that they are not listening at the optimal frequency, when they are detecting a pure tone. A model of threshold immaturity based on inattention also predicts better infant sensitivity to broadband sounds, if one assumes that infants are inattentive less often to broadband than to narrow-band sounds.

The characteristics of the psychometric function for broadband detection could help to distinguish between listening strategy and inattention explanations of the improvement in infants' threshold for broadband noise relative to pure tones. The slope of adults' psychometric function for detection has generally been reported to be independent of bandwidth (Buss *et al.*, 1986; Green and Sewall, 1962; but see de Boer, 1966; Scholl, 1961). If infants are more attentive to broadband noise than to tones, but they listen as adults do, the prediction is that the psychometric function will be steeper than that for detecting tones and have an upper asymptote closer to 1. The inattention model also predicts that correcting or ''rescaling'' the infant psychometric function, using a guessing rate based on the upper asymptote, would make the slope and threshold for tone and noise equivalent, relative to those of adults. An upper asymptote of the infant psychometric function less than or equal to that for tones

would argue that infants are not more attentive to noise than to tones.

Either a wandering listening band or a broadband listening strategy hypothesis would predict an adultlike psychometric function slope and asymptote for detection of broadband noise. In the case of pure-tone detection, if infants listen at a frequency removed from the signal frequency, excitation increases as a function of both the intensity of the tone and the frequency monitored by the infant, presumably a random variable. If infants listen to a broad range of frequencies in pure-tone detection, excitation increases as a function of both the intensity of the tone and the bandwidth monitored by the infant. In the case of broadband noise detection, excitation increases as a function of intensity in the same way at all frequencies. If infants do not match their listening bandwidth to the noise bandwidth, their threshold could still be somewhat elevated relative to adults', but the psychometric function slope in this case should be adultlike.

III. EXPERIMENT 2: PSYCHOMETRIC FUNCTIONS FOR DETECTION OF UNMASKED BROADBAND NOISE

A. Method

1. Subjects

Ten infants and 11 adults, meeting the same inclusion criteria as in experiment 1, provided psychometric function data. At the first test session, the average age of the infants was 32.6 weeks (s.d.=1.7 weeks), and the average age of the adults was 22.8 years (s.d.=3.5 years). Four additional infants participated in the study but did not provide enough data to fit a psychometric function. Eight additional infants and four additional adults produced psychometric functions with slope not significantly greater than zero; these functions were excluded from analysis.[3] Ten infants were tested at just one suprathreshold level to estimate the upper asymptote of the psychometric function. The average age of these infants was 34.6 weeks (s.d.=2.6). Three additional infants did not complete testing in this condition.

2. Stimuli and procedure

The stimulus was the same noise used in experiment 1, low-pass filtered at 6000 Hz, and presented in bursts of 500-ms duration.

The same training procedures were used as in experiment 1. The average number of training trials required to reach criterion was 26.9 (s.d.=8.7) for infants and 13.1 (s.d.=4.2) for adults. Once the subject had reached training criteria, a block of 35 test trials was completed in which 15 trials were no-signal trials, 15 trials were signal trials with signal level fixed, and 5 were probe trials with the signal level at the initial training level. Once the infant had completed the training procedure, only a few reminder trials were presented prior to testing in subsequent visits. Infants who provided psychometric function data were tested at signal spectrum levels of −15 and −10 dB SPL, counterbalanced for order across subjects, in the first two visits to the lab. Depending on performance in these two conditions, the infant was tested at −20 or −5 dB SPL spectrum level, and then at levels 5 dB higher or lower, in order to place as many data points as possible along the slope of the psychometric function. If performance was worse at −10 dB SPL than at −15 dB SPL, −10 dB SPL was retested. At least five sessions were attempted for each infant. A similar procedure was used for adults, except that the initial spectrum levels were −30 and −25 dB SPL. A test block was accepted only if the response rate on probe trials was greater than 0.6. The average response rate on probe trials was 0.8 for infants (s.d.=0.24) and all adults responded on all probe trials. In the final sample, four infants and four adults provided four data points, seven infants and five adults provided three data points, and two infants and two adults provided two data points.

The infants who provided upper asymptote data were tested only at 5-dB SPL spectrum level, with no probe trials. The choice of level was based on the psychometric functions. Although not all psychometric functions had a well-defined upper asymptote, the highest level at which a psychometric function was estimated to reach asymptotic performance was −1.5 dB and most were around −7 dB. A single block of test trials was obtained in one or two visits. The average number of trials required to reach training criterion was 33.4 (s.d.=12.6) for these infants.

Hit rates of 1 were converted to $1-(1/2N)$ or 0.97 and false-alarm rates of 0 were converted to $(1/2N)$ or 0.03 (Macmillan and Creelman, 1991). Performance was described as $p(C)_{\max}$. Psychometric functions were fit to each subject's data using the procedure of Bargones *et al.* (1995). Three linear segments were fit to each subject's data using piecewise regression: a lower asymptote with slope=0 and *y* intercept=0.5; an upper asymptote with slope=0 and *y* intercept equal to the average of all data points within the 95% confidence interval of the best $p(C)_{\max}$ achieved in a block; and a segment defined between the levels at which asymptotic performance was reached with slope and intercept the free parameters. The slope of the psychometric function was taken as the slope of the middle segment.

B. Results

1. Psychometric function slope

Examples of data and psychometric functions fits for two infants and two adults are shown in Fig. 3. In general the fits to the data were good: r^2 ranged from 0.67 to 0.99 for the functions based on 3 or more data points, averaging about 0.9 for both infants and adults.

Average psychometric function slope for detection of broadband noise for infants and adults are shown in Fig. 4. Infants' psychometric function slope was significantly shallower than that of adults ($p<0.05$). The observed psychometric function slope for infants detecting noise in this study was $0.03p(C)_{\max}$/dB, while that reported for infants detecting tones by Bargones *et al.* (1995) was $0.04p(C)_{\max}$/dB. The psychometric function slope for adults detecting noise was $0.06p(C)_{\max}$/dB, compared to $0.05p(C)_{\max}$/dB reported by Bargones *et al.* for adults detecting a tone. Statistical comparison of the results of the present study to those of Bargones *et al.* indicated that the difference between the

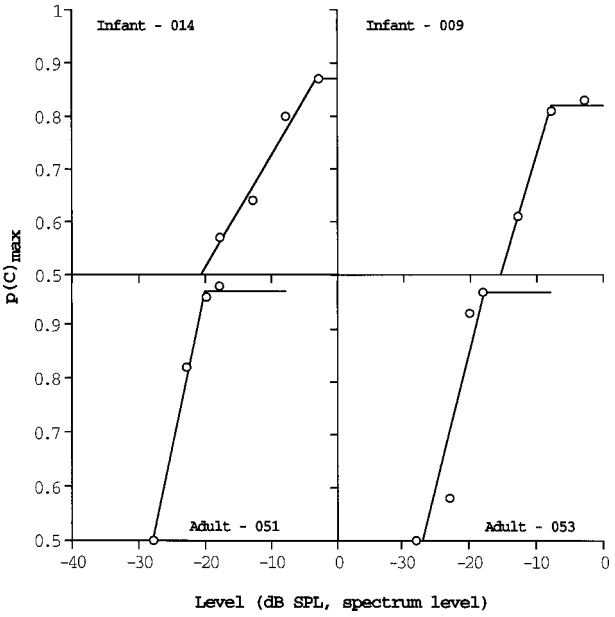


FIG. 3. Examples of psychometric function data (circles) and best-fitting psychometric functions (lines) from experiment 2. The task was detection of broadband noise bursts.

slopes for noise detection and tone detection was not significant for infants ($p>0.05$) or adults ($p>0.10$). The slopes obtained here are within the range typically reported for adults (e.g., Buus *et al.*, 1986).

2. Upper asymptote

The average $p(C)_{max}$ attained by the infants who were tested at a spectrum level of 5 dB SPL was 0.83 (s.d. $=0.06$). This value is very close to that reported by Bargones *et al.* (1995) for infants detecting a tone under various conditions, about 0.85. The infants who provided psychometric function slopes had an average maximum performance of $p(C)_{max}=0.81$ (s.d.$=0.09$); this value was not significantly different from the upper asymptote estimated for the other group of infants ($p>0.05$). Thus, on average, the infants in the psychometric function study achieved asymptotic performance within the range of levels tested. This level of performance, however, was significantly poorer than that attained by adults, $p(C)_{max}=0.96$ (s.d.$=0.01$) ($p<0.001$).

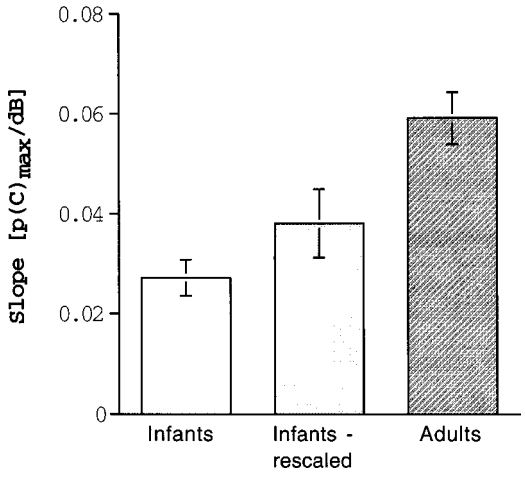


FIG. 4. Average (±1 standard error) slope of the psychometric functions for detection of broadband noise for infants and adults from experiment 2. ''Infants—rescaled'' refers to the average slope of the infants' function after correction for inattentiveness.

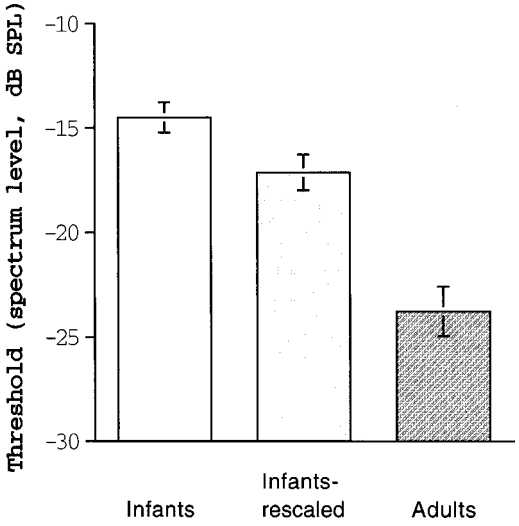


FIG. 5. Average (±1 standard error) unmasked thresholds for a broadband noise, for infants and adults, from experiment 2. ''Infants—rescaled'' refers to the average threshold calculated after the infants' psychometric function had been corrected for inattentiveness.

As the slope and upper asymptote of infants' psychometric function for noise and tone detection are not different, rescaling the psychometric function to correct for inattention would not affect the slope or threshold for noise relative to those for a tone. However, the comparison between the rescaled infant and adult psychometric functions for noise detection may have some relevance for understanding the factors underlying the age differences in slope, upper asymptote, and threshold. Consequently, for each infant it was assumed that the highest $p(C)_{max}$ attained represented the upper asymptote (UA) of the psychometric function. Rescaled $p(C)_{max}$ was defined as $[p(C)_{max}-0.5]/[\mathrm{UA}-0.5]+0.5$. A new psychometric function was then fit to the rescaled data points using the original procedure. Because the adults clearly approached an upper asymptote of 1, their psychometric functions were not rescaled. The average slope of the rescaled infant psychometric function is plotted in Fig. 4 along with the original infant and adult slopes. Although the rescaled infant slopes were steeper than the original infant slopes ($p<0.05$), they were still shallower than adult slopes ($p<0.05$).

3. Threshold

Thresholds were also calculated for each subject who provided psychometric function data. Threshold was defined as the level at which $p(C)_{max}=0.75$. Average thresholds for infants and adults are plotted in Fig. 5. The threshold values estimated from the psychometric functions were very similar to those obtained using the adaptive method in experiment 1 (compare to Fig. 1). Infants' thresholds were, not surprisingly, significantly poorer than those of adults ($p<0.001$),

and the infant–adult difference remained, at 9 dB, smaller than that observed for a 1000-Hz tone in experiment 1 and by Bargones *et al.* (1995).

Thresholds were also calculated for the infants' rescaled psychometric functions. The average of those thresholds is also plotted in Fig. 5. The ''rescaled thresholds'' were about 3 dB lower than the original thresholds estimated for the same infants; this improvement was statistically significant ($p<0.05$). The rescaled thresholds were still significantly poorer than adults' thresholds ($p<0.01$).

C. Discussion

Neither the slope nor the upper asymptote of infants' psychometric function for the detection of broadband noise is different from that for the detection of tones (experiment 2), but infants' thresholds for detecting broadband noise are more adultlike than are their thresholds for detecting tones (experiments 1 and 2). This pattern of results is strong evidence against the inattentiveness explanation for infants' relatively better detection of broadband noise over tones. However, it is not readily explained solely by a listening strategy explanation of infants' pure-tone detection.

If infants' threshold for broadband noise is relatively better than their threshold for pure tones because they are more attentive to broadband noise, then the slope and upper asymptote of the psychometric function for noise detection should be greater than those observed for pure-tone detection. In addition, thresholds calculated from rescaled psychometric functions for tone and for noise detection, using a guessing rate based on the respective upper asymptotes, should be equivalent, relative to adults' threshold. In fact, neither the slope nor the upper asymptote of the infants' psychometric function for noise detection is different from that for tone detection. Thus, there is no evidence that infants are more attentive to noise, and infants' threshold for noise detection, whether corrected for inattention or not, is relatively better than their threshold for tone detection.

The observed results, however, cannot be explained entirely by a model of infants' relative insensitivity to tones based on their listening strategy. The broadband listening model predicts that infants' psychometric function for pure-tone detection will be steeper than that of adults with an upper asymptote of 1. That prediction is inconsistent with published observations (e.g., Bargones *et al.*, 1995). The model based on a wandering listening band predicts that the slope of the infants' psychometric function for pure-tone detection will be shallow and its upper asymptote will only slowly approach 1. That prediction is consistent with published observations (e.g., Bargones *et al.*, 1995). However, both models predict that the slope and upper asymptote of the psychometric function for broadband noise detection should be adultlike even if infants do not match their listening bandwidth to the bandwidth of the signal. Neither model is consistent with the current observation that infants' psychometric function for broadband noise detection is no more adultlike than that for pure-tone detection.

In the interest of parsimony, we began by trying to understand infants' auditory sensitivity in terms of a single immaturity. A realistic model of infants' sound detection apparently must recognize multiple immaturities. The difference between infants' and adults' sound detection could be modeled in terms of conductive efficiency, frequency resolution, inattentiveness, and an internal noise of unspecified origin, as follows:

(1) The difference between masked and unmasked thresholds provides an estimate of the contribution of conductive efficiency to the infant–adult difference in unmasked threshold. Based on the current data, we would estimate the conductive component of infants' absolute threshold immaturity at 3–4 dB. That estimate is consistent with the findings of Keefe *et al.* (1993) with respect to the development of ear-canal conductance levels with an insert earphone.

(2) Because frequency resolution is mature by 6 months of age (Hall and Grose, 1991; Olsho, 1985; Spetner and Olsho, 1990), it should be irrelevant to the age difference in masked threshold.

(3) If the reduced upper asymptote of the infant psychometric function for detection is largely due to inattentiveness, then about 2 dB of the age difference in thresholds and most of the age difference in psychometric function slope for both tones and noise could be accounted for by inattentiveness.

(4) The infants' masked thresholds, relative to adults', were elevated by 5 dB for a broadband noise and 10 dB for a 1000-Hz tone. The corresponding differences for unmasked thresholds were 7 and 14 dB, respectively. Given the observations of Bargones and Werner (1994) that infants do not demonstrate frequency selective listening, this 5–7-dB increase in threshold for the tone could still be consistent with a broadband listening strategy in pure-tone detection. If infants simultaneously monitor the outputs of multiple auditory filters when detecting a pure tone, their pure-tone threshold would still be worse than their broadband noise threshold, but their psychometric function for pure-tone detection would be steeper than that for broadband noise detection (Hübner, 1993). Although not statistically significant, the difference between psychometric function slopes observed here for broadband noise and those reported by Bargones *et al.* (1995) for a 1000-Hz tone is in the direction predicted by such a broadband listening model. Further, Bargones *et al.* reported that infants' rescaled psychometric function slopes in pure-tone detection were consistently steeper than those of adults. Thus, most of the age difference in psychometric function slope could be accounted for by broadband listening combined with inattentiveness. The broadband listening model could be tested directly, for example, by estimating the weights infants assign to different spectral regions when detecting a tone in noise as in Stellmack *et al.* (1997) and Willihnganz *et al.* (1997).

(5) Once inattentiveness is taken into account, a 3-dB age difference in threshold for masked broadband noise detection remains to be explained by age differences in the level of internal noise. The possible sources of the increased level of internal noise include a failure to match

listening bandwidth to stimulus bandwidth, variability in auditory intensity coding, higher levels of spontaneous activity in the auditory system, higher levels of physiological noise, or noise in the measurement of threshold by an adult observer. Besides increasing threshold, any of these factors could account for the difference between infants and adults in psychometric function slope that remains after rescaling the infant function.

The exact value of the relative level of internal noise, of course, depends on our estimates of all the other components. An independent estimate of the level of internal noise in infant detection can be obtained from an analysis by Schneider and Trehub (1992), who calculated the ratio of mechanical advantage to internal noise as a function of age and frequency, based on thresholds for octave-band noise in sound field. If infants' relative mechanical advantage is taken as the relative diffuse-field absorption cross-sectional area level reported by Keefe *et al.* (1994), the infants' relative level of internal noise estimated from Schneider and Trehub's data would be 5–6 dB higher than adults' between 1000 and 4000 Hz. Schneider and Trehub's model does not include an influence of inattentiveness. Their estimate of relative internal noise level is about the same as the combined effects of inattentiveness and internal noise in our model.

IV. SUMMARY AND CONCLUSIONS

Infants' thresholds for broadband noise detection are closer to those of adults than are their thresholds for tone detection. This observation is not consistent with the hypothesis that infant–adult differences in threshold are due to immature nonlinear growth of neural response with intensity in the infant auditory system. A preliminary examination of infants' psychometric function for detecting broadband noise does not support the idea that infants are less inattentive to broadband noise than to tones. Rather, like adults tested in the current procedure, infants produce similar psychometric functions for broadband noise and tone detection. Inattention combined with a broadband listening strategy and an additional source of internal noise may account for the observed pattern of results.

ACKNOWLEDGMENTS

Part of this study is based on the second author's Undergraduate Honors Thesis, Dept. of Speech and Hearing Sciences, University of Washington. Julianne Siebens and Nicole Holmer collected some of the data described here. The authors thank Ed Burns for his comments on an earlier draft of the paper.

[1]If infants were able to listen within a flat spectrum broadband noise at frequencies to which they are more sensitive, their detection thresholds for noise might also be better than for tones. Schneider *et al.* (1989), however, report that masked detection threshold matures at the same rate at all frequencies, and the data of Keefe *et al.* (1993) suggest that infant–adult differences in ear-canal impedance at 7 to 9 months of age are as small at 1000 Hz as at any other frequency. Thus, attempting to listen to more audible frequencies is unlikely to produce an improvement in either masked or unmasked threshold for a broadband noise over a 1000-Hz tone for infants of that age.

[2]Nozza (1995) reported that the minimum level of a broadband noise that would mask a tone was 8 dB higher for infants than for adults. That value is consistent with the 7-dB age difference in unmasked noise threshold reported here,

[3]No attempt was made to retest adults who produced nonmonotonic psychometric functions. Some of these relatively untrained listeners produced inconsistent results.

Allen, P., and Wightman, F. (**1994**). ''Psychometric functions for children's detection of tones in noise,'' J. Speech Hear. Res. **37**, 205–215.

Bargones, J. Y., and Werner, L. A. (**1994**). ''Adults listen selectively; infants do not,'' Psychol. Sci. **5**, 170–174.

Bargones, J. Y., Werner, L. A., and Marean, G. C. (**1995**). ''Infant psychometric functions for detection: Mechanisms of immature sensitivity,'' J. Acoust. Soc. Am. **98**, 99–111.

Berg, K. M. (**1991**). ''Auditory temporal summation in infants and adults: Effects of stimulus bandwidth and masking noise,'' Percept. Psychophys. **50**, 314–320.

Berg, K. M., and Boswell, A. E. (**1995**). ''Temporal summation of 500-Hz tones and octave-band noise bursts in infants and adults,'' Percept. Psychophys. **57**, 183–190.

Berg, K. M., and Boswell, A. E. (**1998**). ''Infants' detection of increments in low- and high-frequency noise,'' Percept. Psychophys. **60**, 1044–1051.

Bull, D., Eilers, R. E., and Oller, D. K. (**1984**). ''Infants' discrimination of intensity variation in multisyllabic stimuli,'' J. Acoust. Soc. Am. **76**, 13–17.

Buus, S., Schorer, E., Florentine, M., and Zwicker, E. (**1986**). ''Decision rules in detection of simple and complex tones,'' J. Acoust. Soc. Am. **80**, 1646–1657.

Dai, H., Scharf, B., and Buus, S. (**1991**). ''Effective attenuation of signals in noise under focused attention,'' J. Acoust. Soc. Am. **89**, 2837–2842.

de Boer, E. (**1966**). ''Intensity discrimination of fluctuating signals,'' J. Acoust. Soc. Am. **40**, 552–560.

Durieux-Smith, A., Edwards, C. G., Picton, T. W., and McMurray, B. (**1985**). ''Auditory brainstem responses to clicks in neonates,'' J. Otolaryngol. **14**, 12–18.

Eggermont, J. J. (**1985**). ''Physiology of the developing auditory system,'' in *Auditory Development in Infancy*, edited by S. E. Trehub and B. A. Schneider (Plenum, New York), pp. 21–46.

Eisenberg, R. (**1976**). *Auditory Competence in Early Life* (University Park Press, Baltimore).

Fabiani, M., Sohmer, H., Tait, C., Gafni, M., and Kinarti, R. (**1979**). ''A functional measure of brain activity: Brain stem transmission time,'' Electroencephalogr. Clin. Neurophysiol. **47**, 483–491.

Gerber, S. E., and Mencher, G. T. (**1979**). ''Arousal responses of neonates to wide band and narrow band noise,'' Paper presented at the Annual Convention of the American Speech-Language-Hearing Association, Atlanta, GA, November.

Green, D., and Sewall, S. (**1962**). ''Effects of background noise on auditory detection of noise bursts,'' J. Acoust. Soc. Am. **34**, 1207–1216.

Green, D. M., and Swets, J. A. (**1966/1988**). *Signal Detection Theory and Psychophysics* (Peninsula, Los Altos, CA).

Hafter, E. R., and Schlauch, R. S. (**1992**). ''Cognitive factors and selection of auditory listening bands,'' in *Noise-induced Hearing Loss*, edited by A. L. Dancer, R. J. Salvi, and R. P. Hamernik (Mosby Year Book, St. Louis), pp. 303–310.

Hall, J. W. I. (**1992**). *Handbook of Auditory Evoked Responses* (Allyn and Bacon, Boston).

Hall III, J. W., and Grose, J. H. (**1991**). ''Notched-noise measures of frequency selectivity in adults and children using fixed-masker-level and fixed-signal-level presentation,'' J. Speech Hear. Res. **34**, 651–660.

Hübner, R. (**1993**). ''On possible models of attention in signal detection,'' J. Math. Psychol. **37**, 266–281.

Keefe, D. H., Bulen, J. C., Arehart, K. H., and Burns, E. M. (**1993**). ''Ear-canal impedance and reflection coefficient in human infants and adults,'' J. Acoust. Soc. Am. **94**, 2617–2638.

Keefe, D. H., Burns, E. M., Bulen, J. C., and Campbell, S. L. (**1994**). ''Pressure transfer function from the diffuse field to the human infant ear canal,'' J. Acoust. Soc. Am. **95**, 355–371.

Kopyar, B. A. (**1997**). ''Intensity discrimination abilities of infants and adults: Implications for underlying processes,'' Unpublished doctoral dissertation, University of Washington.

Macmillan, N. A., and Creelman, C. D. (**1991**). *Detection Theory: A User's Guide* (Cambridge University Press, New York).
Nozza, R. J., and Wilson, W. R. (**1984**). ''Masked and unmasked pure-tone thresholds of infants and adults: Development of auditory frequency selectivity and sensitivity,'' J. Speech Hear. Res. **27**, 613–622.
Nozza, R. J. (**1995**). ''Estimating the contribution of non-sensory factors to infant–adult differences in behavioral threshold,'' Hear. Res. **91**, 72–78.
Olsho, L. W. (**1985**). ''Infant auditory perception: Tonal masking,'' Infant Beh. Dev. **7**, 27–35.
Raab, D. H., Osman, E., and Rich, E. (**1963**). ''Effects of waveform correlation and signal duration on detection of noise bursts in continuous noise,'' J. Acoust. Soc. Am. **35**, 1942–1946.
Sawilowsky, S. S. (**1990**). ''Nonparametric tests of interaction in experimental design,'' Rev. Educ. Res. **60**, 91–126.
Scharf, B. (**1987**). ''Focused auditory attention and frequency selectivity,'' Percept. Psychophys. **42**, 215–223.
Schneider, B. A., Bull, D., and Trehub, S. E. (**1988**). ''Binaural unmasking in infants,'' J. Acoust. Soc. Am. **83**, 1124–1132.
Schneider, B. A., and Trehub, S. E. (**1992**). ''Sources of developmental change in auditory sensitivity,'' in *Developmental Psychoacoustics*, edited by L. A. Werner and E. W. Rubel (American Psychological Association, Washington, DC), pp. 3–4.
Schneider, R. A., Trehub, S. E., Morrongiello, B. A., and Thorpe, L. A. (**1989**). ''Developmental changes in masked thresholds,'' J. Acoust. Soc. Am. **86**, 1733–1742.
Scholl, H. (**1961**). ''Über die Bildung der Hörschwellen und Mithöschwellen von Dauerschallen,'' Frequenz **15**, 580–564.
Spetner, N. B., and Olsho, L. W. (**1990**). ''Auditory frequency resolution in human infancy,'' Child Dev. **61**, 632–652.
Stellmack, M. A., Willihnganz, M. S., Wightman, F. L., and Lutfi, R. A. (**1997**). ''Spectral weights in level discrimination by preschool children: analytic listening conditions,'' J. Acoust. Soc. Am. **101**, 2811–2821.
Taylor, M. M., and Creelman, C. D. (**1967**). ''PEST: Efficient estimates on probability functions,'' J. Acoust. Soc. Am. **41**, 782–787.
Turkewitz, G., Birch, H. G., and Cooper, K. K. (**1972**). ''Responsiveness to simple and complex auditory stimuli in the human newborn,'' Dev. Psychobiol. **5**, 7–19.
Viemeister, N. F., and Schlauch, R. S. (**1992**). ''Issues in infant psychoacoustics,'' in *Developmental Psychoacoustics*, edited by L. A. Werner and E. W. Rubel (American Psychological Association, Washington, DC), pp. 191–210.
Werner, L. A. (**1992**). ''Interpreting developmental psychoacoustics,'' in *Developmental Psychoacoustics*, edited by L. A. Werner and F. W. Rubel (American Psychological Association, Washington, DC), pp. 47–88.
Werner, L. A. (**1995**). ''Observer-based approaches to human infant psychoacoustics,'' in *Methods in Comparative Psychoacoustics*, edited by G. M. Klump, R. J. Dooling, R. R. Fay, and W. C. Stebbins (Birkhauser, Boston), pp. 135–146.
Werner, L. A., and Bargones, J. Y. (**1991**). ''Sources of auditory masking in infants: Distraction effects,'' Percept. Psychophys. **50**, 405–412.
Werner, L. A., and Gray, L. (**1998**). ''Behavioral studies of hearing development,'' in *Development of the Auditory System*, edited by E. W. Rubel, R. R. Fay, and A. N. Popper (Springer, New York), pp. 12–79.
Werner, L. A., and Marean, G. C. (**1991**). ''Methods for estimating infant thresholds,'' J. Acoust. Soc. Am. **90**, 1867–1875.
Wightman, F., and Allen, P. (**1992**). ''Individual differences in auditory capability among preschool children,'' in *Developmental Psychoacoustics*, edited by L. A. Werner and E. W. Rubel (American Psychological Association, Washington, DC), pp. 113–113.
Willihnganz, M. S., Stellmack, M. A., Lutfi, R. A., and Wightman, F. L. (**1997**). ''Spectral weights in level discrimination by preschool children: synthetic listening conditions,'' J. Acoust. Soc. Am. **101**, 2803–2810.
Yost, W. A., and Klein, A. J. (**1979**). ''Thresholds of filtered transients,'' Audiology **18**, 17–23.

Spatial release from informational masking in speech recognition

Richard L. Freyman,[a] Uma Balakrishnan, and Karen S. Helfer
Department of Communication Disorders, University of Massachusetts, Amherst, Massachusetts 01003

(Received 8 June 2000; revised 25 October 2000; accepted 13 January 2001)

Three experiments were conducted to determine the extent to which perceived separation of speech and interference improves speech recognition in the free field. Target speech stimuli were 320 grammatically correct but nonmeaningful sentences spoken by a female talker. In the first experiment the interference was a recording of either one or two female talkers reciting a continuous stream of similar nonmeaningful sentences. The target talker was always presented from a loudspeaker directly in front (0°). The interference was either presented from the front loudspeaker (the F-F condition) or from both a right loudspeaker (60°) and the front loudspeaker, with the right leading the front by 4 ms (the F-RF condition). Due to the precedence effect, the interference in the F-RF condition was perceived to be well to the right, while the target talker was heard from the front. For both the single-talker and two-talker interference, there was a sizable improvement in speech recognition in the F-RF condition compared with the F-F condition. However, a second experiment showed that there was no F-RF advantage when the interference was noise modulated by the single- or multi-channel envelope of the two-talker masker. Results of the third experiment indicated that the advantage of perceived separation is not limited to conditions where the interfering speech is understandable.
[DOI: 10.1121/1.1354984]

PACS numbers: 43.66.Pn, 43.66.Qp, 43.66.Dc, 43.71.Es [DWG]

I. INTRODUCTION

Many studies have demonstrated that speech recognition in noise improves when the source of the speech is separated horizontally from the source of the background interference (e.g., Hirsh, 1950; Carhart *et al.*, 1969; Dirks and Wilson, 1969; Plomp, 1976; Plomp and Mimpen, 1981; Bronkhorst and Plomp, 1988, 1992; Koehnke and Besing, 1996; Yost *et al.*, 1996; Ericson and McKinley, 1997; Noble *et al.*, 1997; Peissig and Kollmeier, 1997; Hawley *et al.*, 1999; Freyman *et al.*, 1999). The improvement is not surprising, because the separation of signal and noise along the horizontal azimuth lowers the threshold for detecting signals in the presence of the noise over a wide range of signal frequencies (Saberi *et al.*, 1991; Good *et al.*, 1997; Freyman *et al.*, 1999). The lower thresholds are due largely to head shadow for high frequency sounds and binaural interaction for low frequency sounds. Zurek's (1993) model of binaural speech recognition that considers only this improvement in audibility accounts successfully for much of the speech-in-noise data obtained before and since the model was published (e.g., Plomp, 1976; Plomp and Mimpen, 1981; Bronkhorst and Plomp, 1988; Freyman *et al.*, 1999).

Zurek's model was specifically limited to the case of anechoic conditions with stationary background noise. In typical rooms, reflections would be expected to disrupt the interaural differences that form the basis of spatial separation advantages. However, due to the precedence effect, reflections have a minimal effect on the *perceived* locations of the presented stimuli (e.g., Shinn-Cunningham *et al.*, 1993; Litovsky and Macmillan, 1994; Chiang and Freyman, 1998). Hirsh (1950) and Broadbent (1954) suggested that perceived differences in location between speech and interference might contribute to improved speech understanding. We suspect that this may be especially true if the interference was also speech, so that the target and interference might be difficult to segregate if they did not appear to be in different locations. The perceived differences in location could assist the listener in attending to the target speech in the presence of interference, possibly through higher level processes. However, there has been very limited work that directly evaluates this hypothesis.[1]

Freyman *et al.* (1999) investigated the role of perceived spatial separation of target and interference by limiting the detection advantages that spatial separation creates. They used the precedence effect to create the illusion of spatial separation while minimizing detection advantages from head shadow and binaural interaction. In a baseline condition, the target speech, which consisted of nonmeaningful sentences spoken by a female talker, was presented from a loudspeaker directly in front, while a continuous speech-spectrum noise was presented from the same loudspeaker. Speech recognition performance improved consistently with increasing signal-to-noise ratio as expected. In one experimental condition, the speech-spectrum noise interference was presented from a loudspeaker 60 deg to the right in addition to the front loudspeaker, with the front loudspeaker delayed by 4 ms relative to the right loudspeaker. This precedence effect condition created the perception that the noise was on the right, well separated from the target that was presented only at 0 deg. Despite the perceived spatial separation of speech and

[a]Author to whom correspondence should be addressed; electronic mail: RLF@comdis.umass.edu

noise, speech recognition did not improve relative to the condition where they were both from the same loudspeaker. This suggests that the perception of separation of speech and noise is not important in the case of listening to speech in stationary background noise.

The results were different when the masker, like the target, was a recording of a (different) female talker reading the same type of sentences. Relative to the case where the target and interference were presented from the same (front) loudspeaker, there was a marked improvement in speech recognition when the female talker was presented from two loudspeakers (and perceived to be well to the right), while the target was presented only from directly in front. Freyman *et al.* (1999) proposed that perceived separation between speech and interference was important in this case, because physical measurements from KEMAR as well as masked detection thresholds suggested that enhanced speech recognition was not due to less masking, at least in the traditional sense.

Why was there an advantage in the precedence effect condition when the masker was speech but not when it was steady noise? Freyman *et al.* (1999) suggested that the large advantage with the female talker interference occurred because this masker contained ''informational masking'' in addition to energetic masking, similar to the ''perceptual masking'' concept described by Carhart *et al.* (1969). Energetic masking is the traditional conceptualization of masking in which the detection of the signal is reduced by the neural response to the masker in the same frequency channels as the signal. Informational masking is assumed to involve higher level processes that interfere with the listener's ability to follow patterns in the target within a complex masker (see, for example, Watson *et al.*, 1976; Leek *et al.*, 1991; Kidd *et al.*, 1994, 1995, 1998). Research using tonal patterns has shown the benefit of spatial separation to be especially large with informational maskers (Kidd *et al.*, 1998). In the case of two speech messages, Freyman *et al.* (1999) proposed that perceived spatial separation provided a cue that helped listeners segregate the speech of the two talkers. This cue was unnecessary for segregating the target speech from the continuous noise.

The goal of the current research was to determine more precisely the characteristics of a masker that are necessary for it to produce informational masking of speech, as evidenced by improved speech recognition when target and interference are perceived to be in separate locations. To determine whether the masker itself must be understandable speech, this study explored whether unintelligible speech, or noise with the temporal and spectral fluctuations of speech, might also create perceived separation advantages. The first experiment, however, was a replication and extension of the previous study (Freyman *et al.*, 1999) with a different interfering talker and with two interfering talkers. The purpose of doing this was to ensure that the effects measured in the earlier experiment could be generalized beyond the single target/interference combination used in that study.

II. EXPERIMENT I: ONE- AND TWO-TALKER SPEECH INTERFERENCE

A. Method

1. Stimuli

Target sentences were 320 nonsense sentences spoken by a college-aged female speaker of standard American English (JF). These were the same recordings used in Freyman *et al.* (1999). They had originally been recorded on digital audiotape. The tape recorder analog output was low-pass filtered at 8.5 kHz and digitally sampled at 20 kHz using a 16-bit analog-to-digital converter (TDT AD1) for storage on computer disk. The sentences were not semantically correct, but had a subject, verb, object syntax, e.g., ''A *shop* could *frame* a *dog*.'' Each sentence contained three key words (*italicized* above) that were scored by the experimenters. The sentences were divided into 16 lists of 20 sentences each. In one condition, the interference was a recording of a second female talker (SS) reading similar sentences to those of the target. The original recordings were tape recorded, then replayed and digitized in a manner similar to the target. Using waveform editing software, each sentence was normalized to equate maximum amplitudes across sentences. Pauses between the sentences that occurred during the recording were removed, resulting in a 35-s stream of uninterrupted sentences. The stream was repeated continuously for approximately 30 min and recorded onto digital audiotape for use during the experiment. In the two-talker condition, SS's recordings were combined on an equal rms basis with a similar recording of talker TK, who had been used as the talker interference in Freyman *et al.* (1999).[2] The two-talker combination also was recorded on digital audio tape.[3] Both the target and competing speech samples were analyzed spectrally by passing the sampled waveforms through a bank of one-third-octave filters in the range of 160–6300 Hz. The long-term average output of the filters is plotted in Fig. 1 (top set of lines).

2. Apparatus

The experiments were conducted in an anechoic chamber measuring $4.9\,\mathrm{m}\times 4.1\,\mathrm{m}\times 3.12\,\mathrm{m}$. The walls, floor, and ceiling are lined with 0.72 m foam wedges. Subjects were seated in the center of the room in front of a foam-covered semicircular arc on which two loudspeakers were positioned. The front loudspeaker was at 0 deg horizontal azimuth; the right loudspeaker was at 60 deg to the right. Both were 1.9 m from the approximate center of the subject's head and were at ear-height for the typical adult.

The target sentences were delivered via TDT System I instrumentation. The output of the 16-bit D/A converter (TDT DA1) running at 20 kHz was low-pass filtered at 8.5 kHz (TDT), attenuated (TDT PA3), and mixed with the interference before being delivered to a Crown D40 amplifier and a Realistic Minimus 7 loudspeaker. The interference was delivered from a digital audiotape player and fed to a delayer (Klark Teknik DN716), one output of which was delayed by 4 ms relative to the other. The delayed output was mixed with the target (TDT SUM3) prior to power amplification and was delivered to the front loudspeaker. The undelayed

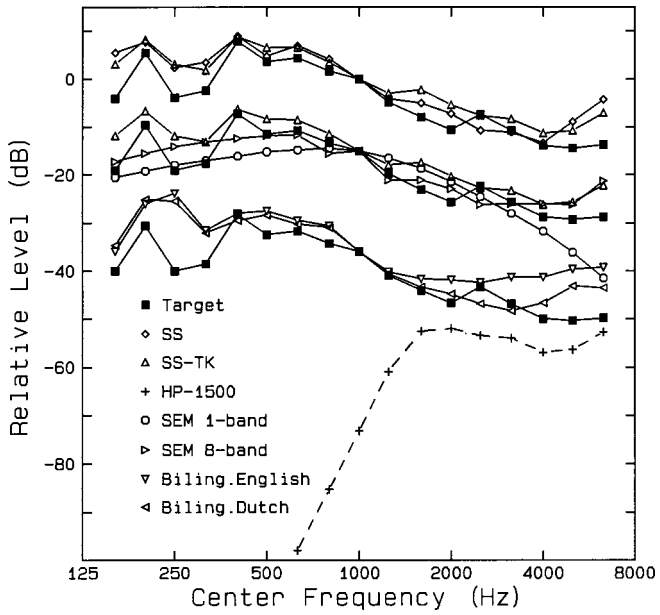


FIG. 1. Long-term averaged outputs of one-third-octave band filters for the target (filled squares) and seven types of interference. The filter outputs have been equated at 1 kHz and offset to show separate sets of lines for each experiment. The lines at the top show the target with the single-talker masker (SS) and the two-talker masker (SS-TK) used in experiment I. The second set shows the target along with two noise maskers: speech-envelope-modulated speech-spectrum noise (SEM one-band) and speech-envelope-modulated eight-channel noise (SEM eight-band). For comparison, the spectrum of the two-talker interference from which the SEM noises were derived has been plotted as well. The third set of lines shows the target along with Dutch and English maskers recorded by bilingual talkers. The dashed line (arbitrarily offset) displays the long-term averaged spectrum of the high-pass filtered SS-TK masker used in experiment I.

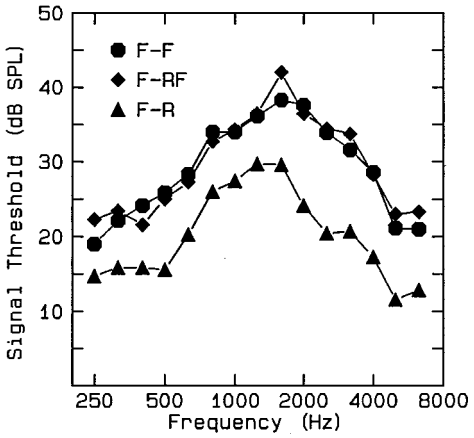


FIG. 2. Thresholds for 15 one-third-octave band noise signals in the presence of speech-shaped noise (from Freyman *et al.*, 1999, Table IV). The data plotted represent the mean thresholds for three subjects, based on three estimates of threshold per subject. Thresholds were obtained using an adaptive 4AFC procedure. The signal was always presented from the front loudspeaker. In the F-F condition, the masker was also presented from the front loudspeaker. In the F-RF condition, the masker was presented from both loudspeakers with the right leading the front by 4 ms. In the F-R condition, the masker was presented from the right loudspeaker.

output was delivered to the right loudspeaker, but was switched off in the conditions in which only the front loudspeaker was to be used. Calibration of presentation level was by means of a 1-in. microphone (B&K 4145) fitted with a random incidence corrector and lowered to the position of the subject's head with the subject absent. A sound level meter (B&K 2204) located outside the chamber measured the microphone output using the A scale and fast meter response. Small differences measured in the level of the target across the 320 sentences were minimized during the experiment using the TDT PA3 attenuator. Daily calibration of target and interference levels was performed using a broadband periodic signal with a fundamental frequency of 100 Hz.

3. Conditions

The target-interference combination could either be F-F (target and interference from the front loudspeaker) or F-RF (target from the front loudspeaker and interference from both loudspeakers, with the right leading the front by 4 ms). In Freyman *et al.* (1999), we measured detection thresholds from three subjects in both of these loudspeaker conditions in the presence of a continuous speech spectrum noise. The data from those threshold measurements are plotted here in Fig. 2, along with the threshold data from the F-R condition for comparison. The results indicate that the F-RF condition does not produce an appreciable release from masking relative to the F-F condition. Thus we attribute any improvements in the F-RF condition to perceived differences in location between target and masker.

There were two types of interference, single-talker (SS) or two-talker (the SS-TK composite). The four conditions were repeated at 4 target-to-interference ratios of -12, -8, -4, and 0 dB, for a total of 16 conditions. Calibration of the target was based on the median of a large sample of sound level meter peak needle readings. It was always presented at 46 dBA both within and across blocks of 20 sentences. The level of the interference, calibrated similarly, was varied across blocks to create the different S-N ratios. The interference level used to compute the S-N ratio refers to the level from a single loudspeaker, i.e., no adjustment in reported S-N ratio was made when the interference was presented from both loudspeakers (which resulted in a 3 dB increase).[4] For example, an S-N ratio of -4 dB was created by presenting the interference at 50 dBA from the front loudspeaker only for the F-F condition, and by presenting at 50 dBA from each of the two loudspeakers in the F-RF condition.

Each of 16 normal-hearing college-aged subjects completed the experiment in a 1-h long experimental session. Their normal-hearing status was confirmed through audiometry, which demonstrated hearing threshold levels $\leqslant$20 dB HL (ANSI, 1996) at 0.5, 1.0, 2.0, 3.0, 4.0, and 6.0 kHz. To minimize the potential influences of subject and sentence list differences, the experiment was conducted using a completely within-subject Latin square design. Each subject listened to all 16 conditions once in a random order, with a different assignment of lists to conditions for each subject. The final data set for each condition consisted of all 16 lists as listened to by all 16 subjects. Thus each data point was based on the percentage of key words perceived correctly on 960 scored items (16 lists$\times$20 sentences per list$\times$3 key words per sentence).

Following initial instructions, subjects were given five

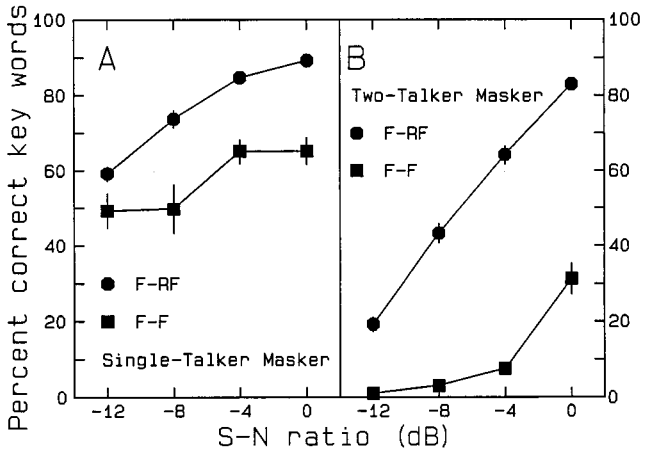


FIG. 3. Mean percent correct key word scores and ± one standard error for experiment I. Panel A shows data for a single-talker masker (SS). Filled squares represent the F-F condition, filled circles the F-RF condition. Panel B shows data for the two-talker interference.

practice trials with the target presented in quiet and then in the presence of interference for the F-RF and F-F loudspeaker conditions. Each subject listened to 16 blocks of 20 sentences with a short break after 8 blocks. Individual trials were initiated by a listener's button press. The interference was gated on first, followed by a single target sentence, which began 0.6 to 1.2 s later. The delay provided a basis for the subject to attend to the target. Because the interference was presented from a continuous loop, the initial words presented when the masker was gated on were variable from trial to trial, occurring at any point during the masker recording. The target and interference ended simultaneously. The listener was instructed to repeat the target sentence to the best of his/her ability after each trial. Listener responses were transmitted to the experimenter's station outside the anechoic chamber via a clip-on microphone. While listeners were free of physical restraint, they were instructed to maintain a head position facing the front loudspeaker.

B. Results

With talker SS as the interference, the single-talker masker results shown in Fig. 3(a) reveal the same general trends that had been found with the TK talker interference in Freyman *et al.* (1999). The slopes of the psychometric functions were shallow, and the F-RF condition produced better results at each signal-to-noise ratio than the F-F condition.[5] The primary difference between SS and TK as maskers was the location of the plateau in the recognition functions. In Fig. 3, the data for masker SS shows no change in performance between −4 and 0 dB S-N ratio. The results for TK had shown this same type of plateau but between −8 and −4 dB S-N ratio.

The two-talker interference results showed even more dramatic improvements in the F-RF conditions relative to the F-F condition [Fig. 3(b)]. In the F-F configuration, listeners appeared to have great difficulty hearing and following the target within the mixture of three total talkers. This is consistent with the findings of Carhart *et al.* (1969), where a complex masker containing two voices was more disruptive than when the masker consisted of one voice. In the present study, it was not until the 0 dB S-N ratio was reached that subjects correctly identified more than an occasional key word out of the 60 in each list. In contrast, the performance in the F-RF condition was considerably better, and improved smoothly and steadily as S-N ratio was increased. The difference in S-N ratio for 30% correct performance was approximately 10 dB.

C. Two follow-up studies

1. High-pass filtering

The absence of substantial threshold differences for the F-F and F-RF condition shown in Fig. 2 supports a conclusion that the improved speech recognition in the F-RF condition is not due to traditional release from masking, but rather to differences in perceived location of target and interference in that condition. The F-RF masker is, after all, the F-F condition plus *additional* masking from the right loudspeaker. The additional masking would clearly not improve the overall S-N ratio at either ear, and measurements made through KEMAR (see Freyman *et al.*, 1999) verified this. However, the F-RF condition is complex with respect to interaural time delay because the interference comes from two loudspeakers. The masker from the right loudspeaker is delayed at the left ear relative to the right whereas the masker from the front loudspeaker, like the target, does not create an interaural delay. There also is a 4-ms delay at the front loudspeaker relative to the right. The configuration apparently does not produce release from masking (see Fig. 2), even in the low-frequency region where binaural processing of time delay is dominant. However, the detection experiment demonstrating this was conducted with an unmodulated broadband noise masker. To be sure that the speech recognition data were not affected by low-frequency release from masking effects particular to the spectrally and temporally complex talker maskers, we conducted an experiment to determine whether the F-RF advantage in speech recognition would occur with high-pass filtered stimuli. Both the target and interference were high-pass filtered at 1500 Hz to minimize the potential for contribution from low-frequency binaural interaction.

The filtering of target sentences and of the SS-TK two-talker masker was accomplished digitally using Matlab with two cascaded four-pole high-pass Butterworth filters for a total of eight poles. The one-third-octave spectrum of the masker is shown with the dashed lines in Fig. 1. The stimuli were calibrated and presented as before with eight conditions (F-F and F-RF at four S-N ratios) delivered to eight subjects in a Latin square design. None of the subjects had participated in the first study. Two of the 16-sentence lists were presented for each condition with the assignment of lists to conditions different for each subject. Across the 8 subjects, all 320 sentences (960 key words) were delivered for each condition. The results displayed in Fig. 4 clearly show that performance was better in the F-RF condition even when the target and interference stimuli had been confined largely to the frequency region above 1500 Hz. These data from the high-pass filtered condition do show some differences from the unfiltered data shown in Fig. 3(b), the most obvious be-

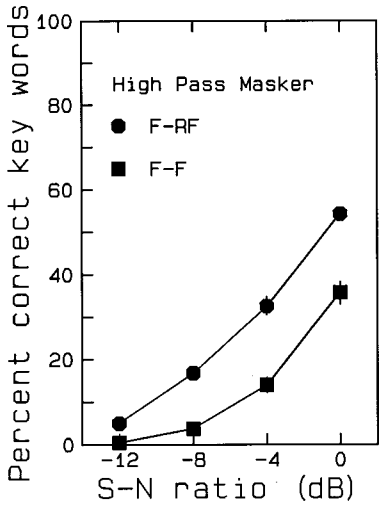


FIG. 4. Mean percent correct key word scores and ± one standard error for high-pass filtered two-talker interference (SS-TK). Filled squares represent the F-F condition, filled circles the F-RF condition.

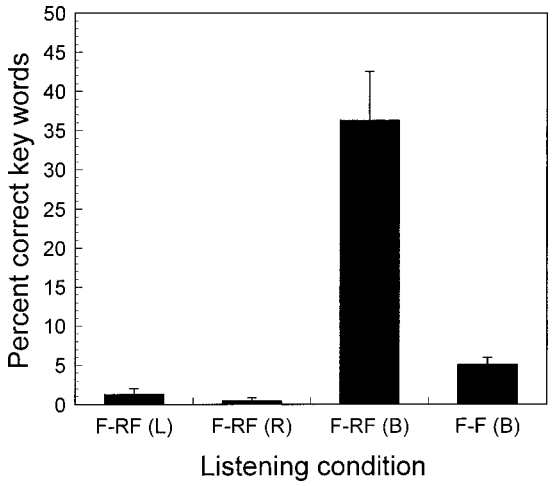


FIG. 5. Mean percent correct key word scores and ± one standard error for four subjects listening under headphones. Target and interference (SS-TK) were recorded via KEMAR in the F-RF and F-F loudspeaker conditions. Four listening conditions are shown: monaural left, F-RF(L), monaural right, F-RF(R), binaural, F-RF(B), and binaural, F-F(B).

ing a smaller difference between the results for the F-RF and F-F conditions. We speculate that the high-passed two-talker interference produced less informational masking than the unfiltered interference because it was not itself as intelligible. The role of the intelligibility of the masker is considered further in experiment III.

2. Monaural versus binaural presentation

A second short follow-up experiment was conducted with a subset of conditions in order to determine whether monaural cues may have played a role in the improvement in the F-RF condition. The 4-ms inter-loudspeaker masker delay produces a rippled spectrum with a 250-Hz periodicity. With the continuous speech spectrum noise masker, this ripple apparently produced little if any effect on the masked audiogram (see Fig. 2). However, the spectral and intensive properties of the talker maskers vary over time and differ from those of the continuous noise masker used in the masked detection experiment. To be certain that the F-RF masker did not create any monaural spectral cues that contributed to the reduced masking found with the speech interferers in that condition, we retested the F-RF conditions monaurally after recording the target and the two-talker interference through KEMAR under a subset of the experimental conditions. The experiment was run at the −4 dB S-N ratio with four of the target lists in both the F-F and F-RF conditions. A KEMAR manikin fitted with Zwislocki couplers and $\frac{1}{2}$-in. microphones (ER-11) was placed in the position where subjects normally were seated. The microphone preamplifier outputs, with the ear canal inverse filters engaged, were recorded onto digital audiotape. The tape recordings were digitized at 22.05 kHz and presented to subjects seated in the anechoic chamber via TDH-39 earphones. Each of four new normal-hearing subjects listened to one 20-sentence list at each of four conditions: F-RF in the left ear only, F-RF in the right ear only, F-RF binaural, and F-F binaural. A Latin square design was again used to assign lists to conditions separately for each subject.

The results across subjects, which were based on 240 scored key words per condition (60 key words×4 lists/subjects), are shown in Fig. 5. Performance was extremely poor in the monaural conditions, suggesting that spectral ripple introduced by the masker delay in the F-RF condition was not responsible for the improved recognition of the target. The data indicate that only in the binaural F-RF condition was a reasonable level of speech recognition achieved. Presumably, subjects used the cue of perceived separation in that condition; i.e., the masker was heard well to the right and the target near the center. Perceived externalization in the binaural conditions was not assessed formally, but informal listening suggested that neither target nor interference was highly externalized, which is not surprising for KEMAR recordings in an anechoic chamber. The lack of externalization may have resulted in a less clear spatial distinction between the target and the two-talker masker, and might explain why performance was not as good in the binaural F-RF condition at the −4 dB S-N ratio as it was in the actual free-field data [Fig. 3(b)].

III. EXPERIMENT II: ENVELOPE MODULATED NOISE

The purpose of this experiment was to determine whether amplitude and/or spectral fluctuations in the speech maskers were responsible for the improvement in the F-RF condition relative to the F-F condition. Improvement was not seen with the unmodulated noise maskers used by Freyman *et al.* (1999), raising the possibility that the envelope modulation of the speech masker was the critical characteristic. In the current experiments the interference was noise modulated by the envelope of the two-talker masker. In one condition, speech-spectrum noise was modulated by the wide-band envelope of the masker. This masker retained the amplitude fluctuations in the speech masker, but not the spectral variation. In the second condition the noise was modulated individually in eight separate frequency channels, then recombined. The number of channels was chosen somewhat arbitrarily to simulate both the amplitude and spectral varia-

tions characteristic of the two-talker masker. We sought to determine whether noise with speechlike amplitude or amplitude/spectral fluctuations would produce results more similar to the actual two-talker masker or to unmodulated noise.

A second question asked in this experiment was the nature and extent of the masking produced by combinations of the two-talker masker and the modulated noise derived from the masker. If added in synchrony with the two-talker masker, the speech-envelope-modulated (SEM) noise could be expected to both reduce the intelligibility of the masking speech and cause it to sound less speechlike. If either or both of these features of the two-talker masker are relevant to the interference that it creates, the effect of the additional energy produced by adding the noise to the speech masker might be offset by a reduction in the informational masking produced by the speech interference. Thus we wanted to know whether less-than-expected reductions in performance would occur with the addition of noise energy to the two-talker masker. A more interesting result would be if no reduction in performance occurred, or, even more provocative, if addition of noise to the speech masker actually enhanced the intelligibility of the target.

A. Method

1. Stimuli

The target stimuli were the same as used in experiment I. The noise interference was created by multiplying noise by the envelopes of the two-talker interference used in experiment I. In experiment II A, white noise was multiplied by the wide-band envelope of this two-talker interference. To derive the envelope, the sampled two-talker waveform was full-wave rectified and low-pass filtered at 20 Hz with a three-pole Butterworth filter. The envelope was then multiplied by a white noise, and the resultant low-pass filtered at 1 kHz with a slope of 12 dB/per octave. The long-term one-third-octave spectrum of this speech-envelope-modulated noise (SEM) masker is compared to the spectrum of the two-talker masker itself and to the target talker in Fig. 1 (second set of lines).

For the eight-channel interference used in experiment II B, the two-talker masker was filtered in eight frequency bands using the following filter cutoff frequencies (all in Hz): 0–707, 707–1000, 1414–2000, 2000–2828, 2828–4000, 4000–5656, 5656–8000. The first filter was a three-pole Butterworth low-pass filter; the others were three-pole Butterworth bandpass filters. The 20-Hz envelope of each of the resulting filters was extracted as described above. The resulting signals were multiplied by white noise, and then filtered again using the cutoff frequencies just described. The eight band-limited noises were then summed to form the eight-channel SEM noise. This procedure is similar in several respects to the methods used by Shannon *et al.* (1995), Dorman *et al.* (1997), and Eisenberg *et al.* (2000). In those experiments, multi-channel speech-envelope-modulated noise was highly intelligible. It was not easy to recognize individual words and sentences when listening informally to the current noise stimuli. Most likely, this is because the processing was conducted on the speech of two talkers mixed together. The long-term spectrum of the eight-band SEM noise is shown with the other spectra in Fig. 1 (second set of lines).

For the speech plus noise portion of the experiments, each SEM noise and the two-talker masker from which it was derived were combined into a two-channel waveform file and were mixed during playback. The envelope extraction process created a short delay such that the noise envelope was shifted by 16 ms relative to the speech envelope. To correct for this, a pad of zeros 16 ms long was added to the beginning of the speech masker so that the noise and speech envelope peaks were aligned. During the experiment, each channel of the two-channel waveform was attenuated individually then mixed (TDT SUM3) prior to presentation through the loudspeakers.

2. Conditions

As in experiment I there were 16 conditions each for experiments II A and II B. Eight of the conditions used the SEM noise at S-N ratios of −12, −8, −4, and 0 dB for both the F-F and F-RF conditions. The other eight conditions were devoted to the study of maskers that were the combination of the two-talker masker and the SEM noise. The two-talker masker was presented in all eight of these conditions at a 0 dB S-N ratio relative to the target. In two of the conditions it was presented alone (either in the F-F or F-RF configuration). For the other conditions the two-talker masker was mixed with the SEM noise. The noise was adjusted to have S-N ratios of −6, −4, or −2 dB relative to the target before it was mixed with the two-talker masker. Thus there were a total of six such combination masker conditions (3 S-N ratios for both F-F and F-RF configurations). A new group of 16 young normal-hearing subjects listened to the 16 conditions of experiment II A and another, completely different group of 16 subjects participated in experiment II B. None of the subjects had participated in experiment I. A Latin square design was used as before for both experiments.

B. Results

The recognition scores for the single-channel envelope-modulated noise, shown in Fig. 6(a), are actually slightly worse in the F-RF condition than in the F-F condition. The small but statistically significant effect (see footnote 4) was probably due to the increase in masker energy when the right loudspeaker was added in the F-RF condition. For the multi-channel noise [Fig. 7(a)] no statistically significant difference was found between the two conditions. In neither case is there evidence that the F-RF condition was easier than the F-F condition. Whereas these results are consistent with the data for unmodulated broadband noise found by Freyman *et al.* (1999), they are far different from results for the actual two-talker masker from which the envelope was derived [Fig. 3(b)]. In that case there was a substantial improvement in the F-RF condition relative to the F-F condition. The similarity between the results for the F-F and F-RF configurations for the modulated noise suggests that the amplitude

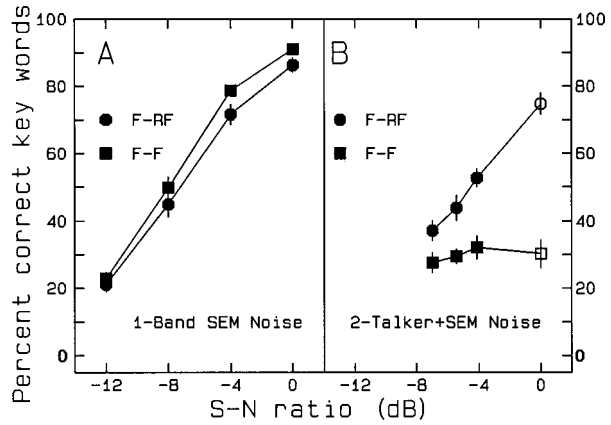


FIG. 6. Mean percent correct key word scores and ± one standard error for experiment II. Panel A shows data for a single-band speech-envelope-modulated (SEM) noise masker. Filled squares represent the F-F condition, filled circles the F-RF condition. Panel B shows data for the SEM noise interference mixed with the two-talker interference (SS-TK) from which it was derived. The abscissa is the target-to-total-interference ratio, the interference consisting of the sum of the talker masker and the SEM noise. Filled symbols represent the conditions where the noise and two-talker interference were mixed. The unfilled symbols at the right in panel B represent data for the speech interference alone condition.

modulations in the talker maskers are not responsible for the improvement in the F-RF condition that occurred in experiment I.

The filled circles and squares in Figs. 6(b) and 7(b) show the results of combining the actual two-talker masker (at 0 dB S-N ratio) with the noises that were derived from the maskers. The unfilled data points at 0 dB S-N ratio indicate the conditions where the two-talker interference was used without the additional noise. The abscissa is the target-to-total-interference ratio, the interference consisting of the sum of the talker masker and the SEM noise. For the single-channel noise [Fig. 6(b)], the F-RF condition demonstrates a steady improvement in performance with increasing S-N ratio, reaching approximately 77% correct at 0 dB (at which point the masker consisted of the two-talker interference with no additional noise). The F-F condition shows essentially no change in performance as a function of S-N ratio. We speculate that two opposing factors contributed to this. As noise was added, the S-N ratio declined, thereby increasing the amount of energetic masking. However, the added noise also reduced the amount of informational masking by partially masking the two-talker interference from which it was derived. These two opposing factors, that is, increased energetic masking and decreased informational masking, may have been responsible for the flat function over a nearly 8-dB range of S-N ratios.

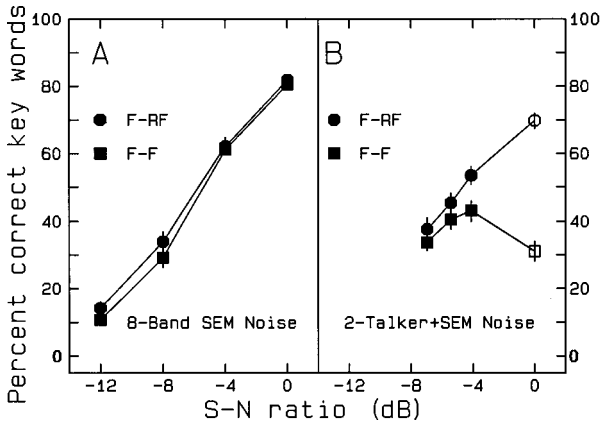


FIG. 7. Mean percent correct key word scores and ± one standard error for the eight-band noise in experiment II. Legend as in Fig. 6.

Figure 7(b) displays the combination masker results for the eight-band noise derived from the two-talker interference. As in Fig. 6(b), the unfilled symbols display the data for the two-talker masker alone, and the filled symbols the results for the combined noise and two-talker masker. Performance in the F-RF condition improved smoothly with S-N ratio. However, percent correct scores for the F-F condition changed nonmonotonically with S-N ratio. Especially in the condition when the SEM noise was added at −2 dB relative to the target, creating a total S-N ratio of −4.12 dB, performance actually improved relative to the two-talker masker only (0-dB S-N ratio condition). We suggest that, because it was closely tailored to the spectral and temporal fluctuations of the two-talker masker, the eight-band SEM noise masked the two-talker masker efficiently, and reduced the informational masking to a greater extent than it added energetic masking.

IV. EXPERIMENT III: THE EFFECT OF THE UNDERSTANDABILITY OF THE MASKER

This experiment examined the degree to which the differences observed between the F-RF and F-F conditions with the talker maskers can be attributed to the fact that the maskers themselves consisted of understandable speech. Two sub-experiments were conducted using speech interference that was not understandable to the listener. In one sub-experiment, the SS-TK two-talker masker was reversed in time. In the other, two bilingual Dutch talkers recorded the interfering speech in Dutch and, for comparison, in English.

A. Method

1. Stimuli

The same target speech recordings were used as in the previous two experiments. For the reversed two-talker interference, the SS-TK masker used in experiment I was reversed digitally. For the Dutch talker interference, two young adult (25 and 30 years of age) native speakers of Dutch recorded nonsense sentences that they had translated into Dutch. For comparison, they recorded the original English sentences as well. Each of the two talkers recorded a different set of 40 nonsense sentences that were also different from those used by the target talker and in the SS-TK masker. Digital audio tape recordings of these sentences were sampled at 22.05 kHz and stored on computer disk. The waveforms of the two bilingual talkers were added to create a two-talker Dutch waveform and a two-talker English waveform, each approximately 35 s long, for continuous playback during the experiment. The long-term average one-third-octave spectra for the two bilingual maskers are shown together with the target in the third set of lines in Fig. 1.

2. Conditions

The design of the experiments was similar to the previous two experiments described. For each sub-experiment, each of 16 subjects listened to all 16 lists of nonsense sentences in 16 different conditions. Thus there were 32 total listeners, none of whom had participated in the earlier studies. In the case of the interference recorded by the bilingual talkers, the 16 conditions were 4 S-N ratios (-12, -8, -4, and 0) $\times$ two loudspeaker conditions (F-F or F-RF) $\times$ two languages (English or Dutch) presented in a Latin square design. The 16 listeners who participated in this experiment all reported that they did not speak or understand Dutch. In the case of the reversed speech conditions a similar design was used but only eight of the conditions were relevant (reversed two-talker masker in the F-F or F-RF conditions at the 4 S-N ratios). The other eight conditions used the same loudspeaker and S-N ratio variations but used a noise masker which consisted of signal-correlated noise (Schroeder, 1968) that had been filtered to have the long-term spectrum of speech. This stimulus is similar to the single-channel speech-envelope-modulated noise used in experiment II. As the results were also highly similar, the data for this condition will not be discussed further.

B. Results

The results for the reversed SS-TK speech interference conditions are shown in Fig. 8, along with the forward SS-TK speech interference data replotted from experiment I with dashed lines. The largest difference was that performance was considerably better in the F-F condition when the masker was reversed than when it was presented normally. This suggests that the understandability of the interfering speech contributed substantially to the difficulty that subjects experienced when identifying the target sentences in the presence of the two-talker masker. By contrast, there was little difference between the forward and reversed maskers in

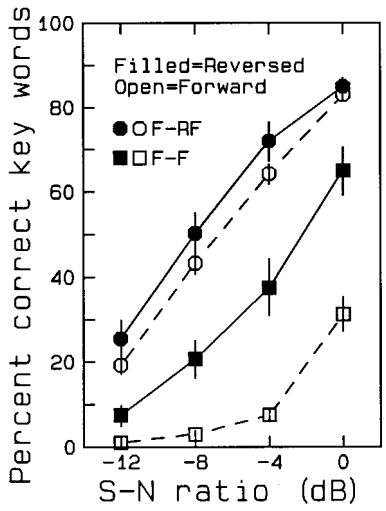


FIG. 8. Mean percent correct key word scores and $\pm$ one standard error for the two-talker (SS-TK) reversed speech interference (filled symbols). For comparison, the SS-TK forward data are replotted from experiment I (dashed lines). Circles represent the F-RF condition and squares the F-F condition.

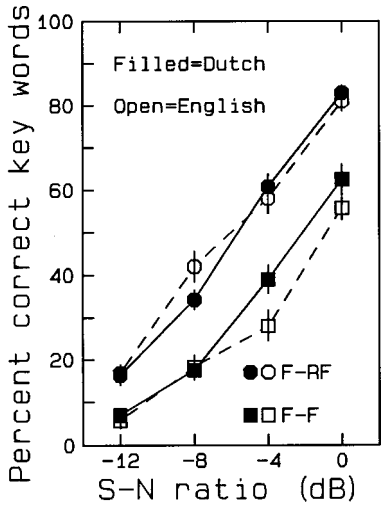


FIG. 9. Mean percent correct key word scores and $\pm$ one standard error for Dutch and English speech interference spoken by two bilingual talkers. Open symbols represent English speech interference and filled symbols, Dutch. Circles represent the F-RF condition and squares the F-F condition.

the F-RF condition. This is similar to results reported by Hygge *et al.* (1992). Apparently, the perceived spatial separation was a sufficient cue to distinguish target from interference, so whether or not the interference was understandable was less important. Even though the reversal caused a greater improvement in the F-F condition than in the F-RF condition, there remained for the reversed speech a substantial difference between the F-F and F-RF conditions. This suggests that perceived spatial separation can lead to improved speech recognition even when the interference is not understood by the listener.

A similar result occurred for the maskers recorded by the native Dutch talkers (Fig. 9). The F-RF vs F-F difference is evident in both the Dutch and English recordings, suggesting that the benefit of the spatial separation occurred even when the speech was not understandable. In the F-F condition, the English recordings by the native Dutch talkers did not interfere with speech recognition as much as did the SS-TK interference [Fig. 9 versus Fig. 3(b)]. However, in the F-RF condition, the difference in the effectiveness of the two types of interference was minimal. This suggests that the English masker spoken by the Dutch talkers produced as much energetic masking as the SS-TK interference (assuming that in the F-RF condition the masking is mostly energetic). The difference in results for the F-F condition between these two masker complexes probably reflects a difference in the amount of informational masking produced.

V. DISCUSSION

We interpret improvements in the F-RF condition relative to the F-F condition as reflecting the contribution of *perceived* spatial separation, which we assume facilitates a release from informational masking. Other explanations that we have considered were ruled out by control experiments, including KEMAR measurements showing that the F-RF condition did not create a headshadow advantage (Freyman *et al.*, 1999), that masked detection thresholds did not improve in the F-RF condition over those in the F-F condition

(Fig. 2), the absence of any measurable advantage for the F-RF condition in either ear monaurally (Fig. 5), and the better performance for the F-RF condition (Fig. 4) even when the target and maskers were restricted to the high-frequency range (where a possible contribution from low-frequency binaural interaction had been minimized). Perceived separation appears to be important in cases in which the speech of one female talker is masked by the speech of one or two other female talkers, whether or not the interfering speech is understood by the listener (Figs. 3, 4, 5, 8, and 9). Perceived separation does not appear to be important when the interference is noise with speechlike amplitude and/or spectral fluctuations over time [Figs. 6(a) and 7(a)].

Presumably, the benefit of perceived separation in releasing informational masking is at its maximum when other cues by themselves are insufficient to isolate and attend to the target in the presence of the interference. Although it has been nearly 50 years since Cherry (1953) listed several factors that might be involved in segregating voices from one another, we are still in the early stages of understanding and quantifying these factors (see Darwin and Carlyon, 1995, and Bronkhorst, 2000 for reviews). One characteristic, similarity in fundamental frequency, has been shown to be important across several studies (e.g., Brokx and Nooteboom, 1982; Summerfield and Assmann, 1991). Although we did not measure fundamental frequency, the fact that the target and interference speech stimuli were recorded by talkers of the same sex probably helped to make them difficult to untangle without spatial separation. Segregation appears to be easier when target and competing speech are from different-sex talkers. In fact, Festen and Plomp (1990) showed that for subjects listening to a male target talker, interference consisting of the same male talker presented backward created considerably more masking than when the interference was female speech presented either forward or backward, although it is not clear how much of the additional masking could be attributed to informational masking. Using competing talkers of different sexes, Duquesnoy (1983) measured spatial separation advantages that were reasonably similar to the advantages measured when the masker was unmodulated noise. Thus there is no suggestion in those results that listeners benefited additionally from *perceived* spatial separation in the competing-speech conditions. Similarly, the advantages seen in the F-RF condition in the current study would probably be smaller if target and interfering speech were from different-sex talkers. Of course, this does not mean that same-sex competing talkers are all difficult to segregate from one another. In this case, the advantages created by perceived separation might be influenced by a host of factors that determine voice quality, including similarity in fundamental frequency.

The role of spatial separation in segregation and auditory object formation has been the subject of considerable study, with most results showing that it is a much weaker cue than, for example, frequency separation (see Bregman, 1990). A simple example is that two tones of identical frequency presented through left and right earphones are perceived as a unitary image within the head, while tones separated sufficiently in frequency and presented to the same ear can be heard as two separate tones. Other examples come from ''double-vowel'' experiments in which different formants are represented with different interaural time delays (ITDs). These studies indicate that listeners do not segregate the images into vowel perceptions based upon formants with a common ITD (Culling and Summerfield, 1995; Hukin and Darwin, 1995). On the other hand, Darwin and Hukin (1999) suggest that listeners attend across time to auditory objects perceived at particular locations, even though the segregation of competing speech stimuli into such objects may be based primarily on other grouping principals, such as harmonicity and temporal synchrony. In the current studies, the target and masking stimuli were asynchronous and differed in a number of dimensions not directly measured (e.g., fundamental frequency contours). The perceived spatial separation created by the precedence effect may have provided listeners with a means to connect the relevant characteristics of the target talker's speech across time, which helped to overcome informational masking.

In our experiments the target and masker were both broadband, so there was clearly energetic masking in addition to informational masking. To estimate the size of the informational masking component, we assume that informational masking is at its maximum when the target and interference are presented from the same loudspeaker and fully released by the perceived separation created in the F-RF condition. The size of the informational masking component may then be determined from the shift in decibels between the F-RF and F-F conditions for equal recognition scores. This difference must be thought of as a range of values because of the differing slopes of growth in performance with S-N ratio. Estimates are 6–10 dB for the SS masker and the SS-TK two-talker masker, 4–6 dB for the Dutch speakers speaking English, 3–4 dB for the Dutch speakers speaking Dutch, and 6 dB for the reversed SS-TK masker. In view of the fact that spatial release from masking for speech in continuous noise with the target at 0 deg is maximally only about 10 dB in anechoic environments (see Zurek, 1993), and considerably less than this in reverberant environments (e.g., Hirsh, 1950; MacKeith and Coles, 1971; Plomp, 1976; Bronkhorst and Plomp, 1988; Koehnke and Besing, 1996), the release from informational masking observed in this study is of significant magnitude.

Mixtures of the SS-TK speech interference and speech-envelope-modulated noise added in synchrony produced the result that speech recognition performance did not decline, and in the case of the eight-channel noise, actually improved, as the overall S-N ratios decreased in the F-F condition [Figs. 8(b) and 9(b)]. Apparently, the noise reduced the informational masking produced by the speech interference. Over a range of conditions this was beneficial for speech recognition, or at least offset the increase in energetic masking produced by the added noise. This result is similar to that found by Kidd *et al.* (1997). They showed a marked increase in the detectability of a 1000-Hz tone in a spectrally complex informational masker when Gaussian noise was added to the masker.

The differences between the shifts in the forward and reversed conditions and between the Dutch and English

speech suggest that a portion, but by no means all, of the informational masking is related to the fact that the masker is itself understandable speech. It is significant that, in the condition using the unintelligible reversed two-talker masker, listeners benefited from the F-RF separation, whereas no benefit of spatial separation was shown using a masker that had a speechlike short- and long-term spectral and temporal envelope (i.e., the eight-channel speech-envelope-modulated noise masker). At this time it is unclear which of several differences between these two types of stimuli (e.g., fine structure, periodicity, pitch variation) are responsible for the differences in results.

Additional study will be necessary to determine whether the size of the putative informational component estimated here (which is as large as 10 dB) is typical of everyday listening conditions. Across this study and that of Freyman *et al.* (1999), we have used only one target talker. However, we have demonstrated a benefit in the F-RF condition with two different individual competing talkers, with the two talkers combined, with the two talkers high-pass filtered, with the two talkers presented backward, and with two other talkers speaking in two different languages. Still, we recognize that we have explored only a small subset of possible target-interference conditions. Further research should identify the critical features of target and interference stimuli that are aided by perceived spatial separation in order to help us understand the nature of informational masking in speech recognition.

ACKNOWLEDGMENTS

This research was supported by a grant from the National Institute of Deafness and other Communicative Disorders (DC01645). The authors would like to thank Elizabeth Adams, Stephanie Sokol, and Shawn Johnson for their assistance with various aspects of the project. We would also like to thank Wes Grantham and Adelbert Bronkhorst for their helpful reviews of an earlier version of this manuscript.

[1]Other than Freyman *et al.* (1999), the only research we are aware of that attempted to isolate the effects of perceived spatial separation was by Broadbent (1954), who found that the target message among two competing speech messages was easier to understand when the target and interference were localized separately. However, the method of creating perceived separation led to unspecified monaural level differences between the two messages, as well as unspecified interaural differences, that could have created actual release from masking. Thus, the degree to which the results obtained by Broadbent (1954) can be attributed to perceived spatial separation is unclear.

[2]The sentences recorded by talker TK had been subjected to broadband automatic gain control at the time of the recording, so there was little variation in peak level across sentences.

[3]In the experiments subsequent to this (i.e., with results shown in Fig. 4 and after) the interference was delivered from a personal computer sound card (Turtle Beach Montego), running at 22.05 kHz with 16 bits of resolution, instead of the digital audio tape player. The interference was always looped continuously.

[4]The 3-dB increase would be expected to occur for a position at the center of the listener's head with the listener absent. For a real listener, the expected increase would depend upon the ear (left or right) as well as the frequency. At the right ear (closest to the 60-deg loudspeaker), the intensity would likely increase by more than 3 dB, particularly at high frequencies. The increase in intensity at the high frequencies would be less than 3 dB at the left (far) ear. Assuming that listeners monitor the ear with the more advantageous S-N ratio (which, in this configuration, is the left ear), the addition of the masking stimulus from the right loudspeaker would result in less than a 3-dB reduction in (effective) S-N ratio.

[5]Data for each experiment were subjected to repeated-measures Analyses of Variance using a 0.01 confidence level. The main effect of listening condition (usually F-F versus F-RF) was significant in each experiment, with the exception of the eight-channel noise data (experiment II). All main effects of S-N ratio were significant. Each listening-condition×S-N ratio interaction also was significant, with the exception of the eight-channel noise experiment (experiment II) and backward speech data (experiment III). The main effect of number of talkers (in experiment I) was statistically significant, as were the interactions between number of talkers and both S-N ratio and listening condition. Finally, in experiment III, the difference in the effectiveness of Dutch versus English maskers did not reach statistical significance, although the interaction between masker language and listening condition was significant.

ANSI (**1996**). ANSI S3.6-1996, ''Specifications for audiometers'' (American National Standards Institute, New York).

Bregman, A. S. (**1990**). *Auditory Scene Analysis* (M.I.T., Cambridge, MA).

Broadbent, D. E. (**1954**) ''The role of auditory localization in attention and memory span,'' J. Exp. Psychol. **47**, 191–196.

Brokx, J. P. L., and Nooteboom, S. G. (**1982**). ''Intonation and the perceptual separation of simultaneous voices,'' J. Phonetics **10**, 23–36.

Bronkhorst, A. W. (**2000**). ''The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions,'' Acta Acust. (China) **86**, 117–128.

Bronkhorst, A. W., and Plomp, R. (**1988**). ''The effect of head-induced interaural time and level differences on speech intelligibility in noise,'' J. Acoust. Soc. Am. **83**, 1508–1516.

Bronkhorst, A. W., and Plomp, R. (**1992**). ''Effect of multiple speech-like maskers on binaural speech recognition in normal and impaired hearing,'' J. Acoust. Soc. Am. **92**, 3132–3139.

Carhart, R., Tillman, T. W., and Greetis, E. S. (**1969**). ''Perceptual masking in multiple sound backgrounds,'' J. Acoust. Soc. Am. **45**, 694–703.

Cherry, E. C. (**1953**). ''Some experiments on the recognition of speech, with one and two ears,'' J. Acoust. Soc. Am. **25**, 975–979.

Chiang, Y. C., and Freyman, R. L. (**1998**). ''The influence of broadband noise on the precedence effect,'' J. Acoust. Soc. Am. **104**, 3039–3047.

Culling, J. F., and Summerfield, Q. (**1995**). ''Perceptual separation of concurrent speech sounds: Absence of across-frequency grouping by common interaural delay,'' J. Acoust. Soc. Am. **98**, 785–797.

Darwin, C. J., and Carlyon, R. P. (**1995**) ''Auditory grouping,'' in *The Handbook of Perception and Cognition (Hearing)*, edited by B. C. J. Moore (Academic, London).

Darwin, C. J., and Hukin, R. W. (**1999**). ''Auditory objects of attention: the role of interaural time differences,'' J. Exp. Psychol. **25**, 617–629.

Dirks, D. D., and Wilson, R. H. (**1969**). ''The effect of spatially separated sound sources on speech intelligibility,'' J. Speech Hear. Res. **12**, 650–664.

Dorman, M. F., Loizou, P. C., and Rainey, D. (**1997**). ''Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs,'' J. Acoust. Soc. Am. **102**, 2403–2411.

Duquesnoy, A. J. (**1983**). ''Effect of a single interfering noise and interfering speech upon the binaural sentence intelligibility of aged persons,'' J. Acoust. Soc. Am. **74**, 739–743.

Eisenberg, L. S., Shannon, R. V., Martinez, A. Schaefer, Wygonski, J., and Boothroyd, A. (**2000**). ''Speech recognition with reduced spectral cues as a function of age,'' J. Acoust. Soc. Am. **107**, 2704–2710.

Ericson, M. A., and McKinley, R. L. (**1997**). ''The intelligibility of multiple talkers separated spatially in noise,'' in *Binaural and Spatial Hearing in Real and Virtual_Environments*, edited by R. H. Gilkey and T. B. Anderson (Lawrence Erlbaum, Hillsdale, NJ).

Festen, J. M., and Plomp, R. (**1990**). ''Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing,'' J. Acoust. Soc. Am. **88**, 1725–1736.

Freyman, R. L., Helfer, K. S., McCall, D. D., and Clifton, R. K. (**1999**). ''The role of perceived spatial separation in the unmasking of speech,'' J. Acoust. Soc. Am. **106**, 3578–3588.

Good, M. D., Gilkey, R. H., and Ball (**1997**). ''The relation between detection in noise and localization in noise in the free field,'' in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. B. Anderson (Lawrence Erlbaum, Hillsdale, NJ).

Hawley, M. L., Litovsky, R. Y., and Colbum, H. S. (**1999**). ''Intelligibility and localization of speech signals in a multi-source environment,'' J. Acoust. Soc. Am. **105**, 3436–3448.
Hirsh, I. J. (**1950**). ''Relation between localization and intelligibility,'' J. Acoust. Soc. Am. **22**, 196–200.
Hukin, R. W., and Darwin, C. J. (**1995**). ''Effects of contralateral presentation and interaural time differences in segregating a harmonic from a vowel,'' J. Acoust. Soc. Am. **98**, 1380–1387.
Hygge, S., Ronnberg, J., Larsby, B., and Arlinger, S. (**1992**). ''Normal-hearing and hearing-impaired subjects' ability to just follow conversation in competing speech, reversed speech, and noise backgrounds,'' J. Speech Hear. Res. **35**, 208–215.
Kidd, G., Jr., and Mason, C. R. (**1997**). ''Combining energetic and informational maskers,'' paper presented at the annual convention of the American Speech-Language-Hearing Association, Boston.
Kidd, Jr., G., Mason, C. R., and Rohtla, T. L. (**1995**). ''Binaural advantage for sound pattern identification,'' J. Acoust. Soc. Am. **98**, 1977–1986.
Kidd, Jr., G., Mason, C. R., Rohtla, T. L., and Deliwala, P. S. (**1998**). ''Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns,'' J. Acoust. Soc. Am. **104**, 422–431.
Kidd, Jr., G., Mason, C. R., Deliwala, P. S., Woods, W. S., and Colburn, H. S. (**1994**). ''Reducing informational masking by sound segregation,'' J. Acoust. Soc. Am. **95**, 3475–3480.
Koehnke, J., and Besing, J. M. (**1996**). ''A procedure for testing speech intelligibility in a virtual listening environment,'' Ear Hear. **17**, 211–217.
Leek, M. R., Brown, M. E., and Dorman, M. F. (**1991**). ''Informational masking and auditory attention,'' Percept. Psychophys. **50**, 205–214.
Litovsky, R. Y., and Macmillan, N. A. (**1994**). ''Sound localization precision under conditions of the precedence effect: effects of azimuth and standard stimuli,'' J. Acoust. Soc. Am. **96**, 752–758.
MacKeith, N. W., and Coles, R. R. A. (**1971**). ''Binaural advantages in the hearing of speech,'' J. Laryngol. Otol. **85**, 213–232.
Noble, W., Byrne, D., and Ter-Horst, K. (**1997**). ''Auditory localization, detection of spatial separateness, and speech hearing in noise by hearing-impaired listeners,'' J. Acoust. Soc. Am. **102**, 2343–2352.
Peissig, J., and Kollmeier, B. (**1997**). ''Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners,'' J. Acoust. Soc. Am. **101**, 1660–1670.
Plomp, R. (**1976**). ''Binaural and monaural speech intelligibility of connected discourse in reverberation as a function of azimuth of a single competing sound source (speech or noise),'' Acustica **34**, 200–211.
Plomp, R., and Mimpen, A. M. (**1981**). ''Effect of the orientation of the speaker's head and the azimuth of a noise source on the speech reception threshold for sentences,'' Acustica **48**, 325–328.
Saberi, K., Dostal, L., Sadralodabai, T., Bull, V., and Perrott, D. R. (**1991**). ''Free field release from masking,'' J. Acoust. Soc. Am. **90**, 1355–1370.
Schroeder, M. R. (**1968**). ''Reference signal for sound quality studies,'' J. Acoust. Soc. Am. **44**, 1735–1736.
Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). ''Speech recognition with primarily temporal cues,'' Science **270**, 303–304.
Shinn-Cunningham, B. G., Zurek, P. M., and Durlach, N. I. (**1993**). ''Adjustment and discrimination measurements of the precedence effect,'' J. Acoust. Soc. Am. **93**, 2923–2932.
Summerfield, Q., and Assmann, P. F. (**1991**). ''Perception of concurrent vowels: Effects of harmonic misalignment and pitch-period asynchrony,'' J. Acoust. Soc. Am. **89**, 1364–1377.
Watson, C. S., Kelly, W. J., and Wroton, H. W. (**1976**). ''Factors in the discrimination of tonal patterns: II. Selective attention and learning under various levels of stimulus uncertainty,'' J. Acoust. Soc. Am. **60**, 1176–1185.
Yost, W. A., Dye, R. H., and Sheft, S. (**1996**). ''A simulated ''cocktail party'' with up to three sound sources,'' Percept. Psychophys. **58**, 1026–1036.
Zurek, P. M. (**1993**). ''Binaural advantages and directional effects in speech intelligibility,'' in *Acoustical Factors Affecting Hearing Aid Performance*, 2nd ed., edited by G. A. Studebaker and I. Hockberg (Allyn and Bacon, Needham Heights, MA).

Sound localization in the presence of one or two distracters

Erno H. A. Langendijk[a)]
TNO Human Factors Research Institute, P.O. Box 23, 3769 ZG Soesterberg, The Netherlands and Waisman Center, University of Wisconsin, Madison, Wisconsin

Doris J. Kistler and Frederic L. Wightman
Waisman Center, University of Wisconsin, Madison, Wisconsin

(Received 3 May 2000; revised 8 November 2000; accepted 23 January 2001)

Localizing a target sound can be a challenge when one or more distracter sounds are present at the same time. This study measured the effect of distracter position on target localization for one distracter (17 positions) and two distracters (21 combinations of 17 positions). Listeners were instructed to point to the apparent position of a train of 30-ms noise bursts, presented at 1 of 85 positions in virtual free field. A harmonic complex and a frequency-swept complex tone served as distracters. The two distracters were turned on 40 and 80 ms after the target onset, had temporal envelopes similar to that of the target, and did not overlap temporally with the target. Virtual sounds were synthesized with individual HRTFs. Localization performance degraded as the number of distracters increased from 0 to 2. When the horizontal distance between target and a single distracter was small (i.e., the interaural differences were almost the same), the influence on the apparent position was greater than when they were far apart. In the vertical dimension, there was not a systematic effect of distracter position on target localizability. However, there was a substantial increase in localization error for targets at high elevations (above 30°) when distracters were present. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1356025]

PACS numbers: 43.66.Qp, 43.66.Pn, 43.66.Yw [DWG]

I. INTRODUCTION

Most previous research on human sound localization has focused on measuring the apparent position of a single source in quiet. This research suggests that the interaural differences in time and intensity of a signal at a listener's two ears are the major cues for the horizontal (lateral) angle of a sound source (Strutt, 1907). Spectral cues, caused by the filtering of a sound by the torso, head, and pinna, provide the cues mainly for the elevation angle of a sound (Butler, 1975). Localization is rather accurate for single broadband sounds. Average errors are about 5° for positions directly in front of the listener (Hofman and Van Opstal, 1998) and up to 20° for more peripheral and rear positions (Oldfield and Parker, 1984; Wightman and Kistler, 1989b; Makous and Middlebrooks, 1990; Carlile *et al.*, 1997).

It remains an open question whether the results obtained from single source localization studies can be generalized to acoustical environments with multiple sources. In everyday life, humans localize sounds in acoustical environments which almost always include more than one sound source. This is also the case with virtual auditory displays, which generate one or more sound sources over headphones. Such displays can be used, for example, to provide pilots and air-traffic controllers with directional information of nearby aircraft or to improve situational awareness of pilots and submarine commanders by presenting an auditory horizon (Bronkhorst *et al.*, 1996; McKinley *et al.*, 1994). In these applications accurate localization of auditory targets in the presence of distracting sounds is essential.

When multiple sources are present, it is important to consider both the localizability and the detectability of the target since detectability can be reduced by the competing sources. Several studies have investigated the audibility of targets in the presence of spatially distributed maskers. Saberi *et al.* (1991) measured detection thresholds for a broadband click-train signal as a function of the angular separation of the signal and a noise masker. Spatial separation of signal and masker sources resulted in a decrease in threshold (relative to spatially coincident sources) of 18 dB in the horizontal (azimuth) dimension and 8 dB in the vertical (elevation) dimension. Gilkey and Good (1995) reported similar results with signal and masker sources that were restricted in bandwidth (low-, mid-, and high-frequency bands). The largest reduction in detection thresholds was observed for high-frequency stimuli (18 dB) and it occurred when the signal and masker were on opposite sides of the head. When the signal and maskers were located on the median plane, the high-frequency stimuli produced improved detection thresholds (9 dB for sources separated by 90°). However, no gain was observed for the other stimuli.

Only a few studies have investigated the influence of one sound source on the localizability of another sound source. For example, Good and Gilkey (1996) asked listeners to judge the apparent position of a click train presented at each of 239 locations surrounding the listener in the presence of a noise masker placed directly in front of the listener. A number of different signal-to-noise ratios (S/N) were tested. They reported an increase in front–back confusions and somewhat poorer elevation performance when the masker source was present, but the effects were large only at low S/N levels. In other studies (Yost *et al.*, 1996; Hawley *et al.*,

[a)]Present address: Philips Research, Professor Holstlaan 4, 5656 AA, Eindhoven, The Netherlands.

1999; Lorenzi *et al.*, 1999; Drullman and Bronkhorst, 2000) localization was investigated using different configurations of target and distracter locations restricted to the frontal horizontal plane. These studies concluded that neither the number of distracters nor the distracter configuration had much effect on localization performance when signal-to-noise ratios were above 0 dB. However, a pilot study by Wightman and Kistler (1997) showed that, for target sounds presented from locations all around the listener, localization performance decreased dramatically with certain distracter configurations.

The present study evaluates the effect of one or two distracters on the localization of virtual auditory targets. Target and distracter positions were uniformly distributed around the listener at elevations between −50° and 90° and over a 360° range in azimuth. Target and distracter stimuli were broadband and scaled to equal A-weighted levels. Each of them created approximately the same spatial percept when presented alone and informal listening convinced us that they were separately identifiable when presented simultaneously. Targets and distracters were gated on and off asynchronously in order to minimize classical masking effects. Under all circumstances listeners reported that they could hear both target and distracter sounds.

II. METHODS

A. Listeners

Five paid listeners (two female, three male) participated in the experiments. Each had normal hearing as verified by audiometric screening and reported no history of hearing problems. The listeners had a small amount of previous experience in sound localization experiments. Data from those experiments suggested that the listeners could localize free-field sounds accurately; their front–back confusion rates were less than 15% and the average localization error (angle between target and response position) was less than 20°.

B. Stimuli

Virtual sources were synthesized by real-time convolution (50 kHz/channel sample rate) of the stimulus with the impulse response of digital filters, one for each ear and each different target position. Each filter was a minimum-phase approximation (Kistler and Wightman, 1992) of the listener's own head-related transfer function (HRTF). The digital filter also compensated for the response of the headphone (Beyer DT990), which was measured on each listener's ears. Details of the synthesis procedures can be found in Wightman and Kistler (1989a). The current HRTF measurement procedure was very similar to that described by Wightman and Kistler (1989a). In the current procedure a 0.95-mm flexible silicone probe tube was inserted in the ear canal with the tip placed very close to the listener's eardrum. A wideband periodic chirplike test sound was then presented from a loudspeaker. The position of the probe tip relative to the eardrum was estimated by an adaptive procedure that searched for the pressure node in the measured response caused by interference between the incoming sound and the sound reflected off the eardrum. A microphone (Etymotic ER7-C) connected to the probe tube recorded the response and a computer averaged 100 responses to boost signal-to-noise ratio. HRTFs were measured simultaneously from both ears at 613 source positions on a 10° azimuth by 10° elevation spherical grid. All measurements from a single listener were completed during a single 30-min session.

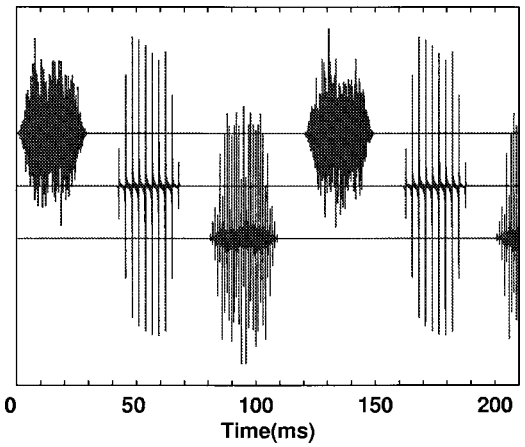


FIG. 1. First 300 ms of the stimuli for the target, Distracter 1 and Distracter 2 (upper, middle, and lower graphs, respectively; stimuli are displaced vertically for better visibility).

All stimuli in this experiment consisted of trains of ten 30-ms bursts (10-ms cosine square on/off ramps) with 90-ms silent intervals between the bursts. Thus each stimulus was a total of 1.2 s long. The specific parameters of each stimulus were chosen to guarantee that the stimuli could be segregated when presented simultaneously. For the target stimulus, each burst was a different sample of white Gaussian noise, bandpass filtered between 200 Hz and 16 kHz. For Distracter 1, each burst was a harmonic complex with a fundamental frequency of 360 Hz and 41 successive equal-amplitude harmonics. For Distracter 2, each burst was a complex tone made up of 30 equal-amplitude components at 500-Hz intervals from 500 Hz to 15 kHz. The frequency of each component was swept by 3000 Hz over the course of the 30-ms burst. The Distracter 1 and 2 stimuli were turned on 40 and 80 ms, respectively, after the target onset and their temporal envelopes were interleaved so that there was no overlap. Target and distracter stimuli were scaled in amplitude such that they generated an A-weighted level of approximately 60 dB at the center of the listener's head. Figure 1 shows the first 210 ms of the target and distracter stimuli.

C. Design

Localization of the target was investigated in three different experimental conditions: a no-distracter condition, a one-distracter condition (Distracter 1 only), and a two-distracter condition. In each condition there were 85 target positions, 5 in each of 17 different spatial regions distributed uniformly on a sphere. The choice of five positions in each region rather than one at the center of the region (which would have been simpler) was an attempt to avoid the response problems which have been reported with small stimulus sets (e.g., Perrett and Noble, 1995). In the no-distracter condition, there were 85 (5×17) unique trials. In the one-

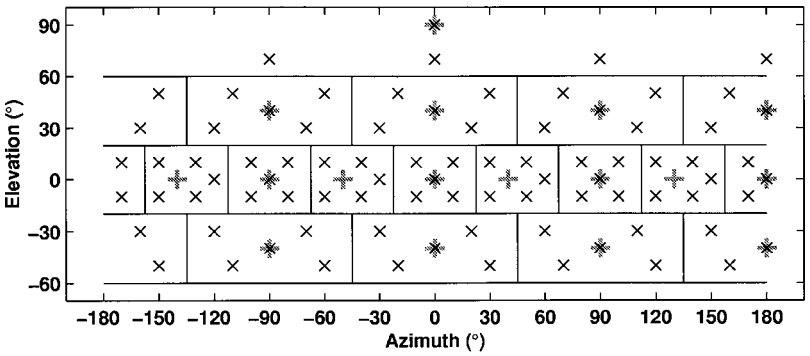


FIG. 2. Target positions (×, $n=85$) and possible distracter positions (+, $n=17$) in the 17 data analysis regions (rectangles).

distracter condition, each of the 85 targets was presented with a distracter presented at a single position selected from each of the 17 regions, resulting in 1445 (85×17) trials. Figure 2 shows the possible target and distracter positions in the 17 spatial regions. In the two-distracter condition, the 85 targets were presented with distracter pairs constructed from 21 combinations of the 17 regions (see Table I), yielding a total of 1785 trials. Trials in each condition were presented in random order. The no-distracter condition was repeated at least 10 times; the one- and two-distracter conditions were repeated two and three times, respectively.

The task for the listener was to point to the apparent position of the target sound with a handheld pointing device (toy gun). An acoustic tracker, developed by the authors (see the Appendix), measured the spatial position of the pointer. The position of the tracker relative to the center of the listener's head determined the perceived direction of the sound. Listeners were instructed to point with a stretched arm, using their left (right) hand for positions on the left (right) side. They were instructed to look at the apparent position of the target when giving a response and could rotate head and body if necessary.

TABLE I. Combinations of distracter positions in the two-distracter condition.

Distracter 1		Distracter 2	
Azimuth	Elevation	Azimuth	Elevation
0	90	−90	40
−90	40	90	40
0	40	180	40
−90	40	−140	0
180	40	−140	0
−90	0	−50	0
0	0	40	0
−90	0	90	0
90	40	130	0
40	0	130	0
0	0	180	0
90	40	−90	−40
−50	0	−90	−40
0	40	0	−40
−50	0	0	−40
90	40	90	−40
0	−40	90	−40
180	40	180	−40
180	0	180	−40
−90	−40	180	−40
90	−40	180	−40

In order to test the accuracy of the pointing method, listeners pointed to each of 17 visual markers on the walls of the room. The markers were placed in the horizontal, median, and lateral vertical planes with elevations ranging from −45° to 90° and a 360° range in azimuth. The average pointing error (angle between marker direction and response direction), collapsed across marker positions and listeners, was about 1.5°. Errors were within 3° for all listeners and all marker positions.

The experiment was carried out in a quiet room (A-weighted ambient noise level below 35 dB) with dimensions $3\times4\times3$ m and fluorescent tube lighting. Listeners were seated in a rotatable chair with the center of their heads in the middle of the room (a laser was used to check this); they were instructed to look straight ahead during stimulus presentation. The entire experiment was divided into ten 2-h sessions. A session started with calibration of the position tracker. Next, the listener was tested either one or two times on the no-distracter condition. Finally, the listener was tested in one of the distracter conditions, completing about 800–1000 trials. Short breaks were provided every 250 trials. On average listeners could complete 100 trials in about 4–5 min.

Localization was also investigated in five control conditions. In the first control condition, listeners localized a 250-ms noise burst (20-ms cosine squared on/off ramps) presented from 1 of 144 different target locations in an anechoic chamber. Listeners responded by calling out azimuth and elevation coordinates of the perceived sound source position. The second control condition was similar to the first except that stimuli were synthesized with individualized HRTFs and were presented over compensated headphones. These first two control conditions were identical to conditions described in earlier studies from this laboratory (Wightman and Kistler, 1989b) and used the same methods to synthesize the virtual sound sources. The purpose of these control conditions was to evaluate the adequacy of the virtual stimulus synthesis procedures. In the third control condition, intended as a test of the utility of the pointer method, listeners used the pointer method to indicate the apparent direction of the same virtual targets that were presented in the second control condition. The fourth and fifth control conditions were included to assess the localizability of the two-distracter stimuli. Listeners used the pointer method to indicate the apparent directions of the harmonic complex (Distracter 1) and the complex sweep (Distracter 2) which were presented

at the same 85 virtual target positions as were used in the main experiment.

D. Data analysis

As in many previous sound localization studies (Wightman and Kistler, 1989b; Makous and Middlebrooks, 1990) listeners made localization errors that we identified as ''front–back confusions.'' These are responses that are in the hemifield (front or rear) opposite the target. In the no-distracter condition and in all five control conditions the percentage of front–back confusions was relatively small (about 5%–10%) and when the confused responses were resolved (i.e., reflected about the lateral-vertical plane), the data formed a compact cloud of positions in the vicinity of the target position. These data were fit with a Kent distribution, a spherical analogue of the elliptical two-dimensional normal distribution (Kent, 1982; see also Carlile *et al.*, 1997), and performance measures for the no-distracter and control conditions were derived from the Kent fits.

In the one- and two-distracter conditions, the number of front–back confusions was significantly larger than in the no-distracter and control conditions. Other types of errors also occurred in the distracter conditions, such as ''attraction'' errors and responses in the region of the distracter, which resulted in multi-modal response distributions. Since these errors were frequent and sometimes hard to distinguish from normal response variability, it seemed inappropriate to describe the data from the one-and two-distracter conditions with the unimodal Kent distribution. To address this problem, the data from the distracter conditions were analyzed in terms of an error score. This score represented the deviation of each judgment (in z-score form) from the centroid of all the judgments to the same target in the no-distracter condition.

III. RESULTS

A. No-distracter and control conditions

Performance in the no-distracter and control conditions (all of which involved localization of a single source, presented in isolation) was analyzed on a listener-by-listener and condition-by-condition basis. First, the percentage of front–back confusions was calculated and then the confusions were resolved by reflecting the response about the front–back plane. A confusion was said to have occurred if reflecting a response to a particular target about the front–back plane reduced the localization error by more than 15°. Next, the responses to all targets in each spatial region were aggregated to form an overall response pattern for that region. Aggregation facilitated comparisons across the various conditions since the first three control conditions tested different spatial positions and different numbers of trials at each position. In the no-distracter condition and in the two control conditions (#4 and #5) involving the gated complex tone stimuli, there were five targets in each region. In the other control conditions there were 4–14 targets in each region. To aggregate the responses, each target-response pair was rotated such that the target position coincided with the center of the region containing the target (see Fig. 2). Then, for each center position, all the responses were fitted with a Kent distribution. A one-standard-deviation (1 s.d.) ellipse around the mean response was estimated from the parameters of the Kent distribution. Figure 3 shows an example of response data and the corresponding 1 s.d. ellipse. The angle between the center of the region and the center of the corresponding Kent distribution was defined as a measure of the systematic error and the length of the major axis of the ellipse was a measure of the dispersion of the data.

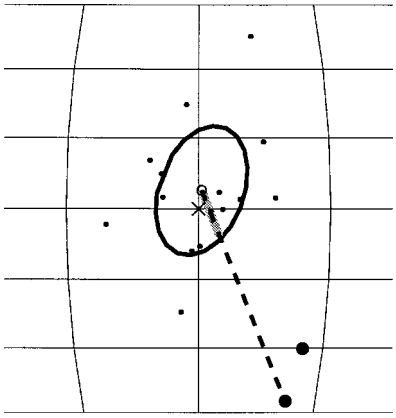


FIG. 3. Illustration of the method of data display and of the calculation of the error score. The figure shows individual responses to a target at 180° azimuth and 0° elevation (indicated by the ×) in the no-distracter condition (·) and in the one-distracter condition (●) with a distracter at 180° azimuth and −40° elevation (just outside the figure). The ellipse represents the one standard deviation contour around the mean response (○) for the no-distracter data. The dashed line connects a response in the one-distracter condition with the mean response in the no-distracter condition. The shaded segment of the line represents the segment of the dashed line that falls within the ellipse. The error score is equal to the length of the dashed line divided by the length of the shaded segment.

Differences among the results from the no-distracter and

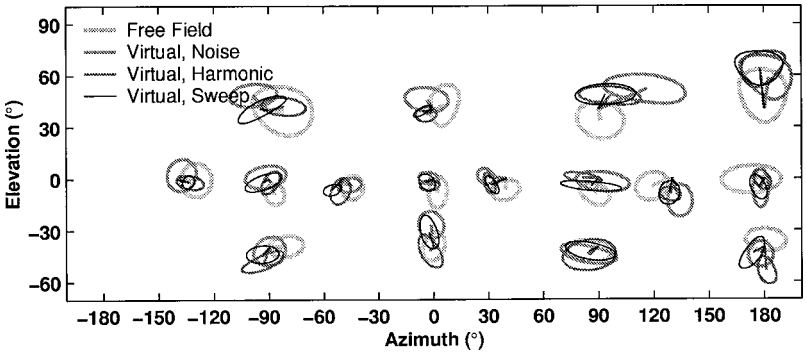


FIG. 4. Illustration of the response distributions in the no-distracter condition (Virtual, Noise) and in three of the control conditions (#1 = Free Field, #4 = Virtual, Harmonic, #5 = Virtual, Sweep) for one listener (SRW). His results are typical for the three best localizers. Lines are drawn from the target position to the mean response. The contours represent 1 s.d. ellipses on a sphere.

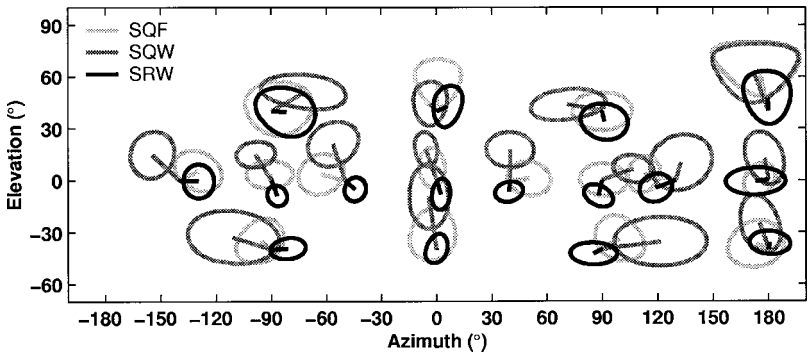


FIG. 5. Illustration of the intersubject variability in the free-field control condition (#1) for three listeners (SQF, SQW, and SRW). The results for the other two listeners (SRQ and SRS) are most similar to those of SRW. Lines are drawn from the target position to the mean response. The contours represent 1 s.d. ellipses on a sphere.

control conditions were generally quite small, as confirmed by inspection of the 1 s.d. ellipse plots. This was expected since all of these conditions involve localization of a single wide-band source presented in isolation. Figure 4 illustrates the similarity for a subset of the conditions for one representative listener. Between-listener variability in dispersion and in the size and direction of the systematic localization error were noticeable for certain target regions. Figure 5 illustrates the intersubject variability for three listeners in the free-field condition.

Figure 6 shows means (across target regions) and 95% confidence intervals for systematic error, dispersion and percentage of front–back confusions from the five control conditions and the no-distracter condition. Each column shows the results from a single listener. Very few significant differences are revealed. Some general observations are: (1) There were no significant differences in the three performance measures between the no-distracter condition and the two control conditions (#4 and #5) in which listeners localized the harmonic complex and the swept complex tone in isolation. (2) Although there were some statistically significant differences in the average dispersion, these differences were small both between conditions and between listeners. (3) For two listeners (SQF, SRS) the percentage of front–back confusions in the virtual verbal and the virtual pointer conditions was somewhat larger than in the free-field condition. In general, the statistical analysis confirmed the conclusions drawn from the graphical analysis, namely that the virtual sources were generated appropriately, the pointer response method was at least as accurate as the verbal response method, and the spatial percepts formed by all three experimental stimuli were roughly similar.

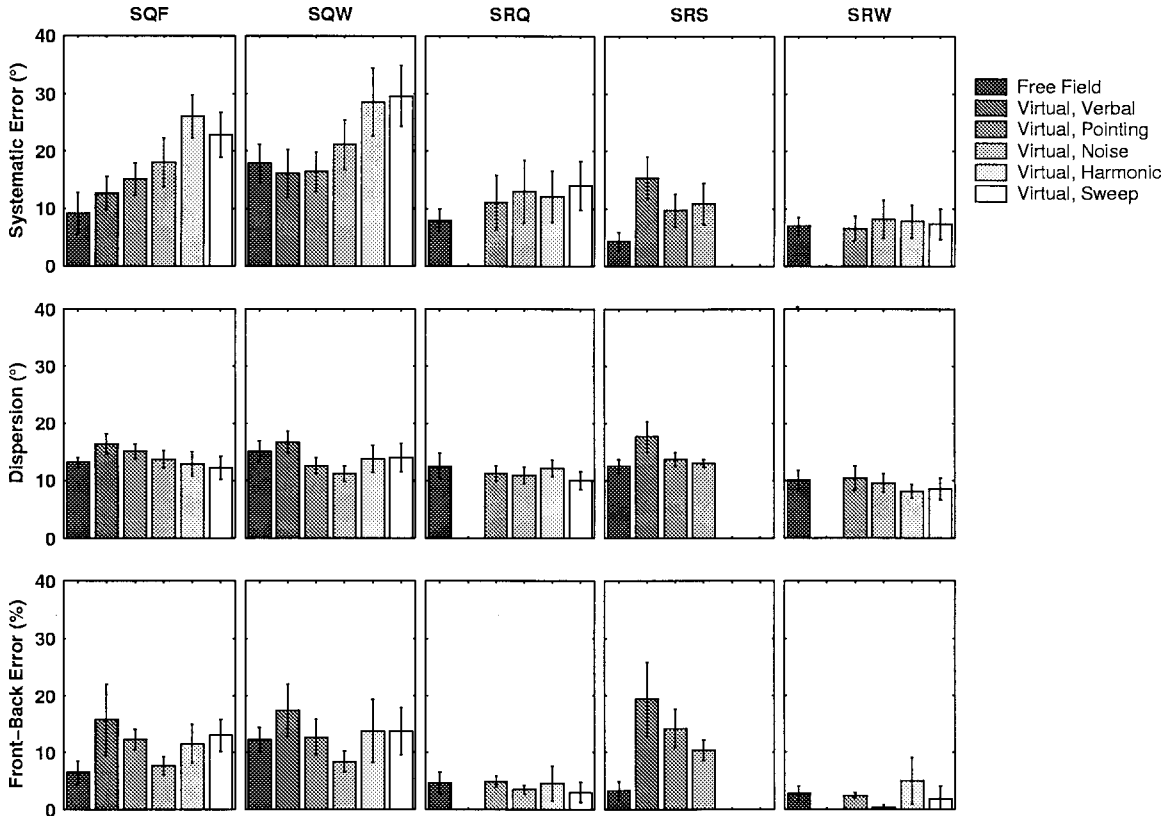


FIG. 6. Means (across target regions) and 95% confidence intervals for systematic error, dispersion and percentage front–back confusions from all five control conditions (#1 = Free Field, #2 = Virtual, Verbal, #3 = Virtual, Pointer, #4 = Virtual, Harmonic, #5 = Virtual, Sweep) and the no-distracter condition (Virtual, Noise). Each column shows the results of a single listener. Note that results for some conditions for some listeners are missing (blank spaces) due to (un)availability of the listener.

B. Distracter conditions

Whereas the distributions of responses in the no-distracter and control conditions were predominantly unimodal, response distributions in the one- and two-distracter conditions were multi-modal due to an increase in front–back confusions and the occurrence of other types of errors. Consequently a different approach to quantifying performance in the distracter conditions was necessary. First a ''global'' measure of error was estimated which enabled a comparison of performance in the distracter conditions to performance in the no-distracter condition. Second the various types of errors were identified and quantified.

The global error was estimated by assigning each judgment in the one- and two-distracter conditions an error score. This error score was comparable to a *z*-score in that it was a deviation score divided by a standard deviation. The deviation score was defined as the angular disparity between the judgment and the mean judgment (centroid) to the same target in the no-distracter condition. The standard deviation was estimated from the parameters of the Kent distribution fitted to the no-distracter data. It was defined as the length of the segment from the centroid to the one standard deviation contour in the direction of the target. Figure 3 illustrates the calculation of the error score. The centroid and standard deviation of the judgments in the no-distracter condition were estimated from data in which front–back confusions had been resolved. However, in order to compare the results from the distracter conditions with those from the no-distracter condition, error scores were also computed for the ''unresolved'' judgments in the no-distracter condition.

In order to facilitate analysis of data from the two distracter conditions, target and distracter positions were grouped into spatial ''areas'' to highlight performance in the Left–Right, Front–Back, and Up–Down dimensions (Kistler and Wightman, 1992). These areas were called Left–Middle–Right, Front–Middle–Back, and Up–Middle–Down areas, respectively. The ''Middle'' areas included all targets within 30° of the median, lateral-vertical, and horizontal planes, for the Left–Right, Front–Back, and Up–Down dimensions, respectively. The remaining areas included all targets beyond this ±30° strip, as appropriate. Figure 7 shows a histogram of a single listener's error scores in the no-distracter condition and in the one- and two-distracter conditions for all target and distracter positions in the Left area. In the distracter conditions more than half of the error scores are larger than 2 (that is, responses were more than 2 standard deviations from the mean response in the no-distracter data) and some error scores are as large as 10. Because of the skewed distributions, a log-transformation was applied to the error scores before computation of the mean and the 95% confidence intervals for each listener for each area. The mean error scores are noted in the legend.

1. Effect of one distracter

The presence of a single distracter substantially degraded listeners' abilities to localize the target sound. For all target positions error scores were larger than in the no-distracter condition, but the size of this effect depended heavily on the target and distracter positions, as illustrated in Fig. 8. An inspection of the error score means and 95% confidence intervals across regions and conditions revealed a similar pattern of results for each of the five listeners, although listeners may have differed slightly in their overall error rates. Because the pattern of results was comparable across listeners, the individual means and standard deviations for each condition and area were averaged across listeners. Significant differences between means is indicated by nonoverlap of the confidence intervals. This figure shows mean error scores for target and distracter positions in the Left–Middle–Right, (upper panel), Front–Middle–Back (middle panel), and Up–Middle–Down (lower panel) areas along with the mean error scores from the no-distracter condition. The upper panel of the figure shows that error scores were considerably larger when target and distracter were in the same lateral area than when they were in different lateral areas. This effect is most prominent when target and distracter were on the same side of the listener (Left or Right area) but also exists for the Middle lateral area.

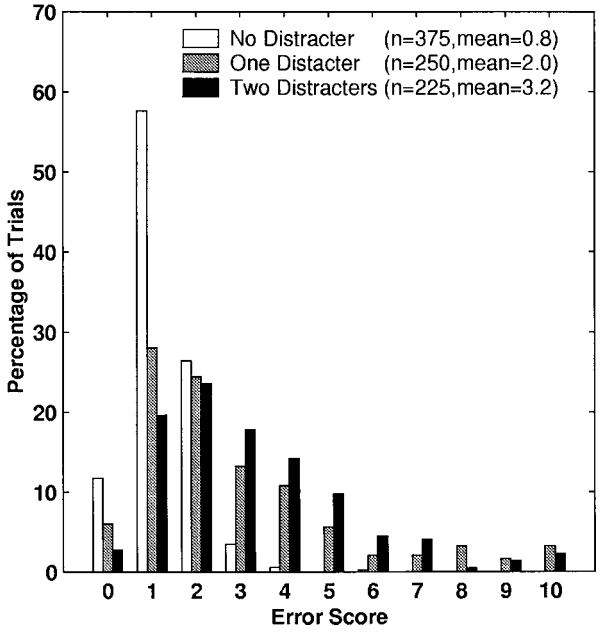


FIG. 7. Histogram of error scores for one listener (SRQ) in the no-distracter condition and in the one- and two-distracter conditions with targets and distracters in the Left area. Note that because of the skewed distributions, a log-transformation was applied before calculation of the mean.

Interestingly, the distracter effect was less dependent on target and distracter position for positions grouped along the other two dimensions. For the Front–Back dimension there were no statistically significant differences in mean error scores between all possible pairs of target and distracter areas. In contrast, for the Up–Down dimension, there was an effect of target position, which is illustrated in the lower panel of Fig. 8; targets in the upper area were affected most by the distracter.

2. Effect of two distracters

Two distracters had greater impact on sound localization ability than did one distracter. Figure 9 summarizes the results from the two-distracter conditions. The data suggest that the distracter effects are in some sense additive. Com-

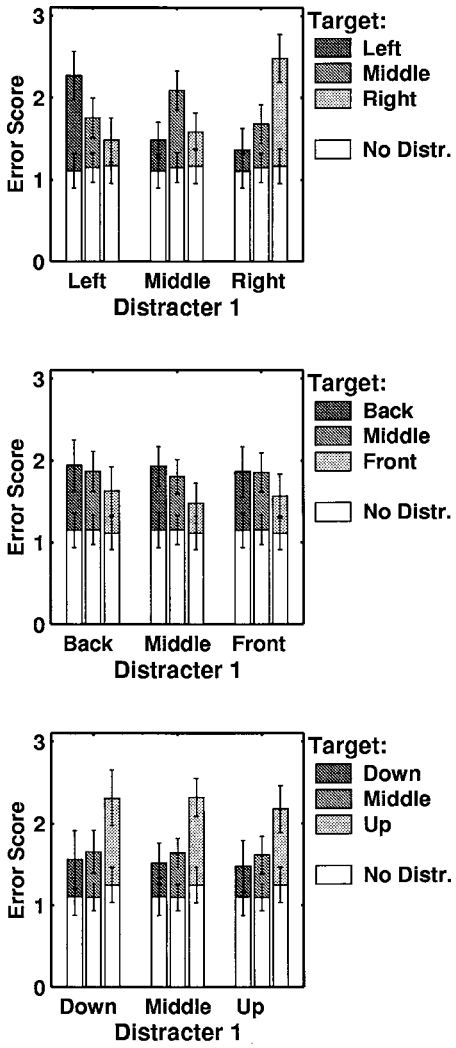


FIG. 8. Mean error scores and 95% confidence intervals (averaged over listeners) from the one-distracter condition and from the no-distracter condition. Different panels show the results for the Left–Right, Front–Back, and Up–Down dimensions.

pare, for example, the upper panel of Fig. 8 (showing the effect of Distracter 1) with the upper-left panel of Fig. 9 (showing the effect of Distracter 1 with Distracter 2 on the left side). The presence of Distracter 2 in the Left area resulted in a significant increase in error scores for targets in the same (left) area, whereas localization performance for targets in the other two areas was only mildly affected. Similar effects occur for Distracter 2 in the Middle and Right areas. An exception occurs when the two distracters were on opposite sides of the target. In that case error scores for targets in the Middle area also increased.

For the Up–Down dimension the effect of two distracters was similar to that of a single distracter; targets in the Up area were more difficult to localize than targets in the other two areas. There were a few statistically significant mean differences in error scores in the Front–Back dimension, but these differences were rather small compared to the differences found in the other two dimensions.

In Fig. 10 the results from the Left–Middle–Right area are replotted in order to demonstrate the effect of having distracters in the same area as the target. It appears that error score increases with the number of distracters in the same target area. Furthermore, the difference between the one-and two-distracter conditions was relatively small when there was an equal number of distracters in the target area. The presence of distracters always resulted in a significant increase of the error score, even when the distracters were outside the target area.

3. Types of errors

Investigation of the response patterns for each target-distracter configuration revealed a number of different types of localization errors. Some types of errors, such as response bias, dispersion, and front–back confusions, occurred in the no-distracter condition as well. Other types of errors occurred only in the distracter conditions, such as target-distracter confusions and attraction of responses toward the distracter. These errors sometimes produced bimodal response patterns in which one subset of the responses was close to the target position and the other close to the distracter position. In order to quantify the relative occurrence of specific types of errors and their magnitudes, the bimodal data were separated from the unimodal data.

In order to isolate the bimodal data in the distracter conditions, the following procedure was applied to the responses to the targets in each of the 17 regions. First, the major and minor axes of the entire response distribution (for all targets in a given region) were determined by fitting the entire data set with a Kent distribution. Then, the responses were divided into two sets. One set consisted of all responses on one side of the minor axis, and the other set consisted of all responses on the other side of the minor axis. Finally, a Kent distribution was fitted to each of the two data sets and 95% confidence ellipses were computed from the parameters of the Kent distributions. If the 95% confidence ellipses for the two sets did not overlap, the data were judged to be bimodal. Figure 11 shows an example of a bimodal response distribution for one condition for one listener.

The outcome of the analysis revealed that in about 8% of the one-distracter conditions the response distributions were bimodal. In those cases the confidence ellipse of one of the modes (Mode 1) overlapped with the target position and the ellipse of the other mode (Mode 2) either overlapped with the distracter position (40% of the cases) or it was within 15° of the line segment from target position to distracter position (60% of the cases). This indicates that for these data there was attraction to or confusion with the distracter.

The results in Table II indicate that in 55% of the bimodal cases the target was in an area above the distracter and in only 14% of the cases it was below the distracter. Further, in most cases (69%) the target and distracter were in the same lateral area. There were slightly more bimodal distributions in cases where the target was in the rear and the distracter in the front than vice versa. These results are in line with those from the error score analysis (cf. Fig. 8).

The bimodal analysis is, in principle, suitable to describe the response distribution when listeners make front–back confusions for a subset of their responses. In the no-distracter condition, however, listeners made very few confusions per

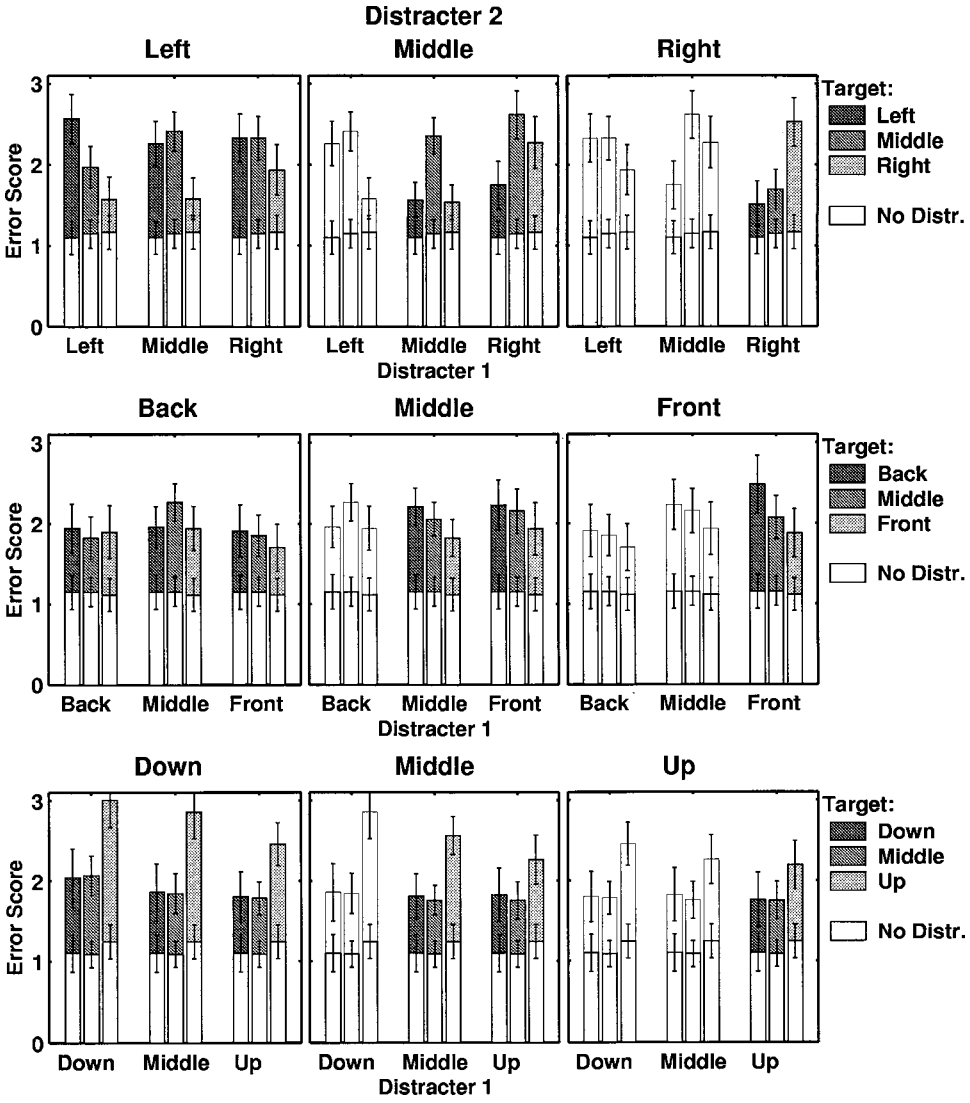


FIG. 9. Mean error scores and 95% confidence intervals (averaged over listeners) from the two-distracter condition and from the no-distracter condition. Different panels show the results for the Left–Right, Front–Back, and Up–Down dimensions. Since some spatial configurations of the distracters are redundant if we assume the distracter stimuli are interchangeable, no data were collected for these configurations. For example, the configuration Distracter 1–Distracter 2 in Left–Middle can be considered identical to Middle–Left. The thin-lined white bars contain such redundant information and were reproduced from the appropriate distracter configurations to facilitate comparisons.

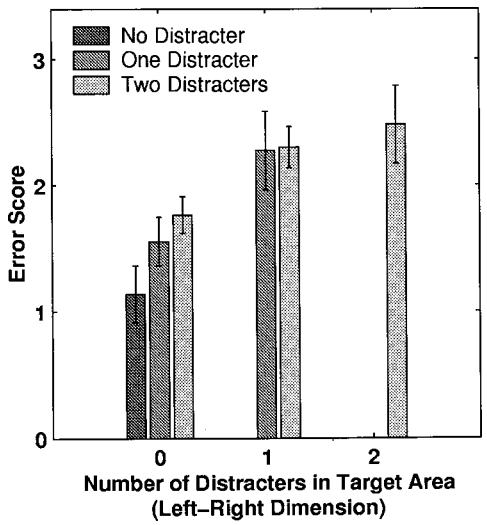


FIG. 10. Effect of the number of distracters in the same lateral area (Left–Middle–Right) as the target for the no-distracter condition and for the one- and two-distracter conditions. Mean error scores and 95% confidence intervals were collapsed across listeners.

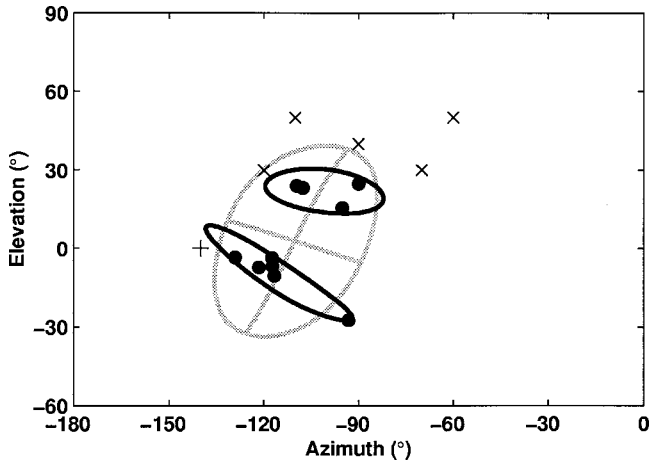


FIG. 11. Illustration of a bimodal response distribution in the one-distracter condition. Target, distracter, and response position(s) are represented with ×, +, and ●, respectively. The gray ellipse represents the 95% confidence region fitted to all ten responses; the major and minor axis of the ellipse are also shown (long and short gray lines, respectively). To test for bimodality the responses were divided into two sets, one on each side of the minor axis. The 95% confidence ellipses were fitted to each set (black ellipses). If the two black ellipses did not overlap the data were judged to be bimodal.

TABLE II. Distribution of bimodal occurrences across relative target and distracter position for the Left–Right, Front–Back, and Up–Down dimensions in the one-distracter condition. *T* and *D* are the target and distracter areas in a given dimension, where Left/Back/Down<Middle <Right/Front/Up. Note that the data in the present table only concern the target-distracter configurations with bimodal response distributions (i.e., 23 out of the all 289 target-distracter configurations).

Dimension	$T<D$	$T=D$	$T>D$
Left–Middle–Right	18%	69%	13%
Back–Middle–Front	44%	32%	24%
Down–Middle–Up	14%	31%	55%

region, which explains why bimodality was not detected in the analysis of the no-distracter data (cf. Table III). In the one-distracter condition about 28% of the bimodal cases had one mode in the Front area and the other in the Back area or vice versa. In virtually all of these cases the ellipse of one of the two modes overlapped with the target and the ellipse of the other mode overlapped with the distracter. When the confused mode was mirrored in the lateral-vertical plane only 5% of the bimodal cases had both modes less than 30° apart. This suggests that the presence of a distracter does not increase the conventional type of front–back confusion, that is, front–back confusions caused by the similarity in the HRTFs of two positions. On the contrary, it appears that the increase in front–back confusions when a distracter is present (cf. Table III) is almost always a result of a confusion with the distracter.

For the unimodal response distributions and for each of the two modes of the bimodal distributions, the systematic error and the dispersion were calculated. The results, collapsed across listeners and distributions, are shown in Table III. The average systematic error in the one-distracter condition for the unimodal distributions was not significantly different from that in the no-distracter condition. For the first mode of the bimodal distributions, it was slightly larger. The average systematic error for Mode 2 was much larger, because the second mode was in general close to the distracter. In fact it is roughly equal to the bimodal target-distracter distance. The average dispersions in the one-distracter condition for the unimodal and in the first modes of the bimodal distributions were only slightly larger than in the no-distracter condition. For the second modes of the bimodal distributions, it was somewhat larger.

TABLE III. Bimodal, unimodal and conventional localization errors for the no-, one-, and two-distracter conditions. An open space (-) means that it was not appropriate to calculate the error (i.e., not enough data available and/or trimodal data expected). For the bimodal distributions Mode 1 and 2 represent the modes closest to the target and distracter position, respectively. The bimodal target-distracter distance is the average target-distracter distance across all bimodal cases. Bimodal and unimodal errors are averaged across all bimodal and unimodal occurrences, respectively. Conventional errors are averaged across all individual responses.

Error type	No-Distracter	One-Distracter	Two-Distracter
Total bimodal	0%	8%	-
Bimodal systematic error			
Mode 1	-	21°	-
Mode 2	-	79°	-
Bimodal target-distracter distance	-	80°	-
Bimodal dispersion			
Mode 1	-	20°	-
Mode 2	-	28°	-
Total unimodal	100%	92%	-
Unimodal systematic error	16°	16°	-
Unimodal dispersion	17°	20°	-
Convention errors			
Front–back confusions	6%	9%	11%
Absolute localization error	20°	24°	28°

Given the results from the one-distracter condition, it might be expected that, in the two-distracter condition, a trimodal distribution with modes near the target and both distracter positions would best describe the data in the most difficult conditions (e.g., with target and both distracters in the same Left–Right area). Due to the limited number of trials, however, such a distribution could not be fitted to the data. Inspection of the raw data revealed that attraction of a part of the responses to one or both distracters did occur, suggesting that the data could indeed be described with a trimodal distribution in some cases.

IV. DISCUSSION

The results of the present study show that localization performance deteriorated as the number of distracters increased from 0 to 2. Poorer localization performance was revealed by a combination of errors, including dispersion, front–back confusions, and attraction to or confusion with the distracters. The disruptive effect of the distracter depended on target and distracter position. When the horizontal distance between target and distracter was small (i.e., small differences in interaural time/intensity cues), the distracter had more impact on the apparent position than when target and distracter were far apart. The error scores were largest for targets at elevations above 30° but the increase in error score caused by the distracter also occurred for targets at lower elevations.

There are few studies in the literature with which these data can be compared. Good and Gilkey (1996) investigated localization of a pulse train in the presence of a single distracter (a broadband noise) positioned directly in front of the listener. In contrast to the present study, they reported a substantial increase in front–back confusions when the distracter was present. There are a number of methodological differences that could explain this discrepancy. For example, Good and Gilkey used a continuous distracter whereas pulsed distracters were used in the present study. A continuous masker could have (and almost certainly did, in the low S/N-ratio conditions) reduced the detectability of the target. Given the relatively large number of front–back confusions in the data from their no-distracter condition (20%–40%), it could also be that their listeners were more susceptible to making front–back confusions. It should be noted that, in previous pilot experiments with continuous distracters (Wightman and Kistler, 1997), front–back confusion rate did not increase.

In other studies (Yost *et al.*, 1996; Hawley *et al.*, 1999; Drullman and Bronkhorst, 2000) target and distracter stimuli were speech signals presented from sources in the frontal horizontal plane. All three of these studies reported either no effect or a relatively small effect of distracters ($n=1-4$) and no significant interactions between target and distracter position. Lorenzi *et al.* (1999) studied the effects of signal-to-noise ratio and distracter location in the frontal horizontal plane using clicks as a target and noise as a distracter. They found that localization accuracy decreased with decreasing signal-to-noise ratios for ratios below 0 dB and related this to the decreasing detectability of the signal in the ear ipsilateral to the noise. For signal-to-noise ratios above 0 dB (i.e., conditions which are more comparable to our conditions), localization performance did not decrease. These findings are in good agreement with those of the present study when results for similar conditions are compared (i.e., for target and distracters positions with azimuths between $-90°$ and 90° and elevations between $-10°$ and 10°).

In contrast to the research cited above, the current study examined a large number of target-distracter spatial configurations and used distracters specifically designed to minimize classical masking effects. There were considerable distracter effects for targets above the horizontal plane when target and distracter(s) were in the same Left–Right region. Interestingly this effect was asymmetric; a target at a high elevation ($>30°$) presented with a distracter at a low elevation ($<-30°$) was more difficult to localize than a target at a low elevation presented with a distracter at a high elevation. Note that this effect cannot be explained by the generally poorer localization performance for high elevation targets when no distracters are present (cf. no-distracter results) since error scores in the distracter conditions were expressed relative to those in the no-distracter condition. A model of sound localization based on spectral correlation of individual HRTFs may explain some of the distracter effects we obtained. Similar models have been used to explain localization of narrowband sounds (Middlebrooks, 1992) and spectro-temporal effects of sound localization (Hofman and Van Opstal, 1998). The basic outline of such a model and preliminary tests in some of the conditions of this experiment are offered in the next section.

A. Spectral correlation model

The spectral correlation model outlined here is based on the assumption that a signal arriving at the eardrum is processed by the auditory system in 3–5 ms time bins in the log-frequency/log-amplitude domain (Viemeister and Wakefield, 1991). Further, it is assumed that the auditory system integrates the spectra of these time bins over a period of about 150 ms in order to obtain a stable estimate of the spectrum of the incoming sound. This estimate is then cross-correlated with a set of internal templates (one for each ear and each direction). The extent to which a template correlates with the signal spectrum determines the probability that the position corresponding to the template will be reported.

Because the stimuli in our experiment were broadband and had nominally flat spectra, the inputs to the model for our preliminary tests were represented by the listener's left and right ear HRTFs. The templates were represented by the HRTFs of both ears for all source directions. There was no attempt to include the additional filtering effects of the middle ear, nor were the binaural cues (interaural time and intensity differences) implemented in the model.

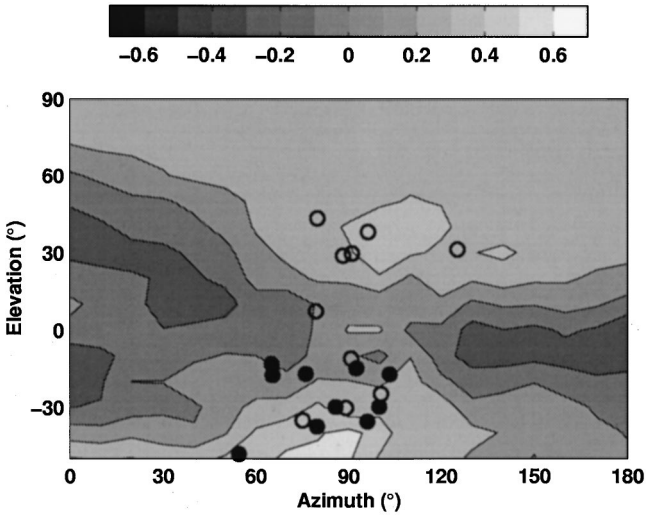


FIG. 12. Model prediction of the response distribution for a target in the one-distracter condition with one source in the upper-right region at 90° azimuth and 40° elevation and the other source in the lower-right region at 90° azimuth and $-40°$ elevation. The model prediction was for a single listener. Shading codes the probability of a response. The figure also shows the listener's responses in these conditions. The circles (○) represent localization responses for targets in the upper-right region with a distracter in the lower-right region. The bullets (●) represent localization responses for targets in the lower-right region with a distracter in the upper-right region.

The predicted probability of a response at a specific position in a given target-distracter configuration was computed in the following way. First, HRTFs (for each ear separately) were smoothed by a 1/12-octave filterbank in the frequency range from 0.5 to 16 kHz. The log of each 1/12 octave filter output was recorded. Then, the smoothed HRTFs from the target and distracter positions were summed, representing the integrated signal spectrum. Finally, this sum spectrum was cross-correlated with each individual smoothed HRTF (i.e., the internal templates). The cross-correlation procedure gave numbers between -1 and $+1$ which were interpreted as low and high response probability, respectively. For the positions in the Left and Right area, the cross-correlation coefficients for the left and right ear were taken individually. For positions in the Middle lateral area, the cross-correlation coefficients were averaged across both ears.

Figure 12 shows an example of the predictions of the model for one listener in the one-distracter condition with one source in the upper-right region with center position 90° azimuth and 40° elevation and the other source in the lower-right region with center position 90° azimuth and $-40°$ elevation. The figure also shows the listener's responses in these conditions. Note that there are ten different one-distracter conditions with the target and distracter in those regions (cf. Fig. 2). In order to compare the response distribution with the model's prediction, the output of the model was averaged across all ten conditions (the predictions of the

model were only slightly different for each of those conditions).

The figure illustrates a typical prediction of the model for conditions with target and distracter in the same lateral area. Note that a bimodal response distribution is predicted with one mode near the target position and the other near the distracter position. This is not self-evident. The prediction is based on the ''sum'' spectrum, and one might expect the sum to correlate highly with the HRTF at a position midway between the target and distracter. The model also predicts a larger number of responses near the lower sound source position than near the higher one. The figure shows that this is supported by the data when the total response distribution is considered.

In its present form the model does not distinguish between target and distracter source. Hence, it predicts the same response distribution independent of which sound source is the target and which is the distracter. The listener, however, is obviously able to make this distinction in most cases. The data show that when the target is in the lower region, almost all responses are in the lower region, but when the target is in the higher region, up to half of the responses go to the lower region while the other half go to the higher region. The model would make no distinction between the two cases. This indicates that there must be an additional process that favors the actual position of the target.

For conditions with target and distracters in different lateral areas, localization performance was only mildly affected. The model, however, predicted up–down confusions in some of these conditions that did not actually occur. Alternatively, it is not unlikely that the auditory system weights the spectrum of each 3–5 ms time bin before doing the summation (otherwise, silent intervals in the signal having a spectrum related to the spontaneous firing of the nerve fibers would disturb the sum spectrum). Hence, in these conditions, there will always be one source closer to the left (right) ear, dominating the sum spectrum. Similar arguments can be given to explain the two-distracter configurations, with one general exception: when the target is in the Middle area with a distracter on either side. In that case the sum spectrum for the left and right ear will be dominated by the spectrum of the distracter in the Left and Right area, respectively. Hence, the error score in these conditions will be larger than, for example, when both distracters are at the same side of the listeners. This conclusion is supported by the data (cf. Fig. 9).

Only a rough outline of the model is presented here. While the preliminary results are suggestive, it is clear that considerable additional work will be necessary to evaluate its potential more fully. That work is currently underway.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the assistance of Shelly Godar, Timothy Read, and Douglas Swiggum in the technical phases of the work and the very helpful editorial comments of Adelbert Bronkhorst. Financial support was provided by ONR (N00014096-1-0675), NIH-NIDCD (P01 DC00116-23), and NASA (Cooperative Agreement No. NCC2-542).

APPENDIX: ACOUSTIC POSITION TRACKING DEVICE

An acoustic position tracking device was developed by the authors to measure the position of a sensor in space. This device consisted of five loudspeakers (Cambridge Soundworks Center/Surround IV speakers, one in each corner of the room, 1 m above the floor and 1.3 m from the center of the room, and one at the ceiling 2.5 m above the floor in the center of the room), a microphone (custom Etymotic Research probe microphone) that acted as the sensor, two amplifiers (Crown D60), an AD/DA converter (Tucker Davis Technologies Power DAC) a signal switcher (Tucker Davis Technologies SS1 Signal Switcher), and a personal computer. On a press of a button a 5-ms chirp (1–15 kHz) was played sequentially from speaker 1 to 5 and recorded with the microphone. The input–output delay was calculated using a cross-correlation technique and a correction was made for the internal delays (AD/DA converter, amplifiers). The distance from each loudspeaker to the sensor was calculated by multiplying the delay by the speed of sound (343 m/s at 20 °C). Then, if the positions of three loudspeakers (L_i) and the distances between these loudspeakers and the sensor (d_i) are known, the position of the sensor (P) can be calculated using the following equations:

$$P_x = \frac{d_1^2 - d_2^2 + L_{2,x}^2}{2L_{2,x}},$$

$$P_y = \frac{d_1^2 - d_3^2 + L_{3,y}^2 + L_{3,x}^2 - 2L_{3,x}P_x}{2L_{3,y}},$$

$$P_z = \pm\sqrt{d_1^2 - P_x^2 - P_y^2},$$

where the coordinate system has its origin at L_1, the x-axis through L_1 and L_2 and the y-axis perpendicular to the x-axis and in the plane spanned by the three loudspeakers. Note that because only three loudspeakers were used in the calculation the sign of P_z is unknown, but this sign can be easily determined by using the distance between the sensor and a fourth loudspeaker with known position.

In the present experiment five loudspeakers were placed such that, when the listener gave a response, the sound paths from the sensor to the four nearest loudspeakers would never be obstructed by the listener. In order to improve the position estimate, the sensor position was calculated for each combination of three loudspeakers (plus a fourth to determine the sign of the z-coordinate) and averaged across all combinations of the nearest four loudspeakers.

The accuracy of the tracking device was tested with a mechanical position tracker (3D-Stylus) with high accuracy (error <0.1 cm). Because the mechanical tracker had a limited range (radius <0.5 m) positions were tested with the base of the stylus at seven different positions. For each base position the tracker was tested for 50 positions chosen randomly. The accuracy of the acoustic tracking device was within 1 cm for all positions tested.

Bronkhorst, A. W., Veltman, J. A., and van Breda, L. (**1996**). ''Application of a three-dimensional auditory display in a flight task,'' Hum. Factors **38**, 23–33.
Butler, R. A. (**1975**). *The Influence of the External and Middle Ear on Auditory Discriminations* (Springer, Berlin).
Carlile, S., Leong, P., and Hyams, S. (**1997**). ''The nature and distribution of errors in sound localization by human listeners,'' Hear. Res. **114**, 179–196.
Drullman, R., and Bronkhorst, A. W. (**2000**). ''Multichannel speech intelligibility and talker recognition using monaural, binaural, and three-dimensional auditory presentation,'' J. Acoust. Soc. Am. **107**, 2224–2235.
Gilkey, R. H., and Good, M. D. (**1995**). ''Effects of frequency on free-field masking,'' Hum. Factors **37**, 835–843.
Good, M. D., and Gilkey, R. H. (**1996**). ''Sound localization in noise: The effect of signal to noise ratio,'' J. Acoust. Soc. Am. **99**, 1108–1117.
Hawley, M. L., Litovsky, R. Y., and Colburn, H. S. (**1999**). ''Speech intelligibility and localization in a multi-source environment,'' J. Acoust. Soc. Am. **105**, 3436–3448.
Hofman, P. M., and Van Opstal, A. J. (**1998**). ''Spectro-temporal factors in two-dimensional human sound localization,'' J. Acoust. Soc. Am. **103**, 2634–2648.
Kent, J. T. (**1982**). ''The Fisher–Bingham distribution on the sphere,'' J. R. Statist. Soc. B **44**, 71–80.
Kistler, D. J., and Wightman, F. L. (**1992**). ''A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction,'' J. Acoust. Soc. Am. **91**, 1637–1647.
Lorenzi, C., Gatehouse, S., and Lever, C. (**1999**). ''Sound localization in noise in normal-hearing listeners,'' J. Acoust. Soc. Am. **105**, 1810–1820.
Makous, J., and Middlebrooks, J. (**1990**). ''Two-dimensional sound localization by human listeners,'' J. Acoust. Soc. Am. **87**, 2188–2200.
McKinley, R. L., Erickson, M. A., and D'Angelo, W. R. (**1994**). ''3-dimensional auditory displays: Development, applications, and performance,'' Aviat., Space Environ. Med. **65**, 31–38.
Middlebrooks, J. (**1992**). ''Narrow-band sound localization related to external ear acoustics,'' J. Acoust. Soc. Am. **92**, 2607–2624.
Oldfield, S., and Parker, S. (**1984**). ''Acuity of sound localization: A topography of auditory space. I. Normal hearing conditions,'' Perception **13**, 581–600.
Perrett, S., and Noble, W. (**1995**). ''Available response choices affect localization of sound,'' Percept. Psychophys. **57**, 150–158.
Saberi, K., Dostal, L., Sadralodabai, T., Bull, V., and Perrot, D. (**1991**). ''Free-field release from masking,'' J. Acoust. Soc. Am. **90**, 1355–1370.
Strutt, J. W. L. (**1907**). ''On our perception of sound direction,'' Philos. Mag. **13**, 214–232.
Viemeister, N. F., and Wakefield, G. H. (**1991**). ''Temporal integration and multiple looks,'' J. Acoust. Soc. Am. **90**, 858–865.
Wightman, F. L., and Kistler, D. J. (**1989a**). ''Headphone simulation of free-field listening. I: Stimulus synthesis,'' J. Acoust. Soc. Am. **85**, 858–867.
Wightman, F. L., and Kistler, D. J. (**1989b**). ''Headphone simulation of free-field listening. II: Psychophysical validation,'' J. Acoust. Soc. Am. **85**, 868–878.
Wightman, F. L., and Kistler, D. J. (**1992**). ''The dominant role of low-frequency interaural time differences in sound localization,'' J. Acoust. Soc. Am. **91**, 1648–1660.
Wightman, F. L., and Kistler, D. J. (**1997**). ''Sound localization in the presence of multiple distracters,'' J. Acoust. Soc. Am. **101**, 3105(A).
Yost, W. A., Dye, R. H., and Sheft, S. (**1996**). ''A simulated ''cocktail party'' with up to three sources,'' Percept. Psychophys. **58**, 1026–1036.

Perceptual "vowel spaces" of cochlear implant users: Implications for the study of auditory adaptation to spectral shift

James D. Harnsberger, Mario A. Svirsky,[a),b)] Adam R. Kaiser,[a)] David B. Pisoni,[a)] Richard Wright,[c)] and Ted A. Meyer[a)]
Speech Research Laboratory, Department of Psychology, Indiana University, Bloomington, Indiana 47405

(Received 2 May 2000; accepted for publication 12 December 2000)

Cochlear implant (CI) users differ in their ability to perceive and recognize speech sounds. Two possible reasons for such individual differences may lie in their ability to discriminate formant frequencies or to adapt to the spectrally shifted information presented by cochlear implants, a basalward shift related to the implant's depth of insertion in the cochlea. In the present study, we examined these two alternatives using a method-of-adjustment (MOA) procedure with 330 synthetic vowel stimuli varying in $F1$ and $F2$ that were arranged in a two-dimensional grid. Subjects were asked to label the synthetic stimuli that matched ten monophthongal vowels in visually presented words. Subjects then provided goodness ratings for the stimuli they had chosen. The subjects' responses to all ten vowels were used to construct individual perceptual ''vowel spaces.'' If CI users fail to adapt completely to the basalward spectral shift, then the formant frequencies of their vowel categories should be shifted lower in both $F1$ and $F2$. However, with one exception, no systematic shifts were observed in the vowel spaces of CI users. Instead, the vowel spaces differed from one another in the relative size of their vowel categories. The results suggest that differences in formant frequency discrimination may account for the individual differences in vowel perception observed in cochlear implant users. [DOI: 10.1121/1.1350403]

PACS numbers: 43.66.Ts, 43.71.Es, 43.71.Ky [RVS]

I. INTRODUCTION

Although cochlear implants allow profoundly deaf people to hear, cochlear implant users show a very wide range of speech perception skills. The most successful cochlear implant users can easily hold a face-to-face conversation, and they can even communicate on the telephone, a difficult task because there are no visual cues available and because the acoustic signal itself is highly degraded (Gstoettner *et al.*, 1997). On the other hand, the least successful cochlear implant users have a difficult time communicating even in a face-to-face situation, and can barely perform above chance on auditory-alone speech perception tasks (Dorman, 1993).

One long term goal of our research is to understand the mechanisms that underlie speech perception by cochlear implant (CI) users and, in so doing, gain an understanding of the individual differences in psychophysical characteristics which may explain individual differences in speech perception with a CI. It is important to remember that electrical hearing as provided by a cochlear implant is quite different from normal acoustic hearing. One important difference lies in listeners' ability to discriminate formant frequencies. Kewley-Port and Watson (1994) report difference limens between 12 and 17 Hz in the $F1$ frequency region for highly practiced normal hearing listeners. In the $F2$ frequency region, they found a frequency resolution of approximately 1.5%. For cochlear implant users, formant frequency discrimination depends on two factors: the frequency-to-electrode map that is programmed into their speech processor, and the individual's ability to discriminate stimulation pulses delivered to different electrodes. It is not uncommon for some cochlear implant users to have formant frequency difference limens that are one order of magnitude larger than those of listeners with normal hearing, or even more (Nelson *et al.*, 1995; Kewley-Port and Zheng, 1998). It is reasonable to hypothesize that cochlear implant users with such limited frequency discrimination skills will find it quite difficult to identify vowels accurately because formant frequencies are important cues for vowel recognition.

Another important difference between acoustic and electric hearing is related to the finding that cochlear implants do not stimulate the entire neural population of the cochlea but only the most basal 25 mm at best, because the electrode array cannot be inserted completely into the cochlea. Therefore, cochlear implants stimulate cochlear locations that are more basal and thus elicit higher pitched percepts than normal acoustic stimuli. For example, when the input speech signal has a low frequency peak (e.g., 300 Hz), the most apical electrode is stimulated, regardless of the particular frequency-to-electrode table employed. The neurons stimulated in response to this signal may have characteristic frequencies of 1000 Hz or even higher. This represents a rather extreme modification of the peripheral frequency map. To the extent that the auditory nervous system of CI users is adaptable enough to successfully ''re-map'' the place fre-

[a)]Also at Devault Otologic Research Laboratory, Department of Otolaryngology-Head and Neck Surgery, Indiana University School of Medicine, Indianapolis, IN 46202.
[b)]Also at Department of Biomedical Engineering, and Department of Electrical Engineering, Purdue University, West Lafayette, IN 47907.
[c)]Now at the University of Washington, Seattle, WA 98195.

quency code in the cochlea, the basalward shift provided by a CI should not hinder speech perception. On the other hand, an inability to adapt and re-map the place frequency code may severely limit speech perception in CI users and may be an important source of individual differences in speech perception (Fu and Shannon, 1999a, b). Although there is consensus in the literature that cochlear implants stimulate neurons with higher characteristic frequencies than those stimulated by the same sound in normal ears, there is controversy about the amount of this possible basalward shift. For example, Blamey *et al.* (1996), in a study where CI users with some residual hearing were asked to match the electrical percepts in one ear to the acoustic percepts in the other ear, concluded that the electrode positions that matched acoustic pure tones were more basal than predicted from the characteristic frequency coordinates of the basilar membrane in a normal human cochlea. However, Blamey *et al.* (1996) also acknowledge that ''the listeners may have adapted to the sounds that they hear through the implant and hearing aid in everyday life so that simultaneously occurring sounds in the two ears are perceived as having the same pitch.'' On the other hand, Eddington *et al.* (1978) conducted the only experiment we are aware of, where a unilaterally deaf volunteer received a cochlear implant and was asked to match the pitch of acoustic and electric stimuli while still in the operating room, before he had much of a chance to adapt to the basalward shift in the way described by Blamey *et al.* (1996). Eddington *et al.* (1978) concluded that their pitch-matching data were ''consistent with frequency versus distance relationships derived from motion of the basilar membrane.''

Several previous studies have addressed the issue of adaptation to changes in frequency-to-electrode assignments for cochlear implant users. Skinner *et al.* (1995) showed that users of the SPEAK stimulation strategy identified vowels better with a frequency-to-electrode table that mapped a more restricted acoustic range into the subject's electrodes than the default frequency-to-electrode table. The experimental table that resulted in better vowel perception represented a more extreme basalward shift in spectral information than the default table, suggesting that listeners with cochlear implants can indeed adapt to such shifts, at least within certain limits. Another study that demonstrates the adaptation ability of human listeners in response to spectral shifts was conducted by Rosen *et al.* (1999), who used acoustic simulations of the information received by a hypothetical cochlear implant user who had a basalward spectral shift of 6.5 mm on the basilar membrane (equivalent to 1.3–2.9 octaves, depending on frequency). Initially, the spectral shift reduced word identification in normal-hearing subjects (1% correct, as compared to 64% for the unshifted condition), but after only three hours of training, subjects' performance improved to 30% correct. Whether or not such performance represents the maximum possible by CI users was not addressed in this study, given the relatively short time spent in training.

Recently, Fu *et al.* (submitted) performed an experiment in which the frequency-to-electrode tables of three cochlear implant users were shifted one octave with respect to the table they had been using daily for at least three years. It is important to note that this one-octave shift was in addition to the original shift imposed by the cochlear implant. After three months of experience with the new table, it was apparent that adaptation was not complete because, on average, subjects did not reach the same levels of speech perception that they had achieved before the table change. Taken together, these previous studies show that auditory adaptation to a modified frequency map is possible but it may be limited, depending on the size of the spectral shift that listeners are asked to adapt to.

In the present study, we investigated the adaptation of human listeners to a basalward shift using a new paradigm, a method-of-adjustment (MOA) procedure. This methodology was used to obtain maps of the perceptual vowel spaces of adult, postlingually deafened cochlear implant users. Similar tasks have been used with normal hearing listeners as well as CI users (Johnson *et al.*, 1993; Hawks and Fourakis, 1998). In this task, subjects select the region of the $F1-F2$ plane that sounds to them like a given vowel, and the procedure is repeated for ten English vowels. This task simultaneously assesses a cochlear implant user's auditory adaptation ability, by comparing the locations of his/her selected regions to those selected by normal hearing listeners, and his/her frequency discrimination skills, by examining the spread of the selected regions. More specifically, a listener who was unable to adapt to the basalward spectral shift introduced by their cochlear implant would be expected to select regions whose centers are systematically shifted to lower frequencies with respect to the regions of the vowel space selected by normal hearing listeners in mapping their vowel categories. The extent to which cochlear implant listeners show relatively normal vowel category centers could be used as a measure of their adaptation to basalward spectral shift.

An alternative to this hypothesis predicts that the spread of selected regions (category sizes), as well as category centers, would increase (or ''smear'') as a result of perceptual adaptation. The resulting vowel categories may be larger (show greater spread) reflecting the cochlear implant listener's need to map a greater range of frequencies to a given vowel. To differentiate between the spectral smearing hypothesis and the frequency discrimination explanation would require a longitudinal study of the changes in vowel spaces and frequency discrimination. For the purposes of this study, a simple shift hypothesis limited to vowel category centers was tested and compared with the frequency discrimination hypothesis.

In addition to the MOA task, two other perceptual tests were administered to CI users, an $F1$ jnd test with synthetic vowel stimuli and a closed set identification test with naturally produced vowels. Taken together, these measures were intended to investigate the role of formant frequency discrimination and auditory adaptation in vowel perception by CI users.

II. EXPERIMENT

A. Methods

1. Participants

Forty-three Indiana University undergraduates with no reported history of speech or hearing problems and eight

TABLE I. Demographic and other information for patients with cochlear implants. CNC=Percentage of correctly identified words on CNC word lists (male talker). Each patient was administered three 50-word lists. *CI6, CI7 Progressive; CI8, Progressive childhood.

Patient	Age	Gender	Age at onset of profound deafness	Age at implantation	Implant use (years)	CNC
CI1	67	F	43	61	6	52%
CI2	35	M	29	31	3	68%
CI3	37	M	34	36	1	46%
CI4	74	F	27	71	2	25%
CI5	63	M	56	57	5	11%
CI6	70	M	*	66	3	32%
CI7	68	F	*	65	2	1%
CI8	58	M	*	52	6	0%

adult cochlear implant (CI) users, all monolingual speakers of English, participated in this experiment. The normal-hearing participants consisted of 20 males and 23 females ranging in age between 18 and 28. The normal-hearing participants were recruited to represent the dialect of American English spoken in central Indiana with a common inventory of vowels. Only normal-hearing listeners who reported living their entire lives in central Indiana were included in this experiment. Central Indiana was defined in terms of a 60-mile radius around Indianapolis, roughly covering the Midland dialect as described by Wolfram and Schilling-Estes (1998). This criterion was used to exclude two other regional dialects found at the northern and southern extremes of the state. These other regional dialects are reported to differ from the Midland dialect in terms of vowel quality and degree and type of dipthongization (Labov, 1991). For participating in two 1-h sessions, the participants received either $7.50 per hour or two credits towards their research requirement if they were enrolled in an undergraduate psychology class.

The CI users were recruited from the population of adult patients served by the Department of Otolaryngology-Head and Neck Surgery at the Indiana University School of Medicine in Indianapolis. The demographics of the CI users are given in Table I, while information concerning their cochlear implants is provided in Table II. All of the CI users were native speakers of American English, with the exception of CI1, who was a native speaker of British English. British and Midland American English are not reported to differ substantively from one another in vowel quality for the ten vowels used in this study (Gimson, 1962; Pilch, 1994). Thus the American English vowel spaces were deemed an acceptable benchmark for CI1 as well as the other CI users.

All of the CI users had received their cochlear implants at least one year prior to participating in this study. The CI users differed from one another in terms of the type of cochlear implant they received: Five were users of the Nucleus-22 or Nucleus-24 device, programmed with either the SPEAK strategy or the MPEAK strategy, while three were users of the Clarion device, programmed with the CIS strategy. The SPEAK strategy (Skinner *et al.*, 1994) filters the incoming speech signal into a maximum of 20 frequency bands, which are associated with different intracochlear stimulation channels. Typically, six channels are sequentially stimulated in a cycle, and this cycle is repeated 250 times per second. The channels to be stimulated during each cycle are chosen based on the frequency bands with the highest output amplitude. In contrast, the CIS strategy (Wilson *et al.*, 1991) as implemented in the Clarion device filters the signal into eight bands, one for each stimulation channel. All channels are sequentially stimulated with pulses whose amplitudes are determined by the filters' outputs. The stimulation cycle is repeated at a rate of at least 833 times per second. The CIS strategy differs from the SPEAK strategy in its use of a different stimulation rate, fewer stimulation channels, and in its stimulation of all channels in a cycle rather than only a subset of the available channels.

TABLE II. Information concerning the patients' cochlear implants. Insertion depth indicates the distance from the most apical electrode to the round window. Frequency range indicates the acoustic range that was mapped into the available stimulation channels. Note: All patients with Clarion devices used the CIS stimulation strategy, while those patients with Nucleus devices used the SPEAK strategy, with the exception of CI5 who used the MPEAK strategy.

Patient	Implant type	Insertion depth (mm)	Frequency range (Hz)	Number of active channels	Stimulation mode
CI1	Clarion 1.0	25	350–5500	8	Bipolar
CI2	Clarion 1.2	25	350–6800	8	Bipolar
CI3	Nucleus 24	21.25	120–7800	21	Monopolar
CI4	Nucleus 22	19.75	170–4800	14	BP+1
CI5	Nucleus 22	22	300–4000	20	BP+1
CI6	Nucleus 22	19.75	150–9200	19	BP+1
CI7	Nucleus 22	19	120–6300	18	BP+3
CI8	Clarion 1.0	25	350–5500	8	Bipolar

TABLE III. The estimated basalward shift for the eight CI users for two cochlear locations corresponding to 475 Hz and 1500 Hz. Shifts are listed either as a frequency shift (Hz), an octave shift, or as the location difference in electrical and acoustic stimulation (mm).

	475 Hz			1500 Hz		
Patient	Hz	Octave	mm	Hz	Octave	mm
CI1	235	0.7	2.7	409	0.3	1.4
CI2	235	0.7	2.7	409	0.3	1.4
CI3	1301	1.9	8.1	2320	1.3	6.1
CI4	1719	2.2	9.2	2068	1.3	6.0
CI5	645	1.3	5.2	1669	1.1	5.1
CI6	1128	1.8	7.6	1542	1.2	5.3
CI7	1707	1.6	9.1	2334	2.1	7.4
CI8	235	0.7	2.7	409	0.3	1.4

The CI users also differed from one another in terms of the depth of insertion of the electrode array in the cochlea. The array's depth of insertion, in turn, determines the magnitude of the basalward spectral shift induced by the implant. It is possible to roughly estimate the size of this basalward shift for an individual CI user with three pieces of information: the location of the electrodes, the frequency to electrode mapping used by the cochlear implant's speech processor, and estimates of the characteristic frequency of the neurons stimulated by a given electrode pair. The speech processors of the participants in this study divide the acoustic frequency spectrum into channels. Each channel is specified by an acoustic frequency range that is assigned to a pair of intracochlear electrodes. Low frequencies are mapped to the apical electrodes, while high frequencies are mapped to the basal electrodes.

In the present study, the basalward spectral shift was calculated for two channels for every subject, the channels corresponding to 475 Hz and 1500 Hz. These frequencies correspond to the $F1$ and $F2$ of a neutral vowel for an average male speaker. First, the place of stimulation for a specific channel was defined as occurring half way between the electrodes for that channel. When stimulation was bipolar (i.e., both electrodes were intracochlear), the place of stimulation was defined as occurring half way between the electrodes for that channel. For patients who received monopolar stimulation (i.e., the return electrode was extracochlear) the location of electrical stimulation was considered to be at the intracochlear electrode. Second, the intraoperative report of insertion depth was used to adjust this place estimate based on the depth of electrode insertion. Cochlear lengths of the subjects were not measured individually. Instead, the basalward shift was calculated assuming that all subjects had average sized cochleas with a length of 35 mm (Hinojosa and Marion, 1983). Third, the electrically assigned frequency for this location was calculated as the geometric mean of the frequency boundaries defining the channel being studied in the speech processor. Finally, the characteristic frequency of the neurons stimulated by a given electrode was calculated from Greenwood (1961) and compared to the frequency calculated in step three. This discrepancy for each CI user, measured in Hz and octave, is shown in Table III. Alternatively, the shift is also reported in terms of the location difference between electrical and acoustic stimulation, in mm. The use of Greenwood's equation is based on the assumption that the average characteristic frequency of neurons stimulated by an electrode pair placed x mm from the round window is the frequency that would cause maximum displacement of the basilar membrane at the same x mm from the round window. Clearly, these are only rough estimates of the amount of basalward shift. In particular, the estimate of a 35-mm cochlea may lead to substantial overestimates or underestimates of the actual basalward shift. Future studies may improve the precision of these estimates by measuring the length of each individual cochlea using 3D reconstructions of CAT scans (Skinner *et al.*, 1994; Ketten *et al.*, 1998); by using the same 3D reconstructions to obtain more precise estimates of electrode location in the cochlea; and by obtaining physiological and behavioral data that may help determine the characteristic frequency of the neurons stimulated by different electrode pairs.

TABLE IV. The formulas for an important subset of the parameters used in generating the synthetic stimulus sets. B_n=the bandwidth of formants $F1$–$F4$.

Parameter	Formulas
Duration	250 ms
$F0$	Male: 120 Hz over the first half, dropping to 105 Hz at the end Female: 186 Hz over first half, dropping to 163 Hz at the end
$F3$	$\{(0.522*F1)+(1.197*F2)+57\}$ or $\{(0.7866*F1)-(0.365*F2)+2341\}$ Hz[a]
$F4$	3500 Hz or ($F3+300$ Hz), whichever higher
$B1$	$\{29.27+(0.061*F1)-(0.027*F2)+(0.02*F3)\}$ Hz
$B2$	$\{-120.22-(0.116*F1)+(0.107*F3)\}$ Hz
$B3$	$\{-432.1+(0.053*F1)+(0.142*F2)+(0.151*F3)\}$ Hz
$B4$	200 Hz

[a]The first formula was used for the half of the grid with higher $F2$ values, while the second was used for the half of the grid with lower $F2$ values.

2. Stimulus materials

a. Method-of-adjustment task. The stimulus set consisted of 330 synthetic, steady-state isolated vowels that varied from one another in their first and second formants in 0.377 Bark increments. The vowels were generated using the Klatt 88 synthesizer (Klatt and Klatt, 1990). The Bark increment size was chosen as a close approximation of the just noticeable difference for vowel formants of Flanagan (1957). The $F1$ and $F2$ values for this stimulus set ranged between 2.63 Z (250 Hz)–7.91 Z (900 Hz) and 7.25 Z (800 Hz)–15.17 Z (2800 Hz). These ranges were chosen to represent the full range of possible values for speakers of American and British English, and were successfully used in piloting the present experiment and in an earlier method-of-adjustment study of vowel perception in normal-hearing listeners (Johnson *et al.*, 1993). All of the other synthesis parameters for this stimulus set also followed Johnson *et al.* (1993). The formulas for calculating the values of the most relevant synthetic parameters are summarized in Table IV. The $F0$ parameter was varied to generate two sets of the 330 stimuli, one representing a male voice and one representing a female voice. All of the synthetic speech sounds were pre-

sented at a 70 dB C-weighted SPL listening level. The stimuli were presented over Beyer Dynamic DT-100 headphones for normal-hearing listeners, and over an Acoustics Research loudspeaker for CI users.

b. Vowel identification task. The vowel identification task employed a closed-set procedure that used nine vowels in an ''h-vowel-d'' format. The stimuli were digitized from the female vowel tokens of the Iowa laserdisc (Tyler *et al.*, 1987). Only steady-state vowels (no diphthongs) were used from this stimulus set. There were three separate productions of each vowel. Listeners were administered three lists that consisted of five repetitions of each vowel and three practice tokens. The stimuli were presented at a level of 70 dB C-weighted SPL over an Acoustics Research loudspeaker.

c. F1 jnd task. The stimuli for this task were seven synthetic three-formant vowels, with an $F2$ value of 1500 Hz and an $F3$ of 2500 Hz. $F1$ varied linearly, from 250 Hz for stimulus 1 to 850 Hz for stimulus 7. Three-formant rather than two-formant vowels were used in order to measure place-pitch discrimination with more realistic and speechlike stimuli. The stimuli were created using the Klatt 88 (Klatt and Klatt, 1990) speech synthesizer software. Voicing amplitude increased linearly in dB from zero to steady state over the first 10 ms and back to zero over the last 10 ms of the stimulus. Total stimulus duration was 1 s. Stimuli were presented at a level of 70–72 dB C-weighted SPL over an Acoustic Research loudspeaker.[1]

3. Procedures

a. Method-of-adjustment task. The procedures varied slightly for each participant group in the study. Normal-hearing participants were tested in a quiet room in two 1-h sessions. The second session took place approximately one week after the first session. In the first session, normal-hearing participants completed the method-of-adjustment task using one of the synthetic stimulus sets, either the male or the female voice set. In the second session, participants completed the method-of-adjustment task with the remaining stimulus set. The experiment was balanced for the order in which the two stimulus sets were presented to the normal-hearing participants. In contrast, the CI users were tested in a quiet room or a sound-attenuated chamber, in a single test session, varying in length by individual CI user between 1 and 3 h. Given the length of time CI users required to complete the MOA task and other demands on their time, they received only one of the two stimulus sets, the male voice set.

Each participant was presented with a two-dimensional (15 rows and 22 columns) visual grid centered in a computer display screen. The grid consisted of the 330 synthetic stimuli described above. A single English word appeared above this grid, constituting the target stimulus for a given trial. The visual target stimulus for a given trial was one of ten words, ''heed,'' ''hid,'' ''aid,'' ''head,'' ''had,'' ''who'd,'' ''hood,'' ''owed,'' ''odd,'' and ''hut,'' each of which contained one of the ten vowels under study, /i/, /ɪ/, /e/, /ɛ/, /æ/, /ʌ/, /ɑ/, /u/, /ʊ/, and /o/. Subjects were instructed to search the grid, playing out individual sounds until they located one or more synthetic sounds that matched the vowel in the visual target stimulus. Subjects were not required to search the grid exhaustively before making any selections. After making their choice(s), subjects were asked to give each synthetic sound a rating on a 1–7 scale, grading how close a match the synthetic sound was to the target. One repetition of the visual target stimulus set was presented to listeners. The order of presentation of the stimulus set varied randomly from participant to participant.

The particular stimuli chosen and their respective ratings were used to calculate *category centers* and *category sizes* for each vowel type for each listener group. The category center for a given vowel was determined by averaging the $F1$ and $F2$ frequencies of the stimuli selected to match the given vowel, with their contribution to the average weighted by their rating. The category sizes were computed from the standard deviation of the selected stimuli in both dimensions, taking the weights into account. Normal-hearing listeners' category centers were expected to appear in the $F1$ by $F2$ space in a similar arrangement to that observed in earlier vowel production studies with American English (i.e., Peterson and Barney, 1952; Hillenbrand *et al.*, 1995). If the category centers of CI patients deviated from those of normal-hearing listeners, the extent of that deviation was expected to be dependent on the magnitude of the basalward shift that individual listeners had to adapt to (see Table III). Failure to adapt completely to this kind of stimulation would result in an overall space that is shifted lower in $F1$ and $F2$ (i.e., the subject would select synthetic stimuli that were lower in $F1$ and $F2$ than those chosen by normal-hearing subjects for the same target vowel).

b. Vowel identification task. The vowel identification task, administered only to the CI users, was a closed-set speech perception task in which three separate tokens of each of nine /hVd/ tokens were presented in random order, one at a time. The CI users had to say which one of the nine stimuli they thought they heard by responding verbally. They were instructed to guess if they did not know which vowel was presented. All subjects heard a total of at least 15 presentations of each vowel (except CI5 who heard 10). The subjects' responses were tabulated and scored for total percentage of correct responses.

c. F1 jnd task. The $F1$ jnd task, administered only to the CI users, required listeners to make an absolute judgment. The stimuli for this task were labeled ''1'' through ''7'' in order of increasing $F1$. In this task, all stimuli were played in sequence several times, so the subjects could become familiar with the stimuli. The stimuli were then presented ten times each in random order and subjects were asked to identify the stimulus that was presented using one of the seven responses. The subject's response and the correct response were displayed on the computer monitor before the presentation of the next stimulus. After each block of 70 trials (10 presentations of each of 7 stimuli), the mean and standard deviations of the responses to each of the 7 stimuli were calculated. The d' for each pair of successive stimuli was calculated as the difference of the two means divided by the average of the two standard deviations. These d' measurements were then cumulated to calculate a cumulative d'

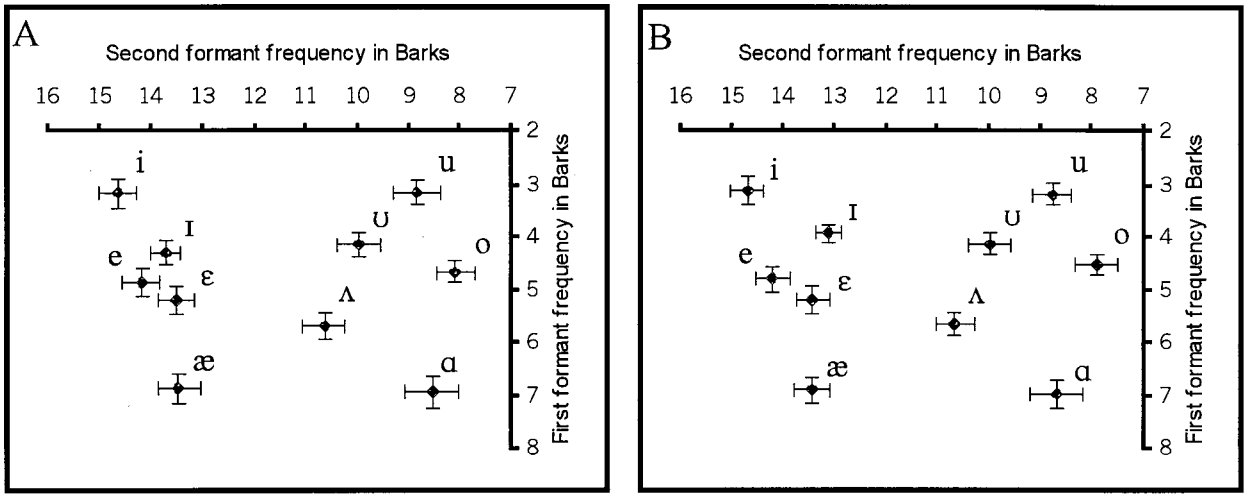


FIG. 1. The mean vowel space of normal-hearing listeners, calculated (a) using all of the ratings provided and (b) using only ratings of four or above.

curve, which provided an overall measure of the subject's ability to discriminate and pitch rank the seven stimuli (Durlach and Braida, 1969; Levitt, 1972). To calculate the cumulative d' curves, we followed the common assumption that the maximum possible value of d' was three. The average jnd (defined as the mean stimulus difference resulting in $d'=1$) was calculated based on the cumulative d' curve. Given that the $F1$ range spanned by the seven stimuli was 600 Hz, the jnd was defined as (600/cumulative d'). At least eight blocks of 70 trials were administered in order for all subjects to reach a plateau in performance as measured by the cumulative d'. The cumulative d' reported here is the average of the best two blocks for each subject. The normative mean cumulative d' from this procedure was calculated to be 53 Hz, based on a pilot study with six normal-hearing listeners.

B. Results

1. Normal-hearing participants

The normal-hearing listeners were expected to select vowel category centers with first and second formant values that corresponded to those typical in vowel production in American English. Figure 1 shows the mean vowel categories obtained from the group of normal-hearing subjects, for the male-voice stimulus set.[2] Each category center is shown along with error bars denoting the ''size'' of each category, that is, the relative spread of the category in both formant dimensions. The error bars represent the standard deviation from the category centers in both formant dimensions. In panel (A) of this figure, on the left, all of the ratings have been used to calculate the center and size of all ten vowel categories. Panel (B) of this figure, on the right, shows the vowel space of normal-hearing subjects calculated using only ratings of four and above. The rating of four was chosen because it was the highest rating that still allowed for category sizes to be calculated for all ten vowels of all of the normal-hearing and CI participants. The center of each category was determined by weighting the Bark values (in each dimension) of all synthetic stimuli that were chosen by their goodness ratings, and then averaging the weighted values.

The perceptual spaces shown in Fig. 1 for the normal-hearing listeners demonstrate that the method-of-adjustment technique for measuring vowel categories can be successfully used to generate vowel spaces that display the typical

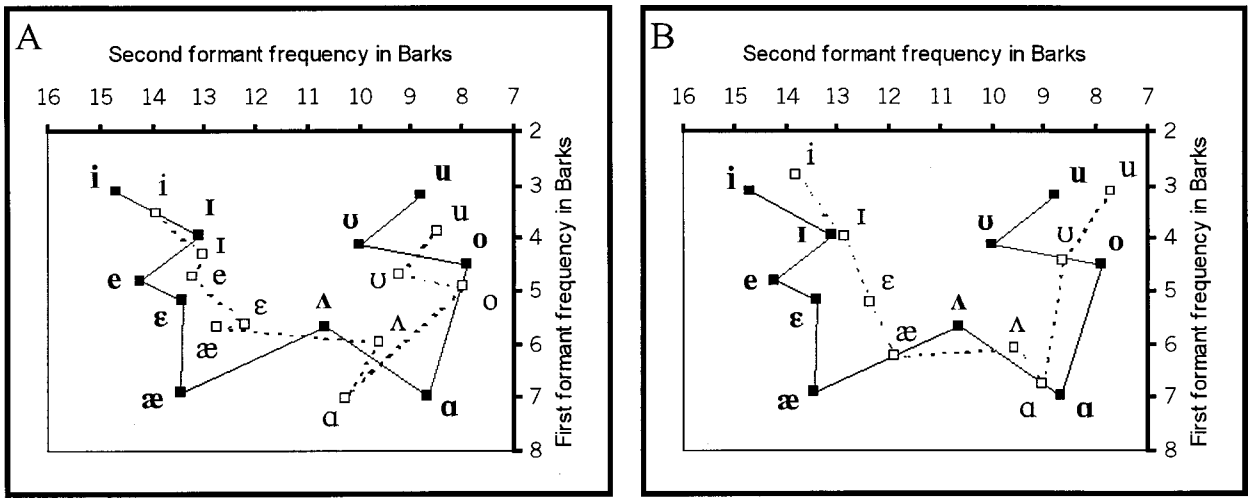


FIG. 2. A comparison of the vowel category centers from the MOA task (in bold, connected by straight lines) with the vowel production centers (in plain text, connected by dotted lines) from (A) Hillenbrand *et al.* (1995) and (B) Peterson and Barney (1952).

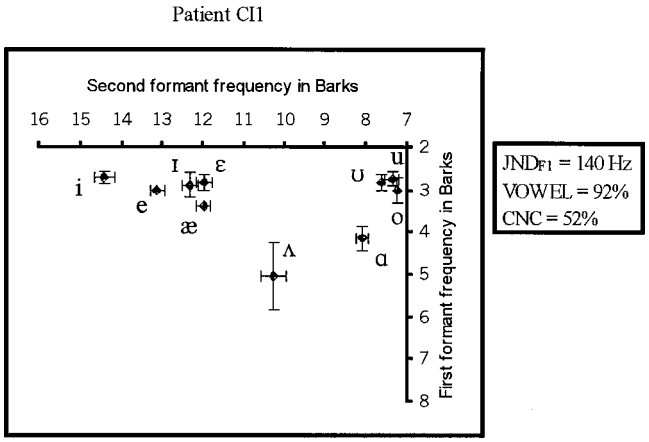


FIG. 3. The vowel space of patient CI1.

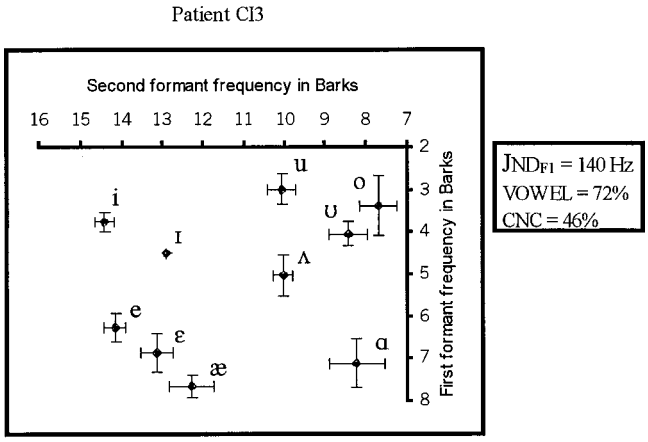


FIG. 5. The vowel space of patient CI3.

intervowel relationships that have been observed in $F1$ by $F2$ spaces generated from vowel production data (Peterson and Barney, 1952; Hillenbrand *et al.*, 1995). For instance, front vowel category centers have a higher $F2$ than back vowel centers; high vowel category centers have a lower $F1$ than low vowel category centers. While no vowel production data have been reported for English speakers from central Indiana,[3] Hillenbrand *et al.* (1995) examined the vowel production spaces of 45 men, 48 women, and 46 children who were native speakers of American English of the variety spoken in southern Michigan. The study by Hillenbrand *et al.* (1995) was a replication and extension of the classic work on the acoustics of American English vowels carried out by Peterson and Barney (1952). The two panels in Fig. 2 show the plots of the vowel production centers from these two studies compared with the MOA vowel category centers obtained from the normal-hearing subjects in the present study (calculated using ratings of 4 and above). Panel (A) shows the vowel category centers from the male talkers in Hillenbrand *et al.* (1995). Panel (B) shows the vowel category centers from the male talkers in Peterson and Barney (1952). While there are differences in the exact locations of the vowel perception and vowel production centers between the two studies, all three vowel spaces [the MOA perceptual vowel space, and the vowel production spaces of Peterson and Barney (1952) and Hillenbrand *et al.* (1995)] show a common set of $F1$ and $F2$ relations between vowels. The first and second formants from the normal-hearing listeners were significantly correlated with their counterparts in both the Hillenbrand *et al.* (1995) set ($r=+0.98$, $p<0.01$ for $F1$; $r=+0.89$, $p<0.01$ for $F2$) and the Peterson and Barney (1995) set ($r=+1.0$, $p<0.01$ for $F1$; $r=+0.83$, $p<0.05$ for $F2$).

2. Cochlear implant patients

The results of the MOA, $F1$ jnd, and vowel perception tasks for the vowel spaces of the eight CI users are shown in Figs. 3–10. Unlike Figs. 1 and 2, Figs. 3–10 show individual CI user spaces, rather than an averaged group space. Individual spaces for CI users were presented separately given the large potential individual differences in the CI vowel spaces and the stated goal of this research in examining individual differences in the CI user population. Each figure shows the labeled centers of the ten vowel categories. The size of each category in each dimension is indicated by error bars. The centers of the categories were computed using the mean in each dimension weighted by the rating given to each synthetic stimulus, using ratings of only four or above, just as in the normal-hearing listeners' vowel space in panel (B) of Fig. 1. The category sizes represent the standard deviation from the category center in the $F1$ and $F2$ dimensions. To

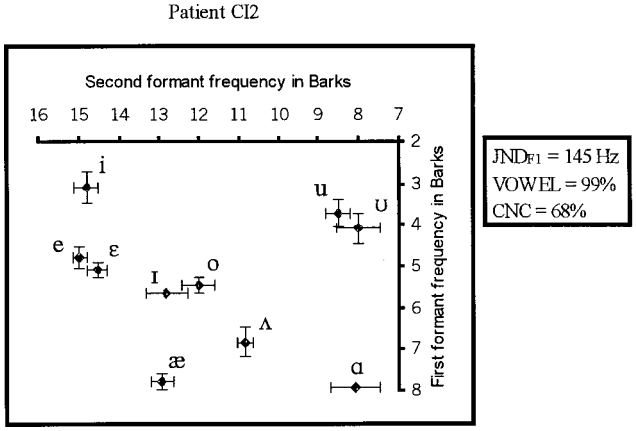


FIG. 4. The vowel space of patient CI2.

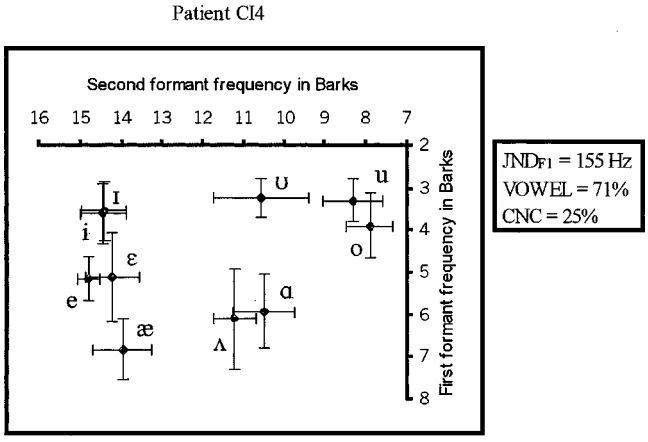


FIG. 6. The vowel space of patient CI4.

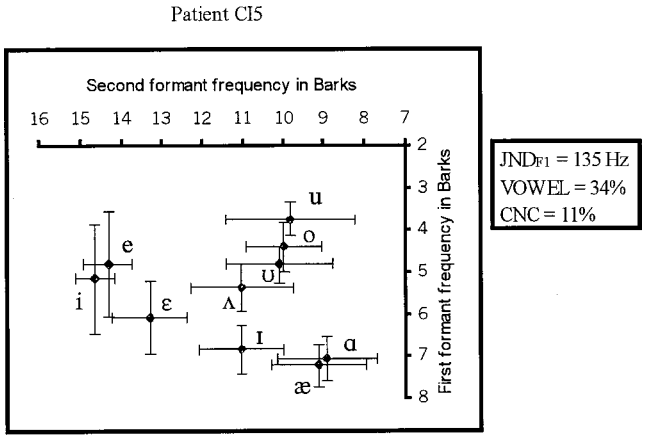


FIG. 7. The vowel space of patient CI5.

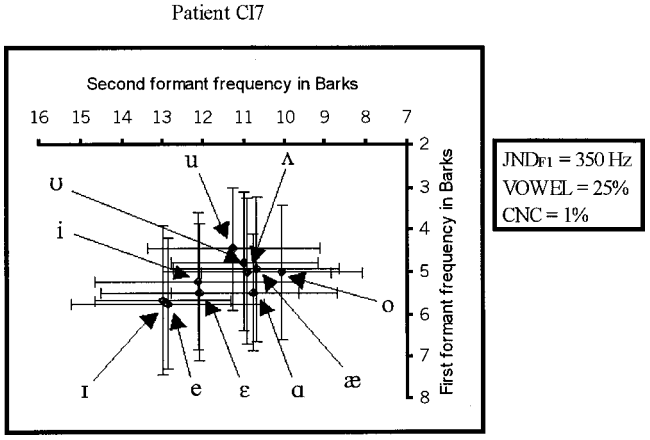


FIG. 9. The vowel space of patient CI7.

the right of each CI user's vowel space, the results of his/her $F1$ jnd (jnd$_{F1}$), vowel perception (VOWEL), and CNC word list (CNC) tests are listed.

An examination of all eight vowel spaces reveals little evidence of any systematic shift due to a lack of adaptation to a basalward spectral shift. The one possible exception is CI1's space. Instead, we find individual CI user spaces that differ from the normal space in terms of the sizes of perceptual categories, their degree of overlap, and the region of perceptual space that particular categories occupy. These observations are supported by measures of the differences between CI and normal vowel categories. Table V lists the absolute differences, in Bark, between the category centers of both normal and individual CI users, in both $F1$ and $F2$. Positive differences indicate that the formant frequency ($F1$ or $F2$) of the normal-hearing listeners for that vowel was greater than the equivalent formant frequency of the CI patient. If a shift were observable with a particular CI user, one would expect to see positive differences in one or both of the formants of most of the vowels of that user. Of the eight CI users, seven showed positive and negative differences, depending on the vowel and formant in question. Only one subject, CI1, showed systematically lowered formants for her category centers, as would be expected for a listener who has not adapted completely to the spectral shift introduced by the cochlear implant. In Fig. 3, CI1's categories are clearly shifted toward the lowest formant values in the upper right corner, resulting in a more compressed space. However, the magnitude of this shift was not the same for all ten vowels, or for both formants. The magnitude of the shift varied from $0.28\,Z$ to $3.53\,Z$, with on average a greater shift observed in $F1$ than $F2$. Surprisingly, this shift was not accompanied by a high jnd$_{F1}$ or low scores on the vowel identification and CNC word recognition tests. This subject did differ from the other seven in terms of her native dialect (British English) and in her professional background as a speech-language pathologist. It is possible, but difficult to determine, that her background biased her in some manner to provide a shifted vowel space in the MOA task.

The results shown in Figs. 3–10 indicate that most CI users adapted to their frequency-shifted input, although they varied widely in the degree to which their vowel spaces were similar to the vowel spaces of normal-hearing listeners. Moreover, this adaptation by CI users did not seem to be related to the depth of insertion of the electrode array of their implant. Table III listed the predicted basalward spectral shifts of the eight CI users in $F1$ and $F2$. If the individual differences in the magnitude of the predicted shifts were responsible for individual differences in vowel spaces, then we would expect to see a correlation between the actual differences between the normal and CI user category centers and the predicted shifts in each formant. In fact, neither of these

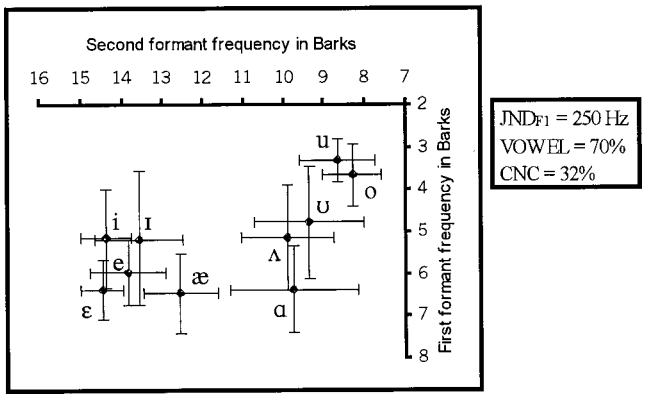


FIG. 8. The vowel space of patient CI6.

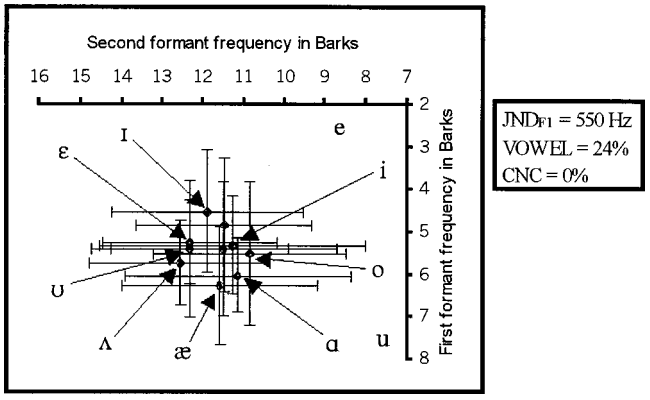


FIG. 10. The vowel space of patient CI8.

TABLE V. The differences in Bark between normal and CI vowel categories, for each formant (F), for individual vowel categories, and the mean difference across all categories.

		Vowel category										
Patient	F	i	ɪ	e	ɛ	æ	ʌ	ɑ	u	ʊ	o	Mean
CI1	$F1$	0.43	1.06	1.79	2.37	3.53	0.62	2.82	0.45	1.32	1.52	1.59
	$F2$	0.28	0.8	1.09	1.46	1.46	0.39	0.57	1.41	2.35	0.66	1.05
CI2	$F1$	0.05	−1.7	0.02	0.14	−0.86	−1.16	−0.93	−0.52	0.07	−0.92	−0.58
	$F2$	−0.13	0.31	−0.77	−1.13	0.52	−0.19	0.57	0.24	1.97	−4.1	−0.27
CI3	$F1$	−0.62	−0.57	−1.47	−1.65	−0.75	0.62	−0.15	0.2	0.09	1.15	−0.32
	$F2$	0.27	0.2	0.06	0.29	1.16	0.64	0.43	−1.29	1.54	0.21	0.35
CI4	$F1$	−0.47	0.39	−0.35	0.08	0.09	−0.44	1.04	−0.11	0.89	0.63	0.18
	$F2$	0.24	−1.3	−0.59	−0.81	−0.53	−0.57	−1.83	0.46	−0.6	0	−0.55
CI5	$F1$	−2.01	−2.91	−0.01	−0.89	−0.31	0.29	−0.1	−0.55	−0.69	0.12	−0.71
	$F2$	0.05	2.08	−0.1	0.13	4.31	−0.38	−0.26	−1.07	−0.14	−2.07	0.26
CI6	$F1$	−2.03	−1.23	−1.16	−1.2	0.44	0.51	0.6	−0.14	−0.65	0.86	−0.4
	$F2$	0.3	−0.45	0.38	−1.04	0.9	0.75	−1.08	0.1	0.6	−0.38	0.01
CI7	$F1$	−2.09	−1.74	−0.94	−0.27	1.97	0.67	1.49	−1.25	−0.62	−0.49	−0.33
	$F2$	2.57	0.15	1.33	1.35	2.73	−0.27	−2.09	−2.49	−1.01	−2.15	0.01
CI8	$F1$	−2.18	−0.6	−0.06	−0.05	0.65	−0.08	0.93	−2.22	−1.28	−1	−0.59
	$F2$	3.42	1.23	2.72	1.1	1.85	−1.89	−2.47	−2.71	−2.34	−2.93	−0.2

two correlations was significant using Spearman rank order correlations. Thus despite the different ''starting points'' of individual CI users in learning to normalize for spectrally shifted input, most patients seem to manage to completely adapt after, at the most, one year of experience.

While most CI users showed little evidence of a lack of adaptation to the basalward spectral shift, many of their vowel category centers differed from those of the normal-hearing listeners in nonsystematic ways. For example, the /e/ and /ɛ/ categories of CI3 are shifted lower with respect to the equivalent normal categories; /i/ and /ɪ/ of CI4 overlap almost completely with one another; and all of CI7 and CI8's categories are centralized and involve massive overlap. There are no obvious explanations for such discrepant category centers. For some categories at the periphery of the space, the centers may have differed from those of normal-hearing listeners if the grid did not encompass the entire relevant $F1*F2$ space for that listener. However, that would not account for the ''interior'' categories (mapped entirely out in the grid provided) that still differ from those of normal-hearing listeners (e.g., CI5's /ʊ/).

The CI users' vowel spaces also differed from the average normal vowel space in terms of vowel category size. This pattern indicates differences in frequency discrimination ability. Among the CI users' vowel spaces, we find several that are composed of categories that overlap very little, and that appear to be in roughly the same regions in perceptual space as those obtained from normal hearing listeners. Examples of such ''normal'' CI spaces are those from CI2 and CI3. In these spaces, front and back vowels occupy their ''normal'' regions of perceptual space. The sizes of the categories in these spaces are not much greater than those of normal hearing listeners. In contrast, the vowel spaces for CI7 and CI8 show much larger categories and a great deal of overlap between front and back vowels. The size of these categories and their arrangement in perceptual space suggest that these users might have great difficulty in using spectral information to discriminate among vowel sounds, or in identifying particular vowels in running speech. The vowel spaces of the four other CI users fall between these two extremes on a continuum characterized by category location and size.

These observations are supported by the difference measures displayed in Table V, and also by the mean sizes of categories (in Bark) for each CI space. These category sizes are shown in Table VI, for both the $F1$ and $F2$ dimension, along with the mean sizes averaged over both $F1$ and $F2$. The sizes for each perceptual dimension were calculated by averaging the Bark distances encompassed by one error bar in each formant dimension of each category center to get a mean category size for a given subject. Below the individual CI user measures are the averages across all CI users (''CI'')

TABLE VI. The mean size of individual CI user's categories in the $F1$ dimension, the $F2$ dimension, and both dimensions, along with the mean category sizes across all normal-hearing (NH) and CI users' categories. Size refers to one standard deviation (one error bar) from the vowel category center.

Patient	$F1$ (Z)	$F2$ (Z)	$F1$ and $F2$ (Z)
CI1	0.23	0.17	0.2
CI2	0.22	0.36	0.29
CI3	0.37	0.37	0.37
CI4	0.75	0.65	0.7
CI5	0.7	1.05	0.88
CI6	1.01	0.98	0.99
CI7	1.6	2.1	1.85
CI8	1.33	2.48	1.9
Mean CI	0.78	1.02	0.9
Mean NH	0.25	0.40	0.33

and the equivalent averages for all normal-hearing listeners (''NH''). The differences in all three measures between the CI users and the normal-hearing listeners was on the order of a 3:1 ratio, and all were significant at the $p<0.01$ level ($F1$: $t=9.1$, $df=81.1$; $F2$: $t=6.9$, $df=81$; $F1$ and $F2$ combined: $t=10.6$, $df=163.6$).

The vowel spaces of CI2 and CI3, the relatively ''normal'' spaces, have the lowest mean category sizes (along with CI1), while CI7 and CI8 have the highest. Large vowel categories typically overlap with one or more neighboring categories in perceptual space, which may be predictive of potential problems in the perception of vowels in natural speech. These findings are consistent with the results of the $F1$ jnd and vowel identification tests: the worst performers on these tests were also the subjects whose vowel spaces included the largest, overlapping categories. The pattern of results is reflected in Spearman rank order correlations between the category size measures and the $F1$ jnd and vowel identification tests. The $F1$ jnd score correlated significantly with category size in $F1$ ($r=+0.76$, $p<0.05$), while the vowel identification score correlated significantly with all three measures ($F1$: $r=-0.91$, $p<0.05$; $F2$: $r=-0.98$, $p<0.01$; $F1$ and $F2$: $r=-0.95$, $p<0.05$). All of the other correlations were only marginally significant. Thus there was a significant positive correlation between category size, as measured in this study, and performance on $F1$ discrimination, and a significant negative correlation between category size and vowel identification.

In addition, the three category size measures also correlated very strongly with the CNC word recognition scores obtained from the CI users ($F1$: $r=-0.88$, $p<0.05$; $F2$: $r=-0.95$, $p<0.05$; $F1$ and $F2$: $r=-0.91$, $p<0.05$). In other words, there was a significant negative correlation between category size and a word identification test. The CNC word lists measure listeners' ability to perform open set word identification. Thus it requires listeners to access their mental lexicons, making the task more similar to speech perception in natural settings than the MOA or phonetic identification tests reported here. The correlations between the CNC word recognition scores and category size results indicate that the MOA procedure provides useful information for characterizing the speech perception abilities of CI users.

III. DISCUSSION AND CONCLUSIONS

Despite the large individual differences observed among cochlear implant users in all three tests used in this study, the results revealed an orderly relationship between the perceptual vowel spaces of these listeners and their performance on the $F1$ jnd and the vowel identification tests. In Sec. I, we discussed two potential factors that may limit vowel identification by cochlear implant users: reduced formant frequency discrimination and lack of adaptation to basalward spectral shift. Overall, the results show that the major problem in vowel identification in these listeners appears to involve formant frequency discrimination. Only one of the eight CI users, CI1, showed a systematic shift of her vowel space that was consistent with limited auditory adaptation. For each vowel, she selected regions with lower $F1$ and $F2$ than those selected by normal hearing listeners. This result is consistent with the hypothesis that she has not completely adapted to the basalward spectral shift and thus selects vowels with very low formants as the best exemplars, to compensate for the frequency shift imposed by the cochlear implant.

Other than CI1, none of the other CI subjects showed evidence of any systematic shifts in their perceptual vowel spaces, suggesting that they were able to adapt to the frequency shift introduced by the cochlear implant. However, CI users clearly differ in the size of their vowel categories, reflecting differences in their ability to discriminate small differences in formant frequency. For example, CI7 and CI8 were the two CI users with the poorest jnd_{F1} values, vowel identification scores, and CNC word recognition scores. Their vowel spaces showed substantial overlap among most vowel regions. In contrast, CI users such as CI1 and CI2 with the best (smallest) jnd_{F1}'s, high vowel identification, and high CNC word recognition scores, tended to have little overlap among vowel regions.

Our results on perceptual vowel spaces are consistent with those reported recently by Hawks and Fourakis (1998). Their cochlear implant subjects also showed widely divergent amounts of overlap among vowel categories obtained using an identification task with synthetic stimuli. In terms of adaptation to basalward spectral shift, the present results are interesting because they suggest that most cochlear implant users are able to adapt to the frequency shifts typically introduced by their devices. However, although CI1's space was abnormal and thus not the best example of incomplete adaptation, her results do remind us that not all listeners may be able to adapt in a similar way. In addition, perhaps more listeners would find it difficult to adapt if the spectral shift was greater than that estimated for the cochlear implant subjects in this study. Moreover, the larger categories of the CI users could themselves reflect incomplete adaptation to basalward spectral shift, as the spectral smear hypothesis suggests.

In summary, the data reported in this paper suggest that vowel perception by cochlear implant users may be limited by the listener's formant frequency discrimination skills, in combination with his/her ability to adapt to basalward spectral shift. The present findings also have implications for understanding the perception of consonant sounds by cochlear implant users, because formant frequencies not only provide cues for vowel perception but also provide important information for consonant recognition. In future studies we intend to explore longitudinal changes in adaptation to basalward shift and in frequency discrimination, in order to investigate the nature of the improvement in speech perception observed in most postlingually deaf cochlear implant users after they receive their devices. An additional avenue for research may involve applying the same methods used in this study to prelingually deafened pediatric CI users. In this case, however, we would not expect to see major frequency shifts in the vowel spaces of children because they would not have prior expectations concerning the location of vowels in the vowel space. Instead, the frequency discrimination skills of prelingually deafened pediatric CI users may be, on average, even worse than those of postlingually deafened adult

CI users. Finally, it may be interesting to determine whether the listeners who show incomplete adaptation have less behavioral plasticity than the other listeners, or whether their lack of adaptation is due to a shallower insertion, which would result in a greater spectral shift to be overcome by the listener. Taken together, these studies should help us account for the enormous individual differences in speech perception by CI users, which should provide the theoretical rationale for improved devices and intervention strategies in this clinical population.

ACKNOWLEDGMENTS

This work was supported by NIH-NIDCD Training Grant No. DC00012, NIH-NIDCD Grant No. R01-DC00111, and NIH-NIDCD Grant No. R01-DC03937. We would like to thank Dr. Richard Miyamoto for providing the electrode insertion information for the cochlear implant users. We would also like to thank two anonymous reviewers for their helpful comments, as well as Chris Quillet for his technical assistance.

[1]Some stimuli were slightly louder than others, but these small differences in loudness should be very hard to perceive for most CI users.

[2]The female stimulus set results were not included in the normal-hearing listeners' averaged space because the CI users were only presented with the male stimulus set.

[3]However, work is in progress in our laboratory on a vowel production study of Central Indiana English, using as talkers the same normal-hearing listeners that participated in this study.

Blamey, P. J., Dooely, G. J., Parisi, E. S., and Clark, G. M. (**1996**). ''Pitch comparisons of acoustically and electrically evoked auditory sensations,'' Hear. Res. **99**, 139–150.

Dorman, M. F. (**1993**). ''Speech perception by adults,'' in *Cochlear Implants*, edited by R. S. Tyler (Singular Publishing, San Diego), pp. 145–190.

Durlach, N. I., and Braida, L. D. (**1969**). ''Intensity perception. I. Preliminary theory of intensity resolution,'' J. Acoust. Soc. Am. **46**, 372–383.

Eddington, D. K., Dobelle, W. H., Brackmann, D. E., Mladejovsky, M. G., and Parking, J. L. (**1978**). ''Auditory prostheses research with multiple channel intracochlear stimulation in man,'' Ann. Otol. Rhinol. Laryngol. **53**, 5–39.

Flanagan, J. (**1957**). ''Estimates of the maximal precision necessary in quantizing certain 'dimensions' of vowel sounds,'' J. Acoust. Soc. Am. **29**, 533–534.

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**). ''Effects of electrode configuration and frequency allocation on vowel recognition with the Nucleus-22 Cochlear Implant,'' Ear Hear. **20**, 332–344.

Fu, Q.-J., Shannon, R. V., and Galvin, J. (submitted). ''Performance over time of Nucleus-22 cochlear implant listeners wearing a speech processor with a shifted frequency-to-electrode assignment,'' J. Acoust. Soc. Am.

Gimson, A. C. (**1962**). *An Introduction to the Pronunciation of English* (Edward Arnold, London).

Greenwood, D. D. (**1961**). ''Critical bandwidth and the frequency coordinates of the basilar membrane,'' J. Acoust. Soc. Am. **33**, 1344–1356.

Gstoettner, W., Hamzavi, J., and Czerny, C. (**1997**). ''Rehabilitation of patients with hearing loss by cochlear implants,'' Radiologe **37**, 991–994.

Hawks, J. W., and Fourakis, M. S. (**1998**). ''Perception of synthetic vowels by cochlear implant recipients,'' J. Acoust. Soc. Am. **104**, 1854–1855.

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.

Hinjosa, R., and Marion, M. (**1983**). ''Histopathology of profound sensorineural deafness,'' Ann. N.Y. Acad. Sci. **405**, 459–484.

Johnson, K., Flemming, E., and Wright, R. (**1993**). ''The hyperspace effect: Phonetic targets are hyperarticulated,'' Language **69**, 505–528.

Ketten, D. R., Skinner, M. W., Wang, G., Vannier, M. W., and Gates, G. A. (**1998**). ''*In vivo* measures of cochlear length and insertion depth of nucleus cochlear implant electrode arrays,'' Ann. Otol. Rhinol. Laryngol. **175**, 1–16.

Kewley-Port, D., and Watson, C. S. (**1994**). ''Formant-frequency discrimination for isolated English vowels,'' J. Acoust. Soc. Am. **95**, 485–496.

Kewley-Port, D., and Zheng, Y. (**1998**). ''Auditory models of formant frequency discrimination for isolated vowels,'' J. Acoust. Soc. Am. **103**, 1654–1666.

Klatt, D. H., and Klatt, L. C. (**1990**). ''Analysis, synthesis, and perception of voice quality variations among female and male talkers,'' J. Acoust. Soc. Am. **87**, 820–857.

Labov, W. (**1991**). ''The three dialects of English,'' in *New Ways of Analyzing Sound Change*, edited by P. Eckert (Academic, New York), pp. 1–44.

Levitt, H. (**1972**). ''Decision theory, signal-detection theory, and psychophysics,'' in *Human Communication: A Unified View*, edited by E. E. David and P. B. Denes (McGraw-Hill, New York), pp. 114–174.

Nelson, D. A., Van Tasell, D. J., Schroder, A. C., Soli, S., and Levine, S. (**1995**). ''Electrode ranking of 'place pitch' and speech recognition in electrical hearing,'' J. Acoust. Soc. Am. **98**, 1987–1999.

Peterson, G. E., and Barney, H. L. (**1952**). ''Control methods used in the study of vowels,'' J. Acoust. Soc. Am. **24**, 175–184.

Pilch, H. (**1994**). *Manual of English Phonetics* (W. Fink, Munchen).

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.

Skinner, M. W., Holden, L. K., and Holden, T. A. (**1995**). ''Effect of frequency boundary assignment on speech recognition with the SPEAK speech-coding strategy,'' Ann. Otol. Rhinol. Laryngol. **166**, 307–311.

Skinner, M. W., Ketten, D. R., Vannier, M. W., Gates, G. A., Yoffie, R. L., and Kalender, W. A. (**1994a**). ''Determination of the position of nucleus cochlear implant electrodes in the inner ear,'' Am. J. Otolaryngol. **15**, 644–651.

Skinner, M. W., Clark, G. M., Whitford, L. A., Seligman, P. M., Staller, S. J., Shipp, D. B., Shallop, J. K., Everingham, C., Menapace, C. M., Arndt, P. L., Antogenelli, T., Brimacombe, J. A., Pijl, S., Daniels, P., George, C. R., McDermott, H., and Beiter, A. L. (**1994b**). ''Evaluation of a new spectral peak coding strategy for the Nucleus 22 channel cochlear implant system,'' Am. J. Otolaryngol. **15** (Suppl. 2), 15–27.

Tyler, R. S., Preece, J. P., and Lowder, M. W. (**1987**). ''The Iowa audiovisual speech perception laser video disk,'' *Laser Videodisk and Laboratory Report*, University of Iowa at Iowa City, Department of Otolaryngology-Head and Neck Surgery.

Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., and Rabinowitz, W. M. (**1991**). ''Better speech recognition with cochlear implants,'' Nature (London) **352**, 236–238.

Wolfram, W., and Schilling-Estes, N. (**1998**). *American English* (Blackwell, Malden, MA).

Effects of delayed auditory feedback (DAF) on the pitch-shift reflex

Timothy C. Hain
Departments of Otolaryngology, Head and Neck Surgery, and Neurology, Northwestern University Medical School, Chicago, Illinois 60611

Theresa A. Burnett
Laryngeal & Speech Section, MNB/NINDS/NIH, Bldg. 10, Rm. 5D38, Bethesda, Maryland 20892-1416

Charles R. Larson[a] and Swathi Kiran
Department of Communication Sciences and Disorders, Northwestern University, Evanston, Illinois 60208

(Received 27 January 2000; accepted for publication 22 February 2001)

Changes in voice pitch auditory feedback to vocalizing subjects elicit compensatory changes in voice fundamental frequency (*F*0). The neural mechanisms responsible for this behavior involve the auditory and vocal-motor systems, collectively known as the audio-vocal system. Previous work [Burnett *et al.*, J. Acoust. Soc. Am. **103**, 3153–3161 (1998); Hain *et al.*, Exp. Brain Res. **130**, 133–141 (2000); Larson *et al.*, J. Acoust. Soc. Am. **107**, 559–564 (2000)] indicated that this system operates using negative feedback to cancel out low-level errors in voice *F*0 output. By introducing delays in the auditory feedback pathway, we hoped to transiently ''open'' the feedback loop and learn which components of the response are most closely related to the timing of the auditory feedback signal. Subjects were presented with pitch-shift stimuli that were paired with a delay of 0, 50, 100, 200, 300, or 500 ms. Delayed auditory feedback did not affect voice *F*0 response latency or magnitude, but it delayed the timing of later parts of the response. As a further test of the audio-vocal control system, a second experiment was conducted in which delays of 100, 200, or 300 ms were combined with stimuli having onset velocities of 1000 or 330 cents/s. Results confirmed earlier reports that the system is sensitive to velocity of stimulus onset. A simple feedback model reproduced most features of both experiments. These results strongly support previous suggestions that the audio-vocal system monitors auditory feedback and, through closed-loop negative feedback incorporating a delay, adjusts voice *F*0 so as to cancel unintentional small magnitude fluctuations in *F*0. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1366319]

PACS numbers: 43.70.Aj [AL]

I. INTRODUCTION

The control of voice fundamental frequency (*F*0) is a complex process involving volition, memory, kinesthetic and proprioceptive feedback, auditory feedback, and the neuromuscular adjustments of the respiratory and laryngeal systems necessary to produce a desired *F*0 level. The ''pitch-shift reflex'' is a vocal response to changes in the pitch of voice auditory feedback. It is thought to help stabilize voice *F*0 in the presence of unintended perturbations. The pitch-shift reflex can be elicited by experimental manipulation of auditory feedback, offering a unique means of studying neural mechanisms of vocal control (Burnett *et al.*, 1998, 1997; Hain *et al.*, 2000; Kawahara, 1995; Kawahara *et al.*, 1996; Larson *et al.*, 1997, 1995, 1996).

Most voice *F*0 responses elicited by a pitch shift are compensatory. That is, a downward shift in feedback pitch (pitch-shift stimulus) results in a rise in voice *F*0, and an upward change in feedback results in a reduction in *F*0. There is also proportionality between stimulus onset velocity and response velocity (Larson *et al.*, 2000). This relation was largely replicated in a mathematical model of the audio-vocal system incorporating negative feedback control incorporating intrinsic delays, and filtering (Larson *et al.*, 2000). These observations indicate the audio-vocal system is sensitive to both the direction and the speed of change in feedback pitch, supporting the hypothesis that the function of the pitch-shift reflex is to null unexpected changes in voice *F*0 output and thus stabilize *F*0.

In studying human behavior using negative feedback control, while the internal components of the system are usually unavailable for experimental modification, it is often possible to modify the gain or timing of the feedback signal. To the extent that negative feedback is used to stabilize the system, these procedures should lead to instabilities. In the present study, pitch-shift stimuli (PSS) were presented in a delayed auditory feedback (DAF) paradigm to alter the timing of the system, and so test the hypothesis that the audio-vocal system operates in a negative feedback mode. While subjects were vocalizing a steady vowel (ah), a harmonizer continuously stored a segment of the voice in a buffer. At some time during the vocalization, that stored segment was unexpectedly shifted upward or downward in pitch and fed back to the subjects over headphones. This caused the subject to hear what sounded to be their own voice at a higher or lower pitch than they were currently producing, but evidence

[a]Electronic mail: clarson@northwestern.edu

of all subsequent vocal responses was delayed by an interval equivalent to that held in the buffer. This paradigm prevented subjects from hearing their *F*0 response to the PSS during the delay interval. Instead, they just heard the pitch-shift stimulus itself. Thus, an error between intended and perceived pitch was created and perception of responses to this error was delayed. By this means, the error correction feedback loop was transiently opened. At the end of the delay interval, the subjects heard their response to the PSS. The resultant behavior was largely replicated with the negative feedback mathematical model previously reported.

II. EXPERIMENT

Subjects: Twenty-two healthy young adults between the ages of 18 and 22 (21 females and 1 male) served as subjects. All subjects passed a hearing screening at 20 dB SPL (500–9 kHz), reported no neurological deficits, had no speech or voice disorder and were not trained singers.

Apparatus and procedures: Subjects were seated in a sound-treated booth, their voices transduced with an AKG boom-set microphone, amplified, recorded, processed for pitch shifting through an Eventide (SE 3000) Ultraharmonizer, mixed with 70 dB (SPL) masking noise (3 dB/oct, 100–5 kHz) and fed back to the subject over AKG earphones (Model K 270 H/C). The harmonizer shifts all frequencies, voice *F*0 as well as formants, and thus the shifted feedback signal sounds like a person's normal voice at a different *F*0 (details of the pitch-shifting algorithm are a trade secret of the manufacturer and are thus unavailable). Throughout the experiment, subjects maintained vocal loudness at approximately 70 dB SPL, resulting in a voice feedback loudness of approximately 80 dB. For additional methodological details see Burnett *et al.* (1998).

Subjects were instructed to vocalize ''ah'' at a comfortable pitch for 5 s, pause for a breath, then repeat. Thirty vocalizations constituted a block. Within each block of vocalizations, 15 experimental trials were mixed pseudorandomly with 15 control trials, in which no PSS was presented. In the experimental trials, a PSS of 100 cents (1200 cents=1 oct) was presented at a random time 500 to 2500 ms after vocalization onset. At the same instant that the pitch shift was initiated, the harmonizer also imposed a delay (DAF) of 0, 50, 100, 200, 300, or 500 ms, which extended the time that the system operates open loop. Thus for the duration of the PSS, auditory feedback was 100 cents displaced from the voice *F*0 of 0–500 ms ago. Given production of a steady note, the PSS created a sudden discontinuity between the *F*0 the subject was producing and the auditory feedback pitch. However, subjects did not hear their response to the PSS until the delay interval had elapsed. After 1 s the PSS was terminated.

The PSS onset and offset were abrupt, i.e., 10,000 cents/s. During a single block of trials, one delay condition was presented, resulting in a total of six blocks per subject. Fifteen subjects were tested with downward PSS and seven with upward PSS. We did not think it necessary to have equal numbers of subjects with upward and downward PSS, since previous studies have failed to show significant differences between these two conditions. (Burnett *et al.*, 1988, 1997; Hain *et al.*, 2000).

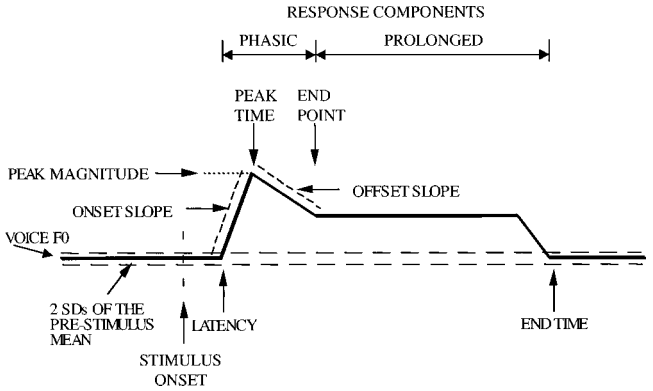


FIG. 1. Illustration of measurement parameters.

Throughout the experiment, the subject's voice, the feedback signal to the subject, and a control pulse to the harmonizer were digitized at a sample frequency of 2 kHz. The subject's voice and the feedback signal were low-pass filtered at 200 Hz (females) or 100 Hz (males) prior to digitization. The *F*0 of digitized acoustical signals was extracted using a software algorithm resulting in dc signals where voltage corresponded to *F*0 (analog *F*0). All signals in each experimental trial were then time-aligned with the PSS-DAF onset and averaged within each block.

From the averaged waveforms, the following set of direct measures was automatically made: response latency, end time, peak magnitude, and peak time (Fig. 1). Response latency was defined as the time poststimulus onset at which the analog *F*0 signal exceeded two standard deviations of the prestimulus mean *F*0. Response end time was that time at which the response returned to within two S.D.'s of the prestimulus mean. Response duration (latency to end time) and onset slope were calculated from these direct measures.

The *F*0 response often peaked and then fell back to a holding value, resulting in two response components: phasic and prolonged. In these cases, several additional measures were calculated. The time when the phasic response ended, ''phasic end time,'' was determined from visual inspection of the records and was defined as a dramatic change in slope of the *F*0 response following the response peak (see Fig. 1). The detection of the slope change was made by eye with the concurrence of two independent observers. Confirmation of the time of the slope change was made by low-pass filtering (5 Hz) and then differentiating the averaged waveform. A change in slope of the differentiated signal was apparent at the offset time of the phasic response. For 61 cases in which the phasic peak end time could clearly be defined, it was measured along with that of the phasic peak offset duration (peak time to end time), slope of phasic peak offset, and phasic peak duration. Identical measures were made for downward responses following increasing pitch shift stimuli. Magnitude measures were converted from Hertz to cents (100 cents=1 semitone) and the absolute values determined in order to permit joint consideration of measures from upward and downward PSS direction conditions. The measures were charted and submitted to significance testing using a

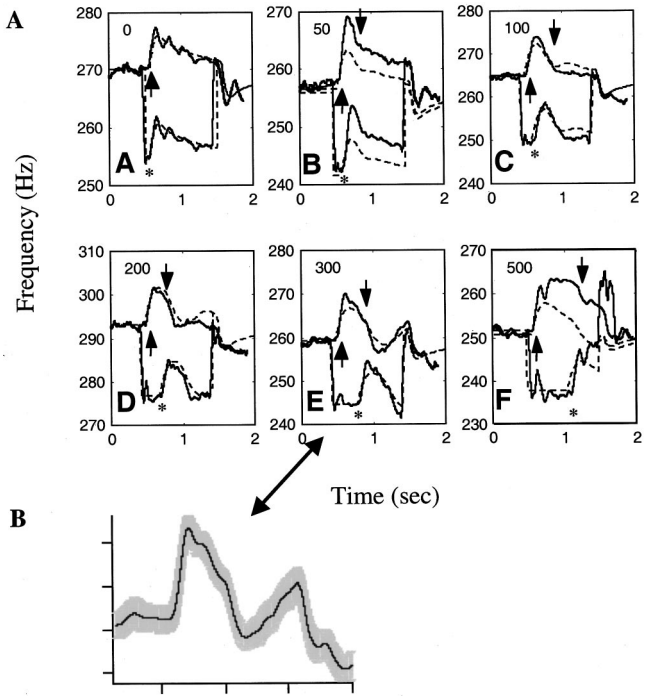


FIG. 2. (a) Examples of averaged voice *F*0 responses to decreasing PSS directions for experiment 1. Each of six boxes represents data for a delay condition. Solid lines are experimental data and dashed lines are simulations. Upper traces are voice *F*0 and lower traces are feedback signals. All traces are from the same subject, and each trace consists of 15 trials. The *Y* axis is frequency in Hz. Upward arrows indicate beginning of *F*0 response. Downward arrows indicate end of phasic component of *F*0 response. A phasic response was not identified for the example in subpanel A of (a), and hence there is no downward arrow to indicate end of the response. Asterisk indicates delayed response of auditory feedback to *F*0 response. Simulations using mathematical model of Fig. 6 parametrized with median values. (b) Enlarged view of average wave and standard error (gray) of the average for 300-ms delay condition.

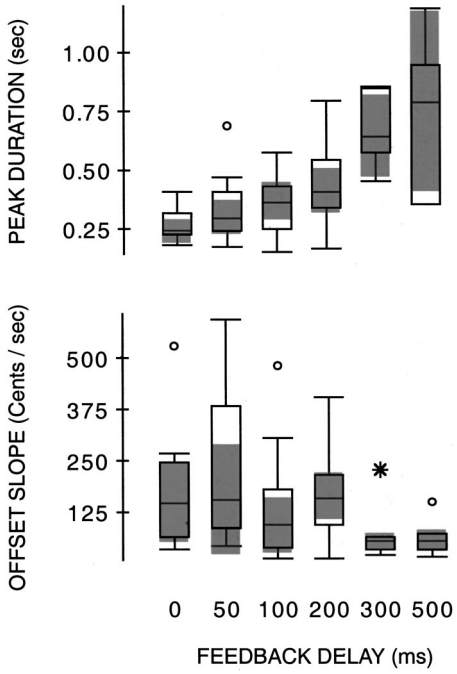


FIG. 3. Boxplots of peak duration and offset slope measures. Boxplot definitions: the upper and lower limits of the boxes (hinges) represent the 75th and 25th percentiles, respectively. Gray bars represent 95% confidence intervals. Horizontal line through box is the median. Whiskers extend to upper and lower limits of main body of data, defined as high hinge+1.5 (high hinge-low hinge) and low hinge−1.5 (high hinge-low hinge). Points depicted as ''o'' extend beyond the above limits, unless they exceed high hinge+3.0 (high hinge−low hinge) or low hinge-3.0 (high hinge-low hinge), in which case they are shown as ''*'' (Data Desk, Data Description, Inc.).

MANOVA in DataDesk™ (Data Description, Inc., Ithaca, NY). Mathematical modeling and simulation of responses was performed using the Matlab/Simulink package (The MathWorks, Inc., Nantick, MA).

III. RESULTS

We measured 126 averaged responses. In all but four responses, subjects changed voice *F*0 in the direction opposite to that of the stimulus (i.e., a ''compensating response''). There were 61 cases in which initial phasic and subsequent prolonged response components could be clearly identified. In the remaining cases, only a prolonged response was measured. No delay condition resulted in significantly higher or lower incidence of phasic responses (chi square=0.6).

Figure 2(a) (A–F) shows representative averaged data from one subject. Data simulations also shown are discussed later. Each part of the graph consists of two traces, the top is the subjet's voice *F*0 and the bottom is the feedback pitch. In each case, the onset of the PSS is seen as a downward deflection in the lower trace, which lasts 1000 ms. Shortly after the PSS onset, the subject's *F*0 response is seen as an upward deflection (upward arrow) in the upper trace. The phasic portion of the response ends at the downward arrow, followed by the prolonged component of the *F*0 response (except 300 and 500-ms conditions), which continues until just beyond the termination of the 1000 ms PSS. Another response occurred at the termination of the PSS, but it was not measured and so will not be discussed further. The subject's *F*0 response onset was reflected in the auditory feedback trace, seen as an upward deflection in the lower trace and indicated by an asterisk (*). The delayed presentation of response feedback (*) apparently constituted a second ''pitch-shift stimulus,'' eliciting an additional compensatory response, this time a reduction in *F*0 that terminated the first *F*0 response (downward arrows). Those subjects who had received PSS that increased in pitch produced patterns opposite in direction to those seen in Fig. 2. Figure F2(b) illustrates the average wave for the 300 cents DAF along with the standard error of the waves making up the average.

The data in Fig. 2 suggest that with an increase in DAF interval, there is a corresponding increase in the duration of the phasic (peak) response. After logarithmically transforming the data to achieve homogeneity of variance, overall significance testing with a multivariate design revealed significant differences among the dependent variables as a function of DAF interval ($F_{\text{approx}}=2.34$, $df=30$, 194, $p<0.005$). Follow-up testing revealed significant differences ($p<0.005$) in phasic peak duration between the 300 ms and 0, 50, and 100 ms, and between 500 ms and 0 ms. The examples in Fig. 2 suggest that the offset slope may also have been affected by the delay conditions, and the boxplots of Fig. 3 also suggest this, however, the differences were not

significant. There were no effects of feedback delay on peak magnitude, latency, peak time, onset slope, or offset slope (see Fig. 1). The overall mean response latency was 130 ms (s.d.=31 ms) and mean peak magnitude was 69 cents (s.d.=69 cents). Although not statistically significant, the delay conditions resulted in a larger peak magnitude (mean=71 cents) compared with the no delay condition (mean=59 cents). There were no differences in any of the response measures as a function of pitch-shift direction.

In our previous work, we proposed a schematic for an *F*0 tracking system, implemented as an explicit mathematical model (Larson *et al.*, 2000). We configured and drove an extension of this model (see Fig. 6) with input waveforms synthesized to match those used in this experiment. We assessed the goodness of fit from the squared difference between experimental data and simulated data normalized to the squared difference between the experimental data and the baseline *F*0 (subsequently called vaf for ''variance accounted for'').

A second-order filter [Eq. (1)] was needed to reproduce our data accurately:

$$\frac{fgain}{sTc+1}\cdot\frac{sTa}{sTa+1}. \tag{1}$$

Using the optimization toolbox of Matlab, we determined the values for four model parameters (*fgain*, *sTa*, *sTc*, *Delay*) that had the largest vaf. When all four parameters were allowed to vary for each DAF delay, the simulations of a set of six averaged experimental traces from one subject accounted for 91.2% (85.2%–96.7%) of the variance in experimental data. Median parameter values from the six fits above were $Tc=0.170$ s, $T=0.235$ s, $fgain=1.15$, and $delay=0.06$. Using these values uniformly for all DAF delays, the model accounted for 76.6% of the variance (see Fig. 2). A first-order filter (i.e., $fgain/Stc+1$) performed much more poorly, accounting for only 18% of the variance. A less constrained second-order filter, described in Eq. (2) under the heading of experiment 2, fit the data equally well as the filter of Eq. (1).

IV. EXPERIMENT 2

Subjects: Twenty-three normal undergraduate students (20 females, 3 males) participated. None reported speech or neurological disorders or were trained singers.

Procedures: All aspects of this experiment were identical to the first with a few exceptions. The first exception was that the pitch-shift stimuli had onset velocities of either 100 or 330 cents/s. These two velocity conditions resulted in ''ramp-shaped'' PSS with durations of 100 or 300 ms. Second, the subject's voice feedback was delayed upon initiation of the PSS by 100, 200, or 300 ms. Thus, there were six experimental conditions. The PSS remained on for 3 s, although only data up to 1.5 s poststimulus onset were analyzed. Within each block of 30 trials, there were 15 upward stimuli pseudo-randomly mixed with 15 trials in which no stimuli were presented.

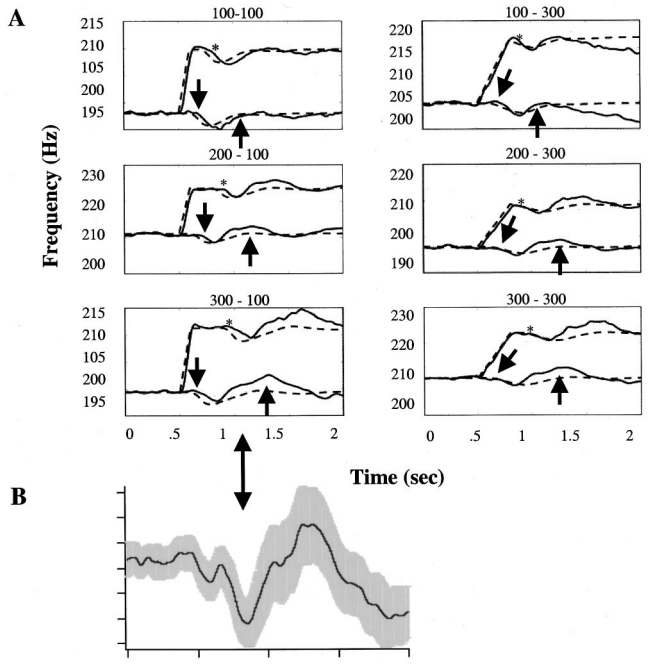


FIG. 4. (a) Examples of averaged voice *F*0 responses to ramped PSS in the increasing direction from experiment 2. Each of six boxes represents data for a combination of ramp duration and a delay condition (duration-delay). The step magnitude was 100 cents in all cases. Solid lines are experimental data and dashed lines are simulations. Upper traces are feedback signals and lower traces are voice *F*0. Simulations are shown using the mathematical model of Fig. 6, parametrized with median values. (b) Enlarged view of average wave and standard error (gray) of the averaged responses for the 300-ms delay, 100-ms ramp condition.

V. RESULTS

Figure 4 shows representative averaged data from one subject as well as simulated data using the mathematical model of Fig. 6, to be discussed later. Of the 138 averaged responses recorded and analyzed, 134 were ''compensatory'' and 4 were ''following'' (the *F*0 changed in the same direction as the stimulus). All 4 of the ''following'' responses occurred when the stimulus ramp duration was 1000 cents/s. Two occurred under DAF of 200 ms and two under DAF of 300 ms. There were 83 cases in which the end of a clearly defined phasic peak could be distinguished from the prolonged *F*0 response (see Fig. 1). Thus we were able to measure phasic response onset variables from all 138 responses, while we could analyze ''end of phasic response'' variables from only 83 cases. Figure 4(b) illustrates the averaged wave from the 300–100 condition along with the standard error of the wave making up the average.

For statistical analysis, latency, peak time, phasic peak duration, and peak magnitude were evaluated with respect to the independent variables. A log transformation of peak magnitude, peak time, and peak duration was done to achieve homogeneity of variance. Offset slope and phasic peak end point were highly correlated with magnitude and duration, so they were not evaluated statistically. There was a main effect for delay ($F_{\text{approx.}}=2.70$, $df=8146$, $p<0.005$) but not for stimulus onset velocity. The DAF condition primarily affected the later parts of the response, as was the case with experiment 1, while there was a tendency for the stimulus velocity to affect earlier parts of the response. Fig-

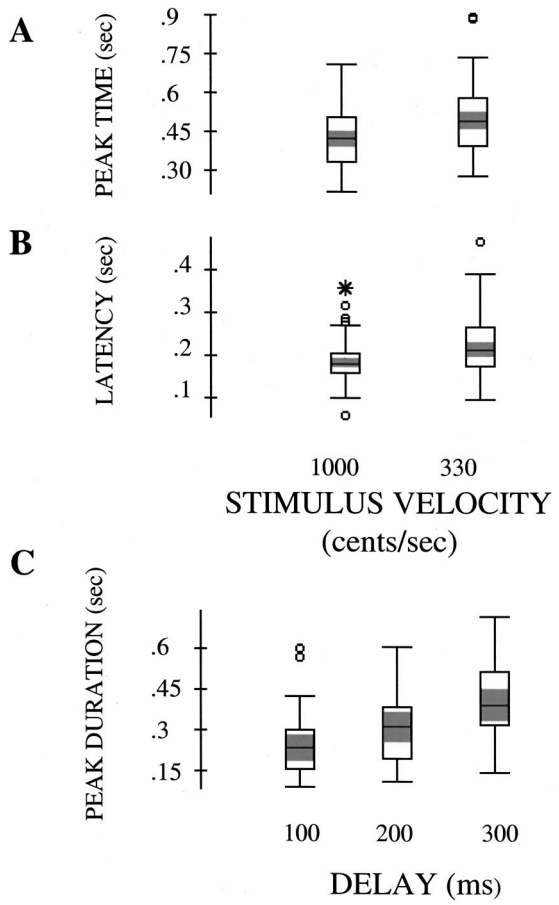


FIG. 5. (a) Boxplots for experiment 2 of peak time (A) and latency (B) as a function of stimulus onset velocity and phasic peak duration (C) as a function of feedback delay.

ures 5(a) and (b) show boxplots of peak time and response latency. Although not significant, peak time increased from 424 ms in the 100 cent/s condition to 506 ms in the 330 cent/s condition. Latency increased from a mean of 186 ms in the 1000 cent/s condition to 222 ms in the 330 cent/s condition. The overall mean latency was 204 ms (s.d.=63). Boxplots in Fig. 5(c) illustrate a progressive increase in duration of the phasic response as the delays became longer. There were significant differences between the 300 ms (415 ms) and 100 ms (256 ms) ($p<0.001$) and 300 and 200 ms (309 ms) ($p<0.05$) comparisons. Mean peak magnitude was weakly affected by the DAF interval, with a decrease from 39 cents in the 100 ms delay to 34 ms with the 300 ms delay ($p<0.05$). The overall mean response magnitude was 36.8 cents (s.d.=20.5 cents).

$$\frac{a \cdot s}{b \cdot s^2 + c \cdot s + 1}. \tag{2}$$

Again we fit parameters of the model of Fig. 6 to the experimental data and determined the values for three filter parameters and the delay parameter that minimized the variance between the experimental data and simulation. The second-order filter of Eq. (1) did not do nearly as well in reproducing data and the average vaf was less than 0%. A less constrained second-order filter [Eq. (2)] still having three parameters did better. When filter parameters were allowed to vary

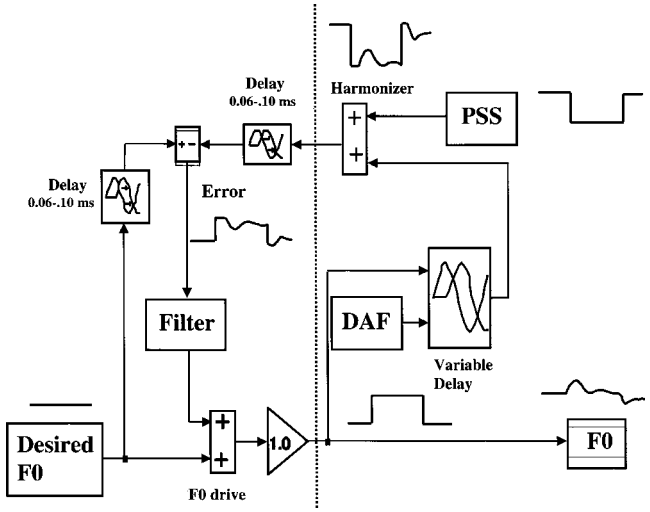


FIG. 6. (a) Mathematical model of compensatory responses composed of an open-loop pathway (Desired $F0 \rightarrow$Gain$\rightarrow F0$), and a feedback pathway (upper portion of diagram). Dynamics are implemented with two matched delays, and a second-order filter [see Eqs. (1) and (2) in the text for details]. Components to the left of the dotted line represent the person and that to the right represent the model.

for each condition, fits were obtained with a mean vaf of 60.3%. On using median values of $a=4.31$, $b=14.9$, $c=70.7$, and delay=0.1 s, the vaf was only 8%.

VI. DISCUSSION

The simultaneous presentation of PSS and DAF is a paradigm in which subjects begin each vocalization trial able to hear their voice feedback in real time. It is only with the onset of the PSS that there is introduction of a delay in the processing of the subject's voice (50–500 ms). The effect is that the PSS is heard by the subject as a change in pitch of the feedback voice signal, but perception of the subject's response is delayed until after the particular delay interval. Thus, the subject hears the PSS in real time but does not immediately hear his/her own response. Subjects were not told ahead of time that their responses would be delayed. Also, the delay was not apparent to the investigators, and it is therefore unlikely that the subjects noticed it and responded to it. Analysis of Fig. 2 illustrates the primary findings of this study; the delayed feedback prolonged the duration of the phasic peak response. Combining a delay with a reduction in velocity of the stimulus also led to an increase in latency and reduction in magnitude of the response.

A key assumption of this study is that in the DAF condition subjects could not hear their true vocal output. While we cannot be certain that their unshifted, undelayed productions were masked by the processed auditory feedback signal relayed via the headphones, we are reasonably sure that whatever perception was present was not a major influence on the responses. The addition of 70 dB SPL masking noise to the processed signal helped reduce perception of the unprocessed signal via bone conduction. Voice loudness was quiet, approximately 70 dB SPL, while the fed back voice loudness was louder, at approximately 80 dB SPL. This difference, as well as the structure of the headphone padding over the ears, reduced hearing of any side-tone signal. In

preliminary tests, the experimenters themselves failed to detect the bone conducted or side-tone signal under these conditions. We are thus reasonably confident that subjects did not hear their own voice response during the delay period.

The model is an elaboration of a structure originally proposed elsewhere (Hain *et al.*, 2000; Larson *et al.*, 2000). In brief, the input to the model is the desired *F*0 and the output is vocalization (*F*0). Internal delays are postulated to represent known response latencies, and the harmonizer/DAF circuitry is implemented with a variable delay. After *F*0 error (disparity between intended and auditory *F*0) is computed, the result is processed with a filter and used to adjust the *F*0 drive signal.

The filter element (Fig. 6) is critical to simulate details of the experimental data. For the step responses of experiment 1, a second-order filter consisting of two cascaded first-order dynamic elements was used [see Eq. (1)]. The first element is a simple low-pass filter, $fgain/sTc+1$, having a median time constant (Tc) of 0.17 s. The second is an adaptation operator, of the form $sTa/sTa+1$, with a median time constant (Ta) of 0.235 s. The adaptation operator causes the overall filter response to a step input to eventually decay to 0, and greatly improved the performance of the filter in simulating experimental data compared to a low-pass filter alone. The frequency response of the entire filter is that of a band-pass with a broad peak centered at 5 Hz where the phase shift is 0, and both low- and high-frequency roll offs. Others have also described the vocal control system in terms of a second-order system (Kawahara, 1995; Ternström *et al.*, 1989).

For the ramp responses, the filter of Eq. (1) did not perform well in fitting experimental data, but the less constrained second-order system as shown in Eq. (2) performed reasonably well. While both Eqs. (1) and (2) are second-order systems, Eq. (1) being a degenerate form of Eq. (2), they differ in that the roots of the second-order system can take on imaginary values in the form of Eq. (2). This allows Eq. (2) to produce responses exhibiting damped oscillations, while Eq. (1) can only produce responses consisting of the sum of two exponentials. The frequency response of the entire filter is again a band-pass filter, but with a higher center frequency than that of Eq. (1) (approximately 10 Hz).

There are several possible reasons why the best-fit models for the step and ramp inputs might be different. The ramp responses were longer than the step responses, due to the more prolonged stimulus. A longer stimulus might allow for participation of other tracking mechanisms, and therefore not be as well fit by a model that reflects early processing. Another possibility is that the ramp stimulus may invoke vocal control systems that make the response more complex than the step. At present there is insufficient data to decide which of the two second-order filters is more accurate. While this mathematical model is doubtless oversimplified and similar behavior might be produced by other constructs, it does demonstrate its feasibility and provides a quantitative hypothesis that can be tested in future experiments.

Returning to the data, the inclusion of the DAF conditions did not affect measures taken from the initial aspects of the responses. The latency, onset slope, and peak time were unchanged by DAF. The model of Fig. 6 reveals that this immunity results from the approximately 100-ms delay between the auditory stimulus and the *F*0 response, adding further evidence that the response to the PSS in at least the first 100–150 ms after the PSS is driven by input during the 100 ms preceding the *F*0 response (indicated by ''*'' in Fig. 2).

There was a nonsignificant trend for greater DAF delay to be accompanied by greater peak magnitude in experiment 1. If significant, this difference would have suggested that for situations where an auditory-vocal disparity is artificially maintained, subjects could have continued to change voice *F*0 indefinitely; a ''runaway'' situation. Instead, it appears that some factors other than auditory input limit the magnitude of the response. These factors might simply include dynamics that adapt to input disparity such as we have proposed in the filter element of our model, or nonauditory sources of feedback, e.g., efference copy or proprioceptive feedback.

The reduced response magnitude observed in experiment 2 (mean=35.8 cents) compared to that in experiment 1 (mean=69 cents) suggests that lower stimulus onset velocities under DAF conditions cause a reduction in response magnitude. Previous findings from experiments without DAF (Larson *et al.*, 2000) suggested that response magnitude *increased* modestly with decreases in onset velocity, and it was suggested that lower stimulus onset velocities might be more effective in eliciting responses to PSS (Larson *et al.*, 2000). Results of the present study indicate this relationship does not apply under DAF conditions. The model of Fig. 6 also supported the present findings, as simulated response magnitude for experiment 2, declined with decreasing stimulus onset velocity. The response decline was caused by the low-frequency rolloff characteristic of the filter of Eqs. (1) and (2).

The features of the PSS response that were affected by the delay conditions were related to the duration and offset of the response or, in other words, later aspects of the response. One clear effect was the increase in phasic peak duration with the DAF interval (Figs. 2, 4, and 5). The event that apparently terminated the phasic peak response was the response to the delayed perception of the initial *F*0 change (marked by asterisks in Fig. 2). This behavior was also replicated in mathematical simulations, and is an example of instability. As the delay period decrease, the duration of the phasic peak change in *F*0 is inappropriately extended. While the instability was not severe, it may be concluded that subjects' sustained perception of pitch-shifted feedback led to a prolonged voice *F*0 response. This again confirms that the voice *F*0 response is achieved through a negative feedback circuit rather than through a more primitive stereotyped reaction to a change in auditory input (Sapir *et al.*, 1983).

In experiment 2, the DAF condition had no effect on the early components of the response; however, PSS onset velocity combined with DAF had a weak affect on these components. There was a decrease in latency and peak time as the PSS stimulus onset velocity increased. Also, responses in experiment 2 had longer overall latencies than those observed in experiment 1, in which sudden pitch-shift onsets were used. This finding is in agreement with previous find-

ings in which low velocity stimuli elicited responses with long latencies (Larson *et al.*, 2000).

To summarize, the results of experiment 1 showed that when a subject's audition of their response to a step of disparity between their intended *F*0 and a fed-back *F*0 is delayed, there is a prolongation of the response, but no change in magnitude. At longer delays, e.g., >300 ms, the effect is an increase in instability of voice *F*0. From experiment 2, when the pitch shift onset velocity is reduced forming a ramp stimulus, along with the introduction in the delay of the subject's response to it, there is likewise an increase in the duration of the subject's response. With reduced pitch shift velocity, latency increases and magnitude declines. Mathematical modeling of the response using a second-order filter reproduced the step responses well, but was not as able to reproduce ramps as accurately. This suggests that other processes not reflected by our mathematical model may complicate the longer responses.

ACKNOWLEDGMENTS

We gratefully acknowledge the assistance of Rokyn Akhavein, who wrote computer programs for data analysis, and of Dr. Mary Kay Kenney and Danielle Lodewyck for help in data analysis. This research was supported by NIH Grant No. DC02764-01.

Burnett, T. A., Freedland, M. B., Larson, C. R., and Hain, T. C. (**1998**). ''Voice *f*0 responses to manipulations in pitch feedback,'' J. Acoust. Soc. Am. **103**, 3153–3161.

Burnett, T. A., Senner, J. E., and Larson, C. R. (**1997**) ''Voice *f*0 responses to pitch-shifted auditory feedback; a preliminary study,'' J. Voice **11**, 202–211.

Hain, T. C., Larson, C. R., Burnett, T. A., Kiran, S., and Singh, S. (**2000**) ''Instructing participants to make a voluntary response reveals the presence of two vocal responses to pitch-shift stimuli,'' Exp. Brain Res. **130**, 133–141.

Kawahara, H. **(1995)**. ''Hearing voice: Transformed auditory feedback effects on voice pitch control,'' presented at Computational Auditory Scene Analysis and International Joint Conference on Artificial Intelligence, Montreal.

Kawahara, H., and Aikawa, K. (**1996**). ''Contributions of auditory feedback frequency components on *f*0 fluctuations,'' J. Acoust. Soc. Am. **100**, 2825.

Larson, C. R., Burnett, T. A., Freedland, M. B., and Hain, T. C. **(1997)**. ''Voice *f*0 responses to manipulations in pitch feedback stimuli,'' presented at First International Conference on Voice Physiology and Biomechanics, Evanston, IL.

Larson, C. R., Burnett, T. A., Kiran, S., and Hain, T. C. (**2000**). ''Effects of pitch-shift onset velocity on voice *f*0 responses,'' J. Acoust. Soc. Am. **107**, 559–564.

Larson, C. R., Carrell, T. D., Senner, J. E., Burnett, T. A., and Nichols, L. L. **(1995)**. ''A proposal for the study of voice of control using the pitch shifting technique,'' in *Vocal Fold Physiology: Voice Quality Control*, edited by O. Fujimura and M. Hirano (Singular, San Diego), pp. 321–331.

Larson, C. R., White, J. P., Freedland, M. B., and Burnett, T. A. **(1996)**. ''Interactions between voluntary modulations and pitch-shifted feedback signals: Implications for neural control of voice pitch,'' in *Vocal Fold Physiology: Controlling Complexity and Chaos*, edited by P. J. Davis and N. H. Fletcher (Singular, San Diego), pp. 279–289.

Sapir, S., McClean, M. D., and Larson, C. R. **(1983)**. ''Human laryngeal responses to auditory stimulation,'' J. Acoust. Soc. Am. **73**, 315–321.

Ternström, S., and Friberg, A. **(1989)**. ''Analysis and simulation of small variations in the fundamental frequency of sustained vowels,'' Report No. 3, Speech Transmission Laboratory, Royal Institute of Technology, Stockholm.

Dynamic consequences of differences in male and female vocal tract dimensions

Adrian P. Simpson[a)]
Institut für Phonetik und Digitale Sprachverarbeitung, Universität Kiel, 24098 Kiel, Germany

(Received 15 August 1999; accepted for publication 4 January 2001)

Phonetic differences between male and female speakers are generally considered in terms of the static acoustic and perceptual consequences of different articulatory dimensions. This article investigates the dynamic acoustic and articulatory implications of differences in mean male and female vocal tract dimensions. The temporal acoustic consequences of time-varying twin-tube resonators of different dimensions are explored, and the possible implications for human speech production are considered. Empirical support for the theoretical predictions is sought by investigating the kinematic and acoustic patterns in diphthong productions from 26 female and 22 male speakers in the *University of Wisconsin X-ray Microbeam Speech Production Database*. Aside from expected acoustic differences, the shape of male and female formant tracks plotted in Bark space is found to be very similar. Male and female patterns of tongue movement, however, are found to be very dissimilar. The mean male diphthong, defined by the tracks of four midsagittal pellets, is characterized by greater pellet excursions, higher pellet speed, and consistently larger dorso-palatal strictures than its female counterpart. The empirical findings suggest that gender-specific dynamic behavior could be an important factor in accounting for nonuniform vowel system differences, but at the same time having more wide-ranging implications for transitional phenomena and undershoot. [DOI: 10.1121/1.1356020]

PACS numbers: 43.70.Aj [AL]

I. INTRODUCTION

It has long been recognized that it is not possible to derive the formant values of male vowels from their female equivalents (or vice versa) by using a single scale factor based on an average female–male vocal tract length difference of 20%. Observations for a number a languages have shown that the relationship is nonuniform, requiring separate scale factors for each formant and vowel category. In particular, the average female F1×F2 vowel space has been found to be consistently larger than its male counterpart. Reasons which have been proposed to account for these differences form four main groups:

(i) anatomical,
(ii) resonance properties of different vowel classes,
(iii) various interactions between voice source and formant frequency, and
(iv) sociophonetic.

Fant (1966, 1975), analyzing vowel data from a number of languages, identifies certain cross-linguistic similarities in the nonuniform relationship between male and female vowel qualities. The size of the scale factor found for unrounded vowels grows with openness of the vowel, frequently exceeding 20% for open vowels. For rounded vowels the scale factors are in general smaller than those found for unrounded vowels. Fant suggests that the nonuniformity is brought about in part by anatomical differences between male and female vocal tracts, male speakers having larger laryngeal cavities and a proportionally longer pharynx than female speakers (Chiba and Kajiyama, 1941), which explains in particular the larger scale factors needed for open vowel categories. Conversely, the relatively small scale factors required to map male close back rounded vowels onto their female counterparts can be accounted for by resonance behavior akin to a double Helmholtz resonator. Unlike a twin-tube model, a Helmholtz resonator is sensitive to changes in volume (rather than length) as well as changes in the length and diameter of its neck. Transferring this to the vocal tract, Fant claims that female speakers may be compensating by creating longer, tighter dorso-velar and bilabial strictures, effecting a lowering of formant frequencies in the direction of the male values.

Attempts at reproducing observed nonuniformity with vocal tract models that incorporate the disproportionate differences in oral to pharyngeal cavity length have had limited success. Nordström (1977) changes oral-pharyngeal cavity length ratios as well as vocal tract diameter. Goldstein (1980) brings together growth data from a wide range of sources which allow modelling of both vocal tract growth and gender-specific differences. While both Nordström's and Goldstein's models succeed in producing nonuniform differences for different vowel categories, both researchers interpret their results as accounting only partially for observed male–female differences.

A number of interactions between voice source and formant frequency have also been proposed. These can been divided into quantitative (relationships between f_0 and formant frequency) and qualitative (phonation type and formant frequency) accounts.

It has been suggested that the larger female acoustic vowel space, reflecting vowel categories which are spaced

a)Electronic mail: as@ipds.uni-kiel.de

further apart, may result from the need to compensate for higher fundamental frequency (Goldstein, 1980; Ryalls and Lieberman, 1982; Diehl *et al.*, 1996). As the harmonic spacing increases with a rising f_0, spectral sampling by the harmonics worsens. If a vowel system being produced with a high f_0 is to maintain the same degree of auditory distinctivity as a system being produced with a low f_0, compensation can be made by increasing the distance between neighboring vowel categories. In a series of forced-choice listening experiments Diehl *et al.* (1996) presented subjects with steady-state [ɪ] and [ʊ] stimuli, synthesized at a range of fundamental frequencies using mean male and female formant data from Peterson and Barney (1952). Their results in general support the hypothesis that identification will decrease as a function of increasing f_0, but there are problems transferring their results to normal speech situations. First, normal speech does not contain acoustically monotone stretches either from the point of view of the source or the filter: both are in constant flux. While House (1990) has studied the way in which pitch perception changes as a function of the degree of spectral change, the extent to which spectral sampling becomes less efficient with an overall higher f_0 when both filter and source are constantly changing has yet to be tested. Second, a number of studies (Titze, 1989; Henton and Bladon, 1985; Holmberg *et al.*, 1988; Klatt and Klatt, 1990) for varieties of British and American English have shown consistent qualitative voice source differences, with female speakers tending to use breathier voice quality than male speakers. Again, it is not clear to what extent this difference is biologically predetermined by anatomical and physiological differences in the vocal folds as suggested by Titze (1989) or whether it is the result of different learned behaviors (Henton and Bladon, 1985). One important acoustic characteristic of breathy voice is that the higher degree of noise in the spectrum would at least partly compensate for poorer harmonic sampling resulting from a higher f_0.

Maurer *et al.* (1991) measured the formant frequencies of adults and children producing isolated vowel utterances at the same pitch. Their results lead them to make the much stronger claim that speakers produce the same formant frequencies for a vowel regardless of differences in anatomical dimension. There are two ways of interpreting these results:

(1) Speakers change the articulatory configuration of a vowel to accommodate changes in f_0, positioning resonance frequencies appropriately with respect to the harmonics.
(2) The energy maxima being measured[1] at higher pitch levels represent a single harmonic located closest to a vocal tract resonance.

Several studies show a relationship between f_0 and F1 that gives rise to nonuniform vowel- and formant-specific differences (Traunmüller, 1984, 1985, 1988; Bladon *et al.*, 1983). F1 is seen to move to accommodate the closest harmonic. Only when F1 is sufficiently high with respect to f_0 [Bladon *et al.* (1983) propose $F1 > 3.5 \times f_0$] does F1 remain constant. In a series of demonstrations using synthetic stimuli, Traunmüller (1985) shows, among other things, how increases in f_0 and F1, which maintain the same tonotopic distance, do not change the perceived vowel quality, i.e., even large changes in F1 do not change the degree of vocalic openness, as long as f_0 is changed to follow F1.

A number of studies (Chiba and Kajiyama, 1941; Potter and Steinberg, 1950; Bladon *et al.*, 1983; Syrdal, 1985; Syrdal and Gopal, 1986; Traunmüller, 1984, 1988) suggest that nonuniformity may in part result from speakers striving to produce similar auditory excitation patterns along the basilar membrane. Bladon *et al.* (1983) propose a normalization procedure, which, following a spectral transformation similar to that carried out by the auditory periphery, allows male and female formants to be mapped onto each other by a simple 1 Bark shift along the basilar membrane. Traunmüller (1984, 1988) shows that the interpeak tonotopic distances stated in Bark between f_0, F1, F2, and F3 exhibit gross interspeaker similarity.

Most of the studies cited thus far admit that their success in accounting for intergender differences is only partial. Most, if only in passing,[2] attribute some of the difference to gender-specific behavioral differences. Several decades of sociolinguistic research (e.g., Labov, 1972, 1990; Eckert, 1989; Milroy and Milroy, 1993) have shown that when performing the same linguistic activity (reading word and sentence lists, interview, informal conversation), female and male speakers score differently with respect to syntactic, morphological, and phonological variables. Females have been found to more readily adopt incoming forms from a prestige variety as well as to be the chief innovators in linguistic change happening within a variety, while males are found to use nonstandard forms more frequently and are slower to adopt the newer forms of linguistic change within a variety (Labov, 1990).

Mean Bark distances between male and female vowel systems for a number of languages (Bladon *et al.*, 1983; Henton, 1992, 1995) show that at least some of the intergender difference must be attributed to different behavioral patterns. Data taken from French, Swedish, three varieties of English, and two varieties of Dutch exhibit mean Bark distances between male and female vowel systems ranging from 1.2 Bark (RP English) to as little as 0.56 Bark (Utrecht Dutch). This shows that for the different languages, or indeed for different varieties of the same language, males and females are behaving differently to achieve greater acoustic similarity (Utrecht Dutch), or to accentuate the intergender difference (RP English).

One factor which would appear to have been neglected when trying to account for acoustic differences between male and female vowel systems is time. Once formant measures have been extracted from the temporal midpoints of a number of vowel tokens, averages calculated, and F2×F1 plots made, the effects of the temporal stringencies imposed by utterance dynamics have long since disappeared. If, however, differences in male and female articulatory dimensions mean that a particular vowel target could be reached in different amounts of time using the same articulatory speed, or, conversely, that equivalent perceptual targets could be attained using different articulatory speeds, then this would have a number of interesting theoretical and practical consequences,

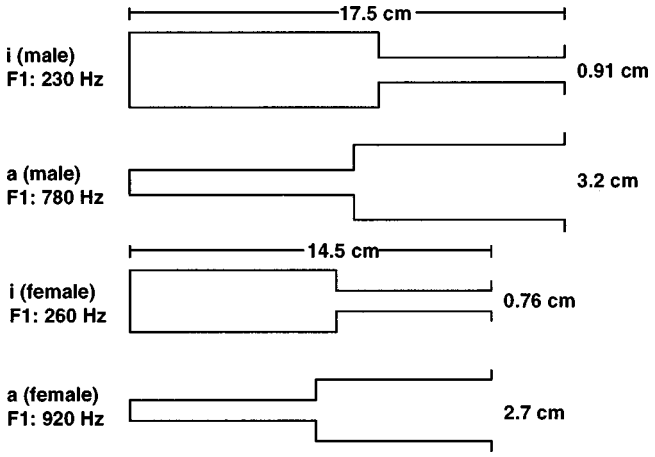


FIG. 1. Twin-tube resonators having [i]- and [a]-like formant frequencies. The top two approximate average length of the male vocal tract, the bottom two those of mean female vocal tract. In contrast to Fant (1960) tube lengths and diameters have been scaled to the same proportions.

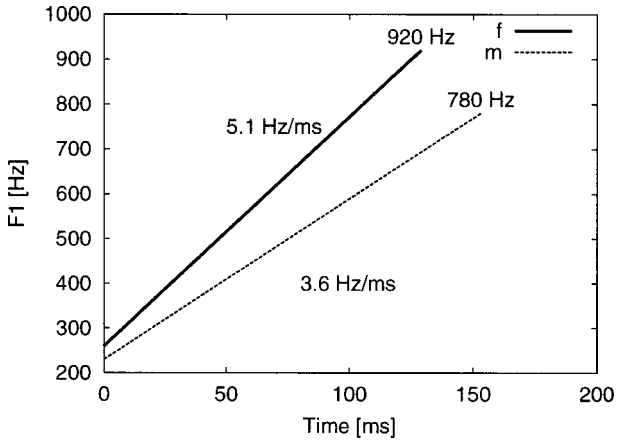


FIG. 2. F1 path from [i] to [a] for the twin-tube resonators in Fig. 1. A nominal rate of change from the [i]- to [a]-configuration of 150 mm/s is assumed. At this speed the smaller tube model reaches the open F1 target in 129 ms, the larger in 153 ms.

one of which being the acoustic dimensions of male and female vowel systems.

The temporal acoustic consequences of varying articulatory dimensions can be illustrated by means of a simple experiment using a slight modification of Fant's (1960) twin-tube resonators. Figure 1 shows twin-tube resonators approximating [i] and [a] for tube lengths of 17.5 and 14.5 cm, respectively. However, in contrast to Fant's model, not only have the tube lengths been scaled down proportionately, but also the tube diameters. The diameters of the 17.5-cm tubes are based on Fant's estimates of male cross-sectional areas of Russian vowels from x-ray photographs (Fant, 1960, p. 115) and those of the 14.5-cm tubes have been scaled down proportionately ($2.7 = 3.2 \times 14.5/17.5$). As the resonant frequencies are dependent on (a) the total length of each twin-tube resonator, (b) the ratios of lengths of the two tubes, and (c) the ratios of their cross-sectional areas, scaling down the cross-sectional area has no effect on the resonant frequencies themselves. The acoustic consequences of scaling diameters and lengths manifest themselves once movement from the close vowel configuration to the open vowel configuration is considered.

Figure 2 plots a linear movement of F1 assuming an appropriate change from the [i]-like configuration to the [a]-like configuration of the malelike and femalelike resonators. The time scale for the completion of the configuration change is based on a realistic nominal articulatory speed of 150 mm/s taken from pellet speeds presented later in this study. At this speed the movement from the [i] to [a] stricture will be completed in 153 ms for the malelike configuration and in just 129 ms for the femalelike configuration. It is also possible to see from Fig. 2 that the same rate of movement in the smaller resonator tube dimensions not only means that the [a]-like configuration is reached more quickly, but also that the rate of change in F1 at 5.1 Hz/ms is higher than for the larger resonator (3.6 Hz/ms).

Extrapolating from the results of such a simple model to human vowel production must be done with great care. Aside from the obvious limitations of modeling vowel resonance with a twin-tube model, any number of other complicating factors such as acceleration and inertia have been neglected. However, experimenting with this simple model does reveal at least four possible acoustic and articulatory consequences of differences in the average dimensions of male and female vocal tracts. The implications are considered from the perspective of the smaller female vocal tract:

(1) A female speaker operating with the same set of vocalic resonance configurations, the same articulatory speed, and within the same temporal structure as a male speaker should exhibit less qualitative undershoot, as a vowel target can be reached in less time. This difference would be negligible for close vowels, becoming increasingly significant with the degree of opening, and would presumably be less for long quantity vowels whose duration is greater, *ceteris paribus*, than that of short quantity vowels.
(2) A female speaker operating with the same set of vocalic resonance configurations and the same articulatory speed as a male speaker should be able to reduce overall vowel duration while maintaining the same degree of undershoot as a male speaker.
(3) A female speaker operating with the same set of vocalic resonance configurations and within the same temporal structure as a male speaker could reduce her articulatory speed, reaching the same resonance configuration in the same time period as a male. As in the previous example the female and male speakers would exhibit similar degrees of undershoot.
(4) A female speaker operating with the same articulatory speed and within the same temporal structure as a male speaker could maintain greater differentiation in her acoustic vowel space by traveling a greater articulatory and acoustic distance in the same time frame. The difference would increase with the degree of vocalic opening. One consequence of maintaining a more widely dispersed vowel space within the same temporal confines would be similar amounts of male and female undershoot.

These consequences of differences in vocal tract dimen-

sions fit in well with both the observation that female–male vowel mappings are nonuniform, as well as the observation that the female vowel space is on average larger than its male counterpart. As can be seen from Fig. 2, the further we move along the time axis, the greater the acoustic distance between between the F1 values, which in turn would predict greater differences between male and female open vowels than close vowels. If, in accordance with the second consequence, female speakers take advantage of the extra time by aiming at more distant targets, then the product is a greater vowel space, compensating for the poorer sampling of the spectral envelope by more widely spaced harmonics. Interestingly, both less undershoot and a greater acoustic vowel spacing are features which can be associated with more careful articulation, but are achieved by operating within the same temporal and articulatory constraints as a male speaker, i.e., without investing any extra time or energy. In other words, female speakers may be conforming to a stereotype of carefulness often associated with their speech, without there being any intention to actually speak more carefully.

However, the fact that the predictions being made here are in line with many of the acoustic observations made in previous studies does not in itself provide hard evidence that what we are observing are the temporal acoustic consequences of different articulatory dimensions. This study seeks to provide such evidence by investigating the relationship between articulatory and acoustic patterns in American English diphthongs. Diphthongs have been chosen as they represent a convenient environment to observe changes in vocal tract configuration which give rise to smooth, easily measurable acoustic patterns.

The results suggest that, at least for the read diphthongs investigated, male and female speakers attain similar perceptual targets—displaced in absolute acoustic terms in line with the differences in articulatory dimension. However, for the male speakers this requires higher articulatory speeds, greater lingual excursion, and larger dorso-palatal strictures. In addition, it is argued that longer female diphthong durations, which were initially unexpected (see predictions made earlier in this work), but which are consistent with the findings of a large survey of German vowels, may be directly related to the observed differences in articulatory speed.

II. DATA

The *University of Wisconsin X-ray Microbeam Speech Production Database* (Westbury, 1994, henceforth *UWDB*) contains synchronous recordings of articulatory data and acoustic signals of 26 female and 22 male speakers carrying out a range of linguistic (reading telephone numbers, word lists, and short texts) and nonlinguistic (swallowing, tongue protrusion) activities. Speakers and tasks will be referred to as they are on the publicly available database, i.e., speakers with the prefix ''jw'' followed by two or three digits; tasks with the prefix ''tp'' followed by three digits. The speakers were between 18 and 37 years of age. This speaker sample can be treated as being dialectally homogeneous. The majority speak an Upper Midwest dialect of American English, with more than 80% having a dialect base in Wisconsin (28 speakers) or an adjacent state. This homogeneous treatment, at least for the data under consideration, is supported by the formant data for diphthong tracks below.

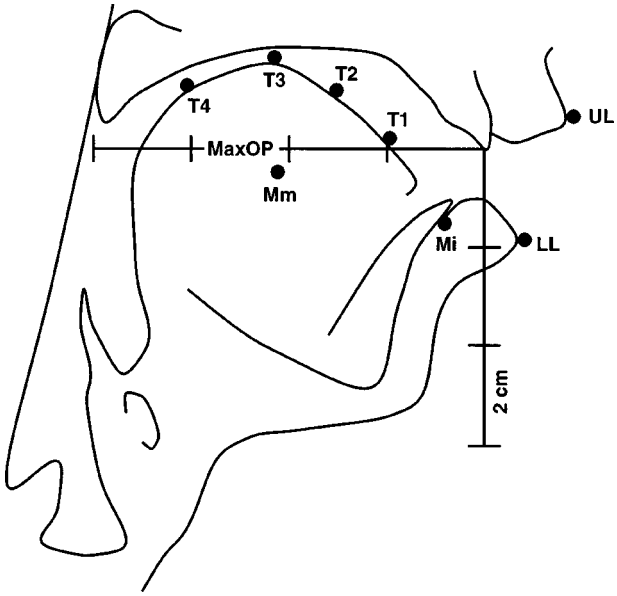


FIG. 3. Schematic representation of pellet placement (from Hashi *et al.*, 1998—reproduced with kind permission of the authors and the Acoustical Society of America).

The articulatory data in the *UWDB* are the two-dimensional coordinates of eight gold pellets captured using the X-ray microbeam technique (Fujimura *et al.*, 1973; Westbury, 1991). Four pellets are positioned equidistantly along the midline of the tongue, two on the lower jaw, and one each on the upper and lower lip. A fixed reference for these movable points is provided by three further pellets, two on the nose and one on the upper incisor. The positions of the eight pellets in the midsagittal plane are illustrated schematically in Fig. 3. The first of the lingual pellets (T1) is placed approximately 1 cm behind the tip of the tongue, the fourth (T4) as far back on the tongue dorsum as the subject would allow. The remaining two pellets are placed equidistantly between these two extremes. During recording itself pellet positions were sampled at different frequencies. Levels of radiation could be reduced by sampling slower moving articulators such as the tongue dorsum and lips at a lower rate than that required to capture tongue tip movement. However, in order to facilitate subsequent data analysis, pellet positions were furnished with synchronous time-stamps by resampling all pellet positions at 145.645 Hz. The sound pressure wave was captured using a directional microphone (Shure SM81 Condenser) and digitized at 21 379 Hz. For a detailed description of technical setup, recording, and elicitation procedures, the reader is referred to Westbury (1994).

The subset of data used in this study is diphthong tokens from the word ''light'' drawn from the five repetitions by each speaker of the sentence ''The coat has a blend of both light and dark fibers.'' In addition to being repeated several times by each speaker, the diphthong in this word seemed an ideal object for observing dynamic articulatory-acoustic relationships. It was expected to exhibit large lingual and acoustic excursions. A certain degree of sentence stress to-

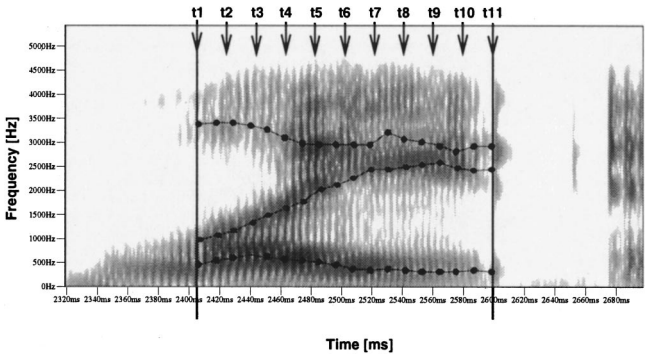


FIG. 4. Spectrogram showing delimitation of the diphthongal stretch overlaid with the results of automatic formant identification. Arrows labeled t1–t11 indicate the temporally equidistant points at which formant tracks are resampled. The extract is taken from female subject jw20 carrying out task tp019.

gether with an absence of linguistic vowel nasality both ensure readily analyzable source and filter characteristics. Because of strict time limits imposed on the recording, any task not completed within the allotted time period had to be broken off. For this reason 1 male (jw51, tp055) and 2 female (jw46, tp059 and tp069) tokens are missing, giving a total of 128 female tokens and 109 male tokens.

III. METHOD

A. Segmentation of the acoustic record

With the aid of oscillograms of the sound pressure wave and wideband spectrographic displays in *xassp* (IPDS, 1997), typical auditory and visual criteria were used to delimit each diphthong token temporally. The onset of the vocalic portion was identified by the audible release of the central closure of the lateral which is spectrally marked by an abrupt increase in amplitude of all formant frequencies. Delimiting the end of the vocalic portion was slightly more problematic, and there was both inter- as well as intraspeaker variability in the realization of the postvocalic stop. This ranged from synchronous apical closure and onset of voicelessness, to differing degrees of glottalization. The synchronous onset of voicelessness and closure is marked spectrally by a sharp attenuation of spectral energy clearly visible in the spectrogram. Glottalization was often accompanied by subsequent oral closure or the maintenance of a stricture of open approximation into the following syllable. In such cases the onset of glottalization was used as the right-hand delimiter. The vertical lines in Fig. 4 show the segmentation of a female diphthong, illustrating such a case of glottalization.

B. Automatic formant extraction

The manually segmented acoustic record provided the basis for automatic extraction of the positions and speeds of the four lingual pellets T1–T4, as well as for tracking the first three formant frequencies.

To ensure stable formant extraction all relevant signals were downsampled from original sampling frequency of 21 739 to 10 870 Hz. The delimited vocalic stretches were subjected to formant analysis using the program *formant* from *Entropic Signal Processing System*. Measurements of the first four formants were carried out using the following settings: 49-ms $\cos^4$ window, 10-ms step size between frames, and an LPC order of 12. The nominal default value of F1 was increased by 10% to 550 Hz for female speakers. The analysis times were extended by 25 ms beyond the segmented start and end times of each diphthong to ensure that the first and last analysis windows were centred around the edges of the vocalic stretch. Formant tracks for each of the 237 tokens were visually checked and any tracking errors corrected using manual measurements made from DFT and LPC spectrograms. Tracking errors mainly occurred at the onset of the vocalic portion with the identification of F4 as F3. Following correction each track was then resampled to provide 11 temporally equidistant formant records for each diphthong using linear interpolation. Eleven points provide a more than adequate definition of formant movement throughout the vocalic stretch and allows average formant tracks to be calculated from durationally different stretches. Figure 4 shows the tracks of F1–F3 calculated by *formant* and arrows indicate positions of the 11 temporally equidistant measurement points used for resampling.

C. Pellet positions

Normalized positions of the four lingual pellets T1–T4 were also ascertained at the same 11 temporally equidistant points. First, raw pellet positions were extracted for each vocalic stretch, and as with the formant records these were resampled, again using linear interpolation.

The resampled raw pellet coordinates were normalized following the procedure described in Hashi *et al.* (1998). Raw pellet positions are stated within a head-defined coordinate system (see Fig. 3). For the lingual pellets this means that the size of linguotectal stricture is ignored since positions of the pellets with respect to hard and soft palate are not available. The normalization procedure attempts to overcome this problem by redefining each pellet position in terms of

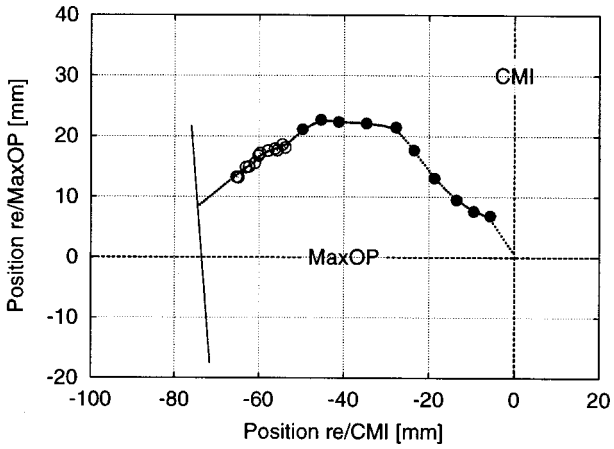


FIG. 5. Steps in the construction of the pseudo-palatal outline. Large filled circles show positions of gold pellets placed along cast taken from regular dental impression. Large open circles indicate extreme positions of dorsal pellet T4, posterior to the original palatal outline. The dotted line shows continuous outline sampled at 0.5-mm intervals following smoothing and extension forward to the origin and backward to the intersection with the rear wall of the oropharynx.

the shortest distance to an individual's palate and this in turn with respect to a mean group palate.

The normalization method requires a continuous midsagittal palatal outline which begins at the tip of the central maxillary incisors and ends at its intersection with the rear wall of the oropharynx. For each speaker in the *UWDB* the rear oropharyngeal wall is defined by a straight line segment. The palatal vault is defined by the coordinates of gold pellets laid along a plaster cast taken from a regular dental impression (large filled circles in Fig. 5). At best this cast extends only a short way along the soft palate so that parts of the roof of the mouth remain undefined. The continuous approximation to the palatal outline, shown by the dotted line in Fig. 5, is constructed as follows. The anterior portion is a straight line segment running from the tip of the central maxillary incisors to the first pellet. Extreme positions (large open circles) of the dorsal pellet (T4) identified from a number of speaking tasks extend the posterior line along the soft palate. The extended outline is resampled at 0.5-mm intervals and smoothed using a 17-point weighted average. Finally, linear regression over the last 5 mm of the outline is used to create a straight line segment which intersects with the pharyngeal line segment. A mean palatal length for the subject group under examination is then calculated from all the continuous outlines.[3]

Each raw pellet position is normalized as follows:

(1) The shortest distance between the pellet and the palatal outline becomes the new y coordinate [y' in Fig. 6(a)].
(2) The distance along the palatal outline from the origin to the location of maximal constriction is calculated [x' in Fig. 6(a)].
(3) x' is divided by the speaker's total palatal outline length and multiplied by the length of the mean palatal outline (87.45 mm).

Figure 6 shows lingual pellet positions from the second and tenth measurement points from the vocalic portion shown in Fig. 4 prior to [Fig. 6(a)] and following normalization [Fig. 6(b)].

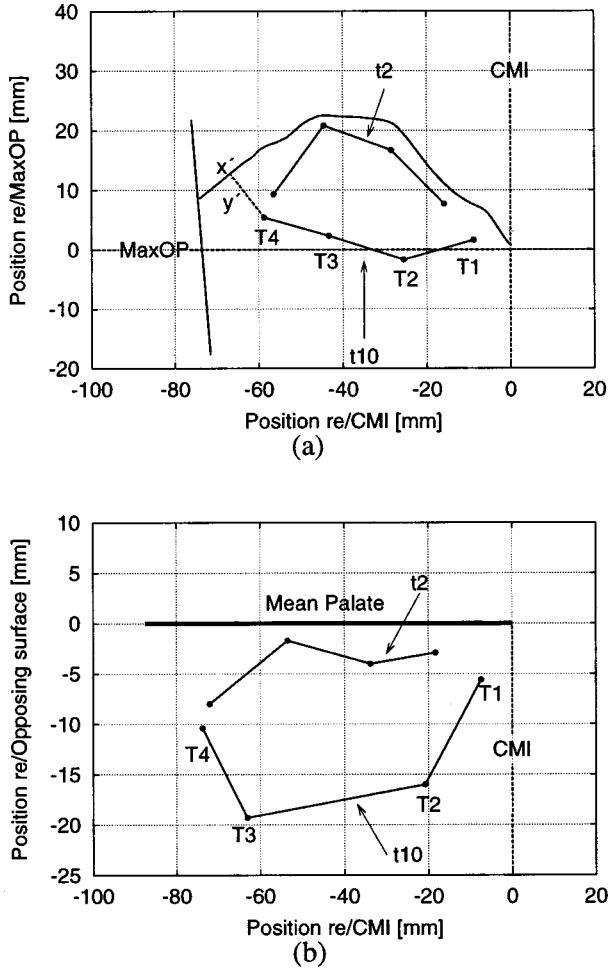


FIG. 6. Raw (a) and normalized (b) positions of the lingual pellets from the second (t2) and tenth (t10) measurement points (2427 and 2563 ms) of the vocalic portion shown in Fig. 4.

TABLE I. Means and standard deviations of male and female F1–F3 in Hertz at each of the 11 measurement points. Values rounded to the nearest integer.

	$\overline{F1}$	s	$\overline{F2}$	s	$\overline{F3}$	s
Female $n=128$						
t1	532	73	1102	118	3164	258
t2	660	64	1269	132	3144	237
t3	715	76	1428	150	3096	220
t4	738	85	1608	143	3035	212
t5	719	93	1774	142	2993	216
t6	668	85	1939	149	2962	220
t7	603	83	2103	171	2936	208
t8	541	82	2227	178	2936	208
t9	497	75	2314	177	2939	190
t10	458	64	2358	173	2952	197
t11	429	58	2321	182	2962	204
Male $n=109$						
t1	440	53	840	84	2682	272
t2	542	47	981	77	2682	235
t3	583	44	1097	76	2636	202
t4	598	45	1230	99	2579	170
t5	587	47	1386	117	2517	141
t6	553	53	1548	131	2476	126
t7	508	55	1698	139	2466	124
t8	459	54	1822	143	2482	122
t9	418	49	1909	137	2514	119
t10	383	44	1950	127	2544	122
t11	349	44	1941	125	2544	138

D. Pellet speeds

In line with formant and pellet position data, pellet speeds were calculated at the same 11 temporally equidistant points in each diphthong token. The distance traveled by a pellet in the 6.866-ms interval between two records was used to calculate interrecord pellet speeds throughout a diphthongal stretch, and subsequently linearly resampled at 11 temporally equidistant points.

IV. RESULTS

A. Duration

The female and male diphthong tokens have significantly (two-tailed t-test, $p<0.05$) different mean durations of

TABLE II. Mean rates of F1 change (and standard deviations) for two stretches during the glide phase of the diphthong portions, expressed in Hz/ms and Bark/ms.

	Female ($n=128$)				Male ($n=109$)			
	Hz/ms		Bark/ms		Hz/ms		Bark/ms	
	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s
t5–t8	3.68	1.392	0.0288	0.0104	2.88	0.845	0.0246	0.0075
t4–t11	2.78	0.829	0.0226	0.0066	2.43	0.574	0.0216	0.0052

165 ms ($s=31.6$) and 151 ms ($s=20.7$), respectively. The longer female diphthong duration of 8.9% is in line with vowel duration differences found in a large survey of German monophthongs and diphthongs (Simpson, 1998). Possible reasons for this difference will be considered in Sec. V, once the articulatory results have been presented.

B. Formant tracks

Table I shows the mean and standard deviations of female and male F1–F3 in Hz at each of the 11 measurement points t1–t11. Differences in linearly scaled F1 are of particular interest here. The maximum F1 mean for both males and females is found at measurement point t4, the minimum at the last measurement point t11. The male F1 difference between these two extremes (249 Hz) is only 80% of that found for female F1 (309 Hz).

The mean rate of F1 change was calculated for two stretches, t5–t8 and t4–t11. The former gives a good estimate for the stretch over which the glide movement is at its fastest, the latter an estimate for the entire closing glide phase. Table II contains means and standard deviations of rates of change in Hz/ms and Bark/ms. Three of the differences between the male and female rates of change are highly significant ($p<0.001$), with only the difference in Bark rate of change for the entire glide stretch t4–t11 being insignificant. The largest difference between the male and female values is for the stretch t5–t8 where the female linear (Hz) rate of change is approximately 28% greater than the male value. But even this is still less than the hypothetical calculations of differences made in Sec. I.

TABLE III. Differences between male and female formant means at each of the 11 measurement points in Hertz and Bark. Figures in brackets are female–male scale factors in percent (see text). All Hertz values have been rounded to the nearest integer.

	Hertz						Bark		
	F1		F2		F3		F1	F2	F3
t1	92	(21)	262	(31)	482	(18)	0.81	1.61	1.06
t2	118	(22)	288	(29)	462	(17)	0.95	1.59	1.02
t3	132	(23)	331	(30)	460	(17)	1.02	1.68	1.04
t4	140	(23)	378	(31)	456	(18)	1.07	1.75	1.06
t5	132	(22)	388	(28)	476	(19)	1.02	1.63	1.13
t6	115	(21)	391	(25)	486	(20)	0.92	1.50	1.17
t7	95	(19)	405	(24)	470	(19)	0.79	1.43	1.14
t8	82	(18)	405	(22)	454	(18)	0.71	1.34	1.10
t9	79	(19)	405	(21)	425	(17)	0.71	1.29	1.02
t10	75	(20)	408	(21)	408	(16)	0.70	1.27	0.97
t11	80	(23)	380	(20)	418	(16)	0.76	1.20	0.99

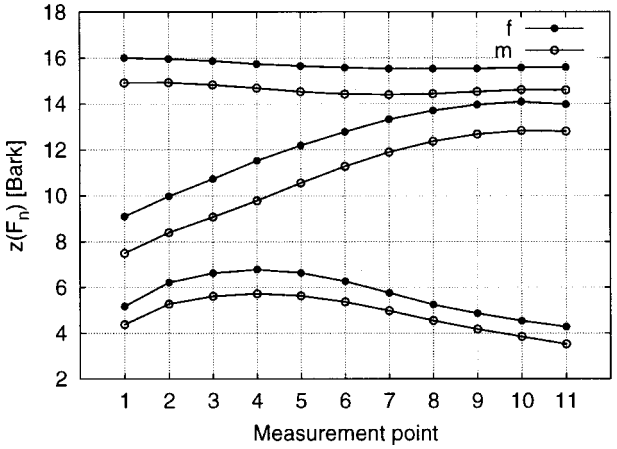


FIG. 7. Mean female and male F1–F3 values plotted as a function of measurement point in Bark.

Table III shows absolute differences between the mean female and male formant values in Hertz as well as in Bark [using Traunmüller's (1990) conversion]. Also for Hertz values each difference is represented as a scale factor using Fant's (1966) formula in Eq. (1). Similarities in the magnitude of standard deviations when seen in percentual terms of the mean would seem to provide *post hoc* support for treating subjects as a dialectally homogeneous group. This does not deny dialectal difference, but large intersubject variation in diphthong production, such as consistent monophthongal realization found in some varieties of American English, would seem to be absent.

$$k_n=\left(\frac{F_n\text{female}}{F_n\text{male}}-1\right)\times 100\%. \tag{1}$$

Visual comparison of the formant tracks using the perceptually more appropriate Bark scale are presented in Fig. 7. The relationships between male and female formant values expressed in Bark exhibit interesting differences which are presented graphically in Fig. 8. Each mean female formant value is plotted as a function of its difference from the mean male value. The tight cluster of F3 values around 1 Bark reflects the parallel course of this formant throughout the diphthong. For both F1 and F2, however, the difference in

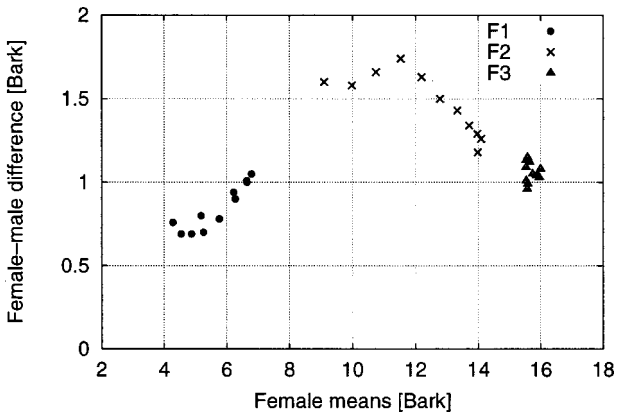


FIG. 8. Differences between mean female and male formant values at each of the 11 measurement points plotted as a function of the female means. All values in Bark.

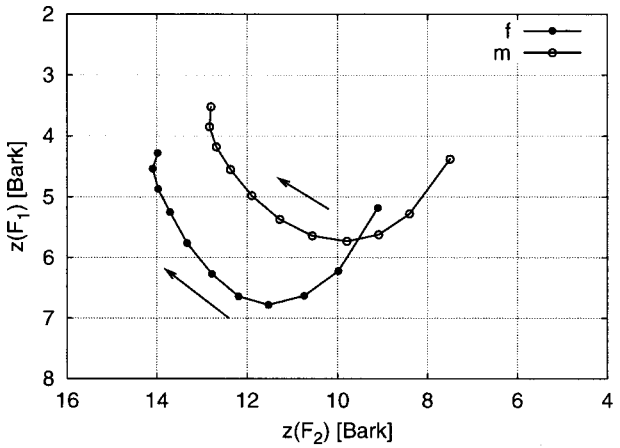


FIG. 9. Mean female (filled circles) and male (open circles) F1 values plotted as a function of mean F2 values for each of the 11 measurement points. Arrows indicate the direction of movement.

Bark between male and female values increases with vocalic openness and backness. At high values (vocalic openness), the F1 differences are more than 1 Bark, falling to approximately 0.7 Bark as F1 reaches its mean minimum (vocalic closeness). The difference between F2 values is largest of all the formants, rising to between 1.6 and 1.7 Bark at low values (vocalic openness and backness), and dropping to approximately 1.1 Bark when F2 is at its highest (vocalic closeness and frontness). Differences in F1 and F2 Bark values suggest a female diphthong which is more peripheral in perceptual vowel space. The larger overall difference in F2 may be attributable to the different pharynx-oral cavity length ratios (Chiba and Kajiyama, 1941).

Despite absolute and relative differences, it is equally important to recognize the similarities in perceptual shape which can be represented by plotting mean F1 values as a function of mean F2 values (see Fig. 9). Not only are the F2×F1 paths very similar in shape, there is also no significant difference in distance traversed through F2×F1 space, with the female and male lengths at 6.88 and 6.84 Bark, respectively.

C. Pellet tracks

Table IV contains means and standard deviations of the normalized *x* and *y* coordinates of the four lingual pellets T1–T4 at each of the measurement points. Differences in *n* originate from the failure of the microbeam system to track a pellet during part or all of the utterance containing a diphthong token. The tabulated pellet positions are summarized graphically in Fig. 10, with the ventral pellet T1 on the right, the dorsal pellet T4 on the left.

We begin by dealing with an artifact of the normalization procedure. As can be seen from Table IV the standard variations for the normalized *x* coordinates are twice to three times as great as the standard deviations for the *y* coordinates. This is in line with the findings of Hashi *et al.* (1998), who report a marked increase in variance when comparing raw and normalized *x* coordinates. The reason for the difference lies with the method for determining the normalized *x* coordinate: find the shortest distance between the pellet and

TABLE IV. Means and standard deviations of normalized *x* and *y* coordinates in mm for the four lingual pellets T1–T4 at each of the 11 measurement points t1–t11.

	T1 $n=124$				T2 $n=125$				T3 $n=124$				T4 $n=125$			
	$\bar{x}$	*s*	$\bar{y}$	*s*	$\bar{x}$	*s*	$\bar{y}$	*s*	$\bar{x}$	*s*	$\bar{y}$	*s*	$\bar{x}$	*s*	$\bar{y}$	*s*
Female																
t1	−10.0	2.3	−4.6	1.6	−21.1	6.1	−14.4	2.5	−50.9	6.4	−13.9	3.2	−66.1	7.0	−9.0	3.1
t2	−9.5	2.6	−6.5	1.6	−23.0	7.6	−14.7	2.5	−51.0	6.4	−13.6	3.1	−66.2	6.9	−9.4	3.1
t3	−9.1	2.9	−8.5	1.7	−24.0	7.8	−14.4	2.5	−50.6	6.3	−12.7	3.0	−66.1	6.7	−9.7	2.9
t4	−9.0	3.2	−10.0	1.9	−25.4	7.8	−13.4	2.5	−49.9	6.1	−11.3	2.8	−65.7	6.6	−9.7	2.7
t5	−9.7	3.3	−10.7	1.9	−26.7	7.2	−11.9	2.4	−49.3	5.7	−9.5	2.6	−65.1	6.5	−9.6	2.5
t6	−10.9	3.0	−10.5	1.9	−28.1	6.3	−10.1	2.4	−48.6	5.5	−7.7	2.4	−64.1	6.3	−9.4	2.5
t7	−11.8	2.8	−9.4	2.0	−29.1	5.7	−8.3	2.3	−48.1	5.0	−6.0	2.3	−63.1	6.2	−9.1	2.7
t8	−13.0	2.6	−7.8	2.0	−30.1	5.1	−6.7	2.1	−47.8	4.8	−4.7	2.2	−62.2	6.0	−9.0	2.9
t9	−14.1	2.5	−6.1	1.8	−30.9	4.6	−5.4	1.9	−47.5	4.6	−3.8	2.0	−61.6	5.7	−8.8	3.1
t10	−15.2	2.4	−4.4	1.7	−31.6	4.3	−4.7	1.7	−47.4	4.5	−3.4	1.9	−61.3	5.6	−8.8	3.2
t11	−16.1	2.4	−3.2	1.5	−32.0	4.1	−4.3	1.6	−47.4	4.4	−3.3	1.8	−61.1	5.6	−8.8	3.3
	$n=108$				$n=102$				$n=103$				$n=102$			
Male																
t1	−9.0	3.1	−4.9	1.7	−19.5	5.4	−15.4	2.6	−50.9	9.5	−19.3	3.1	−68.2	7.0	−14.7	4.0
t2	−8.1	3.4	−7.1	1.7	−20.5	6.3	−15.7	2.4	−51.6	8.5	−18.5	3.0	−68.3	7.0	−14.6	3.8
t3	−7.3	3.7	−9.4	1.9	−21.9	7.2	−15.5	2.1	−52.0	7.7	−17.1	2.8	−68.3	6.7	−14.1	3.5
t4	−7.3	4.1	−11.3	2.2	−23.1	7.2	−14.7	2.0	−51.8	7.2	−15.1	2.6	−67.9	6.6	−13.3	3.1
t5	−8.4	4.4	−12.2	2.4	−24.4	7.2	−13.4	1.9	−51.1	6.9	−12.8	2.4	−66.9	6.6	−12.3	2.9
t6	−9.8	4.4	−11.9	2.5	−25.7	7.1	−11.7	1.8	−50.3	6.8	−10.4	2.3	−65.9	6.5	−11.3	2.8
t7	−11.2	4.1	−10.7	2.4	−26.8	6.8	−9.8	1.8	−49.4	6.6	−8.3	2.3	−64.5	6.5	−10.3	2.9
t8	−12.5	3.6	−8.8	2.3	−27.9	6.3	−8.1	1.7	−48.7	6.5	−6.6	2.2	−63.3	6.5	−9.5	3.0
t9	−13.6	3.3	−6.6	1.9	−28.9	5.9	−6.7	1.6	−48.1	6.3	−5.4	2.1	−62.5	6.4	−9.0	3.1
t10	−14.6	3.4	−4.5	1.5	−29.6	5.6	−5.6	1.6	−47.8	6.2	−4.6	2.0	−62.0	6.5	−8.7	3.2
t11	−15.4	3.5	−3.0	1.4	−30.1	5.4	−5.1	1.5	−47.6	6.1	−4.3	1.9	−61.7	6.4	−8.6	3.2

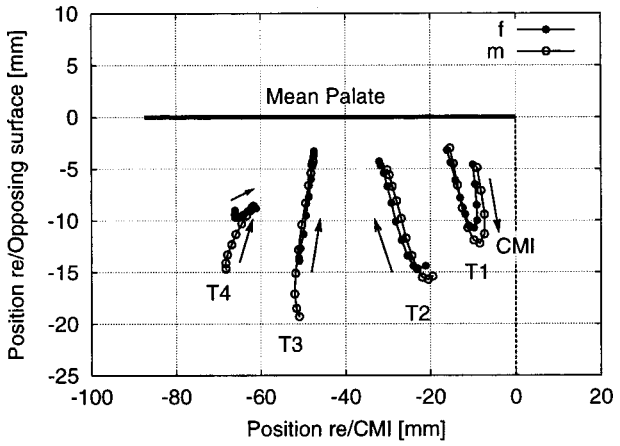


FIG. 10. Mean normalized positions of the four lingual pellets T1 (right) to T4 (left) for each of the 11 measurement points, t1–t11, for female (filled circles) and male (open circles) diphthongs. Arrows indicate direction of movement.

the palate and calculate the distance along the palate to the location of minimum constriction. This location can jump abruptly when a pellet approximates a different part of the palate. Since the normalized *y* coordinate represents the minimal stricture size, such abrupt jumps do not occur and the variances found for raw and normalized *y* values are the same.

The biggest similarity between the male and female patterns apparent from Fig. 10 is the shape of the pellet tracks for the first three lingual pellets T1–T3. This similarity is in marked contrast to the tracks of the dorsal pellet T4. Although each of the male pellets moves through a greater space than its female counterparts, the difference for T4 is disproportionate, with the female pellet showing scarcely any vertical displacement. As the most posterior pellet this possibly reflects the greatest difference between the male and female vocal tracts, namely the difference in pharynx length. The longer male pharynx may require greater vertical excursions of the tongue dorsum for open vowels, or, conversely, the relatively short female pharynx may restrict the degree of vertical freedom of the tongue dorsum which would lead to a further shortening of this cavity.

There are two further gender-specific differences to be seen in all pellet tracks. First, for the tongue-tip pellet T1 and the two posterior pellets T3 and T4 the male pellet tracks are larger, i.e., the mean path length of each pellet trajectory, is greater than its female counterpart (Table V). This is not the case for T2, although it is not clear why. Second, the largest mean pellet distances from the palate indicate a smaller female linguo-tectal stricture size when the diphthong is at its most open. At the point of minimum stricture the mean positions of the mid-ventral and mid-dorsal pellets (T2 and T3) also indicate a smaller female stricture, and for T1 and T4 there is no difference.

TABLE V. Distances in mm traveled by the four lingual pellets T1–T4 for female and male speakers measured from the mean pellet tracks in Fig. 10.

Pellet	Female	Male
T1	16.7	20.2
T2	16.0	15.8
T3	11.3	16.1
T4	5.8	9.4

TABLE VI. Mean speeds in mm/s and standard deviations of the four lingual pellets T1–T4 at the 11 measurement points for female and male speakers. Nonsignificant male-female speed differences are indicated with[ns].

	T1		T2		T3		T4	
Female	$n=124$		$n=125$		$n=124$		$n=125$	
t1	115.1	32.1	46.2	20.9	35.4	19.2	40.7[ns]	21.6
t2	136.0	34.8	48.7	20.5	46.8	25.2	40.3[ns]	21.1
t3	128.3	36.0	66.3	28.5	73.5	36.3	42.8	23.5
t4	97.3	34.6	97.5[ns]	33.8	103.5	41.6	53.8	30.3
t5	70.6	34.1	124.0[ns]	34.3	122.0	38.6	66.2	34.9
t6	80.2[ns]	40.1	135.0	30.0	122.4	31.4	71.3	33.2
t7	109.6	36.8	128.9	28.6	106.9	28.7	66.3	29.6
t8	132.3	37.2	110.6	32.5	82.1	30.2	54.6	28.2
t9	142.3	46.2	85.7	35.4	55.3	29.0	42.2	26.2
t10	133.5	54.9	59.6	34.3	34.8	23.0	32.3	22.1
t11	107.5	57.2	41.1	29.3	26.1	17.8	26.4	17.3
Male	$n=108$		$n=102$		$n=103$		$n=102$	
t1	138.0	37.8	52.4	21.2	49.7	29.6	39.7[ns]	17.3
t2	169.5	42.6	64.6	27.7	76.6	43.7	45.1[ns]	22.2
t3	166.5	41.5	83.6	33.4	116.8	46.8	59.6	33.1
t4	128.8	37.8	106.3[ns]	39.5	152.9	45.4	81.7	41.9
t5	86.9	34.0	129.9[ns]	38.4	173.9	40.7	101.7	45.2
t6	88.1[ns]	45.7	145.6	32.6	174.2	35.4	111.5	42.0
t7	129.1	43.4	146.3	28.4	154.8	34.3	106.5	37.6
t8	161.6	41.8	132.9	28.1	122.8	36.3	89.2	36.4
t9	175.9	49.4	110.2	28.8	86.6	36.8	67.2	35.9
t10	165.7	53.8	82.4	26.9	54.4	33.4	48.1	32.1
t11	133.6	47.3	58.7	23.3	36.0	27.0	38.0	26.8

D. Pellet speeds

Since the mean female pellet positions exhibit smaller excursions and this is done within a greater mean duration (see previous section), we should expect to find that the male tongue is moving faster. Means and standard deviations of interrecord speeds at each of the 11 measurement points are presented in Table VI. Differences between male and female mean speeds found to be nonsignificant ($p>0.05$) using at *t*-test have been indicated using n5. Visual comparison of mean pellet speeds can be made in Fig. 11.

The largest mean speeds are found for tongue tip pellet T1 during the movement away from and into the consonantal strictures, with a minimum being reached at the middle of the vocalic stretch. The mean speeds of the remaining three pellets exhibit their largest values around the middle of the vocalic stretch. As one might expect from differences in mean pellet positions and excursion sizes, it is the two posterior pellets T3 and T4 which exhibit the greatest gender-specific differences with T3 showing mean speeds similar to those found for the tongue-tip pellet T1.

V. DISCUSSION

The beginning of this study proposed a number of possible consequences arising from differences in vocal tract dimensions, which might underlie male–female differences. An examination of acoustic-articulatory relations in 237

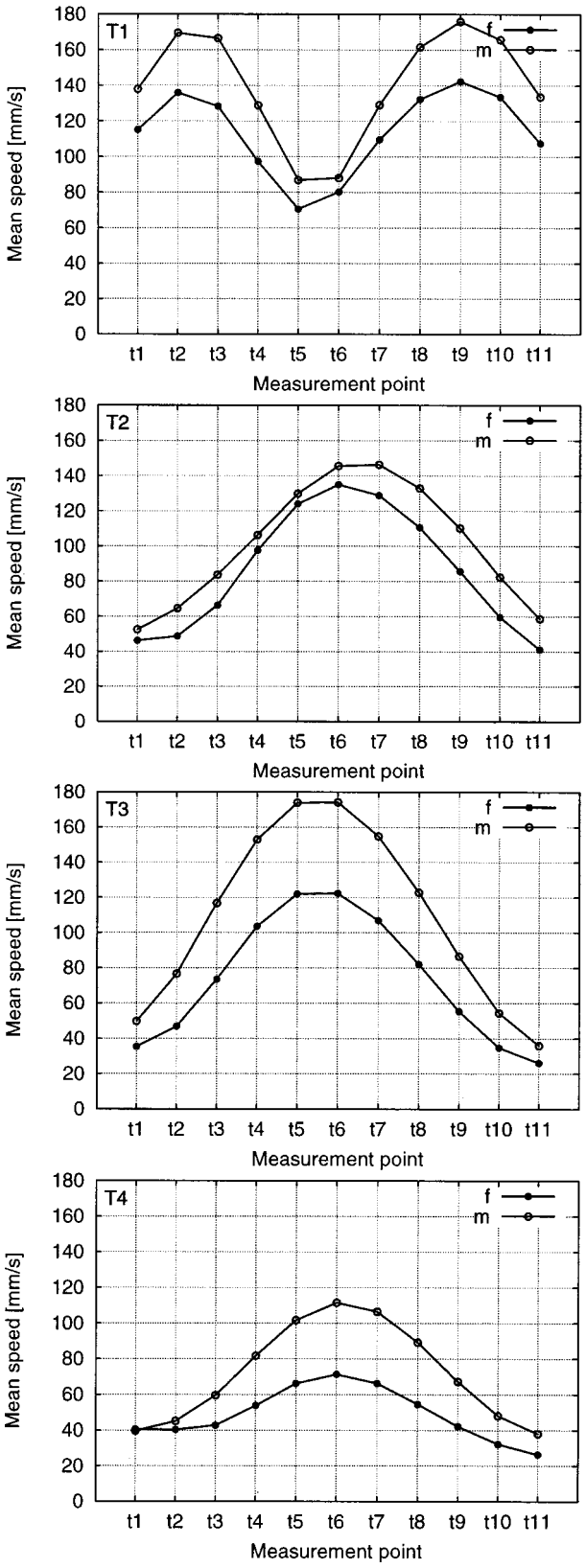


FIG. 11. Mean female (filled circles) and male (open circles) speeds for each of the lingual pellets T1–T4 at each of the measurement points, t1–t11.

diphthong tokens produced by approximately equally sized groups of female and male subjects in the *UWDB* has begun to substantiate the hypothetical predictions. The results show that there is a trade-off between tongue movement and its acoustic product. To achieve a similar perceptual product the male tongue must traverse a greater articulatory distance at greater speed, having overall larger dorso-palatal strictures throughout the vocalic stretch. However, the use of ''similar'' with respect to the perceptual products must be treated with care, since it is not clear to what extent the observed differences between formant values reflect gender-related behavioral differences, i.e., males are producing less, females more peripheral vowel qualities.

TABLE VII. Means and standard deviations of durations of the initial lateral and diphthongal portions in ''light'' together with mean duration of utterance containing the ''light'' tokens (all values in ms).

	[l]		[aɪ]		Utterance		
	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s	n
Female	82	30	165	32	2911	349	128
Male	90	25	152	21	2872	333	109

One aspect of the data presented here which appears to run contrary to the predictions made at the outset is the mean durations of the diphthong tokens mentioned at the beginning of the results section. With a mean duration of 165 ms the female diphthong is approximately 8.9% longer than the mean male diphthong duration of 151 ms. Similar observations have been made in other studies for Dutch (Koopmans-van Beinum, 1980) and British English (Whiteside, 1996). One of the hypothetical predictions was that a female speaker, having a smaller distance to travel to attain the same phonetic target, should be able to do so in a shorter space of time. Why then should male diphthongs on average be shorter despite the greater distance covered? One possible explanation is sociophonetic. Producing longer segment durations can be seen as one of the correlates of speaking clearly, an attribute often assigned to female speech (e.g., Elyan, 1978). However, two pieces of evidence speak against a direct behavioral difference. First, measurement of other durational aspects of the sentences containing the diphthong in question suggest a more complex picture. Table VII contains means and standard deviations of durations of the initial lateral and diphthongal portions in *light* as well as mean durations of the utterances containing the *light* tokens. While the difference in mean diphthong duration is highly significant ($p<0.0001$), neither the initial lateral nor the utterance durations are significantly different. The second piece of evidence comes from vowel duration differences in German. Simpson (1998) examined acoustic and durational patterns in a sample of more than 60 000 vowel tokens from read and spontaneous speech material contained in the Kiel Corpus (Simpson *et al.*, 1997). Table VIII contains means and standard deviations of durations of the 5062 spontaneous and 2160 read diphthong tokens from the vowel sample. The final column contains the female-male durational differences in percent. As might have been expected, the durations of read vowels are longer than those of the spontaneous vowels. However the female vowels are consistently longer than the male counterparts in similar proportions within each corpus, and the average differences (spontaneous: 9.1%, read: 11.2%) between female and male durations for both corpora

TABLE VIII. Mean and standard deviations in ms of male and female diphthong durations in spontaneous and read German from the Kiel Corpus. Right-hand column contains female–male differences in percent.

	Female			Male			
	$\bar{x}$	s	n	$\bar{x}$	s	n	% difference
			Spontaneous				
aɪ	133	58	1543	120	55	2165	10.1
aʊ	129	63	484	124	56	610	4.0
ʏ	141	40	113	124	35	147	13.7
All	132	58	2140	121	54	2922	9.1
			Read				
aɪ	135	51	678	122	47	678	10.7
aʊ	144	55	276	132	48	276	9.1
ʏ	146	39	126	127	35	126	15.0
All	139	51	1080	125	46	1080	11.2

are very similar to those found in the *light* diphthongs reported here. Simpson (1998, p. 86) only lists durational differences for the ten most frequently occurring monophthongs, but the intra- and intercorpus differences between the genders are comparable to those found for the diphthongs: read durations are always greater than spontaneous durations and female durations are always greater than male durations. Simpson cites sociophonetic reasons for the durational differences found in the German monophthong material, an account which has also been offered for greater female vowel durations found in other studies (e.g., Koopmans-van Beinum, 1980; Whiteside, 1996). However, in light of the striking similarities found in the English and German data, a sociophonetic account becomes increasingly unattractive, since it would not only have to claim that male speakers exhibited consistently shorter vowel durations across different linguistic activities, but also that the degree to which this difference manifested itself against two cultural backgrounds was the same.

It seems necessary, therefore, to look for possible phonetic accounts, which relate the speed of articulatory movement to its overall duration. There are at least two possible biomechanical accounts. First, smaller average female dimensions may require higher articulatory accuracy, in turn demanding slower movement and greater temporal extent of the movement.[4] Lack of gender-specific difference in positional variance for the *y* coordinate (leaving aside the higher overall *x* variance due to the normalization procedure) in Table IV does not, however, seem to suggest any differences in accuracy. The second possible explanation is that speed of movement within an articulatory gesture is positively correlated with the duration of the gesture. Such relationship is provided by the mass–spring model which has commonly been used to model articulatory movements (e.g., Saltzman, 1986; Browman and Goldstein, 1990). In such a model time is not an explicit component, rather the form of oscillation is governed by factors such as stiffness, equilibrium position, or rest length of the spring and damping. Crucial to the present discussion is that the stiffness component creates a relationship between the frequency of oscillation and speed of movement: the higher the stiffness, the higher the frequency (translating into shorter segment duration) and the higher the rate of movement (for the same equilibrium position). It is, however, tenuous, on the basis of the available evidence, to suggest that male and female speakers are executing the same gesture with different degrees of stiffness.

The possible trade-off between speed of articulatory movement and segment duration in the attainment of similar perceptual targets has a number of implications. Striving for equivalent perceptual targets, vocal tract length should be inversely proportional to vowel duration, large males on average having the shortest, young children the longest durations. Of course, tokens of only one diphthong have been investigated here and it is to be expected that the acoustic-articulatory trade-off will manifest itself differently in different utterance environments. So, for instance, consonant–vowel and vowel–consonant transitions may be carried out at similar articulatory speeds producing marked acoustic and perceptual differences for various vocal tract dimensions. Children, who exhibit negligible intergender differences in vocal tract dimension, may begin to imitate aspects of adult behavior at an early stage, e.g., boys may imitate shorter adult male durations. This is one interpretation of production and perception data from Günzburger *et al.* (1987). Normally sighted and visually handicapped listeners were asked to identify the gender of prepubertal boys and girls on the basis of utterances of isolated vowels and sentences. Gender identification on the basis of isolated vowel stimuli was poorer than for read sentences. While differences in f_0 and formants were found, the results of qualitative judgments provided by the visually handicapped listeners are of particular interest, since the sentence utterances of girls were in general judged to be less sloppy and clearer, which in turn could be directly related to durational differences. Unfortunately, these were not addressed in their study.

VI. CONCLUSIONS

This study has shown that differences in the average vocal tract dimensions of males and females have temporospatial consequences for articulation and its acoustic and auditory products. This introduces an important dynamic aspect to accounting for the nonuniform differences between male and female vowel systems. They do not preclude the other factors discussed in Sec. I, rather the dynamic advantages of smaller vocal tract dimensions facilitate a possible need or desire for a more widely spaced acoustic vowel system. The implications of the findings here do, however, go beyond accounting for nonuniform differences between vowel systems. We must expect gender-specific differences in consonant–vowel transitions and undershoot.

Speech synthesis of the female voice, whose quality still lags behind that of the male voice (Henton, 1999), has up until now disregarded gender-specific dynamic differences in filter behavior, concentrating almost exclusively on improving source models (Klatt and Klatt, 1990). However, the results of this study indicate that different timing models are required for segmental duration and possibly for formant transitions in consonant–vowel and vowel–consonant transitions.

ACKNOWLEDGMENTS

I would like to thank Jonathan Rodgers, Michel Scheffers, Anders Löfqvist, and the two anonymous reviewers for their careful and constructive comments. I would also like to thank Hartmut Traunmüller for his invaluable discussion on many of the perceptual and acoustic matters addressed. Any errors which remain are of course my own. This work was supported (in part) by Research Grant No. R01 DC 00820 from the National Institute of Deafness and Other Communicative Disorders, U.S. National Institutes of Health.

[1]Maurer *et al.* define formants as energy maxima in the acoustic signal as identified from spectrographic representations.

[2]E.g., 'Finally, we cannot quite exclude the possibility of universal ''feministic'' preference in vowel qualities which might have influenced the average data' (Fant 1975, p. 18).

[3]For more details of the normalization procedure, see Hashi *et al.* (1998).

[4]Christopher Harris, personal communication.

Bladon, R., Henton, G., and Pickering, J. (**1983**). ''Towards an auditory theory of speaker normalization,'' Language and Communication **4**, 59–69.

Browman, C. P., and Goldstein, L. M. **(1990)**. ''Tiers in articulatory phonology, with some implications for casual speech,'' in *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech*, edited by J. Kingston and M. Beckman (Cambridge U. P., Cambridge), pp. 341–376.

Chiba, T., and Kajiyama, M. **(1941)**. *The Vowel—Its Nature and Structure* (Tokyo-Kaiseikan, Tokyo).

Diehl, R. L., Lindblom, B., Hoemeke, K. A., and Fahey, R. P. (**1996**). ''On explaining certain male-female differences in the phonetic realization of vowel categories,'' J. Phonetics **24**, 187–208.

Eckert, P. (**1989**). ''The whole woman: Sex and gender differences in variation,'' Language Variation and Change **1**, 245–267.

Elyan, O. (**1978**). ''Sex differences in speech style,'' Women Speaking **4**, 4–8.

Fant, G. **(1960)**. *Acoustic Theory of Speech Production* (Mouton, The Hague).

Fant, G. (**1966**). ''A note on vocal tract size factors and non-uniform F-pattern scalings,'' STL-QPSR **4**, 22–30.

Fant, G. (**1975**). ''Non-uniform vowel normalization,'' STL-QPSR **2**(3), 1–19.

Fujimura, O., Kiritani, S., and Ishida, H. (**1973**). ''Computer-controlled radiography for observation of movements of articulatory and other human organs,'' Comput. Biol. Med. **3**, 371–384.

Goldstein, U. **(1980)**. ''An articulatory model for the vocal tracts of growing children,'' Ph.D. thesis, MIT, Cambridge, MA.

Günzburger, D., Bresser, A., and Kers, M. t. (**1987**). ''Voice identification of prepubertal boys and girls by normally sighted and visually handicapped subjects,'' Language and Speech **30**, 47–58.

Hashi, M., Westbury, J. R., and Honda, K. (**1998**). ''Vowel posture normalization,'' J. Acoust. Soc. Am. **104**, 2426–2437.

Henton, C. G. **(1992)**. ''The abnormality of male speech,'' in *New Departures in Linguistics*, edited by G. Wolf (Garland, New York), pp. 27–59.

Henton, C. G. **(1995)**. ''Cross-language variation in the vowels of female and male speakers,'' in *Proc. XIIIth ICPhS*, Vol. 4, Stockholm, pp. 420–423.

Henton, C. G. (**1999**). ''Where is female synthetic speech?'' J. Int. Phonetic Assoc. **29**, 51–62.

Henton, C. G., and Bladon, R. A. W. (**1985**). ''Breathiness in normal female speech: Inefficiency versus desirability,'' Language and Communication **5**, 221–227.

Holmberg, E. B., Hillman, R. E., and Perkell, J. S. (**1988**). ''Glottal airflow and pressure measurements for soft, normal and loud voice by male and female speakers,'' J. Acoust. Soc. Am. **84**, 511–529.

House, D. **(1990)**. *Tonal Perception in Speech*, Travaux der l'Institut de Linguistique de Lund (Lund U. P., Lund).

IPDS **(1997)**. ''xassp User's Manual (Advanced Speech Signal Processor under the X Window System),'' in *The Kiel Corpus of Read/Spontaneous Speech—Acoustic data base, processing tools and analysis results*, edited by A. P. Simpson, K. J. Kohler, and T. Rettstadt, Arbeitsberichte des Instituts für Phonetik und digitale Sprachver-arbeitung der Universität Kiel (AIPUK) 32, pp. 31–115.

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.

Koopmans-van Beinum, F. J. **(1980)**. *Vowel Contrast Reduction: An Acoustic and Perceptual Study of Dutch Vowels in Various Speech Conditions* (Academische Pers B. V., Amsterdam).

Labov, W. **(1972)**. *Sociolinguistic Patterns* (Blackwell, Oxford).

Labov, W. (**1990**). ''The intersection of sex and social class in the course of linguistic change,'' Language Variation and Change **2**, 205–254.

Maurer, D., Cook, N., Landis, T., and d'Heureuse, C. (**1991**). ''Are measured differences between the formants of men, women and children due to f_0 differences?'' J. Int. Phonetic Assoc. **21**, 66–79.

Milroy, J., and Milroy, L. (**1993**). ''Mechanisms of change in urban dialects: The role of class, social network and gender,'' Int. J. Appl. Linguistics **3**, 57–77.

Nordström, P.-E. (**1977**). ''Female and infant vocal tracts simulated from male area functions,'' J. Phonetics **5**, 81–92.

Peterson, G., and Barney, H. (**1952**). ''Control methods used in the study of vowels,'' J. Acoust. Soc. Am. **24**, 175–184.

Potter, R. K., and Steinberg, J. C. (**1950**). ''Toward the specification of speech,'' J. Acoust. Soc. Am. **22**, 807–820.

Ryalls, J. H., and Lieberman, P. (**1982**). ''Fundamental frequency and vowel perception,'' J. Acoust. Soc. Am. **72**, 1631–1634.

Saltzman, E. **(1986)**. ''Task dynamic coordination of the speech articulators: a preliminary model,'' in *Generation and Modulation of Action Patterns*, edited by H. Heuer and C. Fromm, Experimental Brain Research Series 15 (Springer, New York).

Simpson, A. P. **(1998)**. *Phonetische Datenbanken des Deutschen in der empirischen Sprachforschung und der phonologischen Theoriebildung*, Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel (AIPUK) 33.

Simpson, A. P., Kohler, K. J., and Rettstadt, T. (eds.) **(1997)**. *The Kiel Corpus of Read/Spontaneous Speech—Acoustic data base, processing tools and analysis results*, Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel (AIPUK) 32.

Syrdal, A. K. (**1985**). ''Aspects of a model of the auditory representation of American English vowels,'' Speech Commun. **4**, 121–135.

Syrdal, A. K., and Gopal, H. S. (**1986**). ''A perceptual model of vowel recognition based on the auditory representation of American English vowels,'' J. Acoust. Soc. Am. **79**, 1086–1100.

Titze, I. R. (**1989**). ''Physiologic and acoustic differences between male and female voices,'' J. Acoust. Soc. Am. **85**, 1699–1707.

Traunmüller, H. (**1984**). ''Articulatory and perceptual factors controlling the age- and sex-conditioned variability in formant frequencies of vowels,'' Speech Commun. **3**, 49–61.

Traunmüller, H. **(1985)**. ''The role of the fundamental and the higher formants in the perception of speaker size, vocal effort, and vowel openness,'' *Phonetic Experimental Research at the Institute of Linguistics University of Stockholm (Perilus)* **4**, 92–102.

Traunmüller, H. (**1988**). ''Paralinguistic variation and invariance in the characteristic frequencies of vowels,'' Phonetica **45**, 1–29.

Traunmüller, H. (**1990**). ''Analytical expressions for the tonotopic sensory scale,'' J. Acoust. Soc. Am. **88**, 97–100.

Westbury, J. R. (**1991**). ''The significance and measurement of head position during speech production experiments using the x-ray microbeam system,'' J. Acoust. Soc. Am. **89**, 1782–1791.

Westbury, J. R. **(1994)**. *X-ray Microbeam Speech Production Database User's Handbook, Version 1.0*, Madison, WI.

Whiteside, S. P. (**1996**). ''Temporal-based acoustic-phonetic patterns in read speech: Some evidence for speaker sex differences,'' J. Int. Phonetic Assoc. **26**, 23–40.

Linear degrees of freedom in speech production: Analysis of cineradio- and labio-film data and articulatory-acoustic modeling

Denis Beautemps,[a] Pierre Badin,[b] and Gérard Bailly[c]
Institut de la Communication Parlée, UMR CNRS 5009, INPG–Université Stendhal, 46, Av. Félix Viallet, F-38031 Grenoble Cedex 1, France

(Received 15 August 1999; accepted for publication 8 February 2001)

The following contribution addresses several issues concerning speech degrees of freedom in French oral vowels, stop, and fricative consonants based on an analysis of tongue and lip shapes extracted from cineradio- and labio-films. The midsagittal tongue shapes have been submitted to a linear decomposition where some of the loading factors were selected such as jaw and larynx position while four other components were derived from principal component analysis (PCA). For the lips, in addition to the more traditional protrusion and opening components, a supplementary component was extracted to explain the upward movement of both the upper and lower lips in [v] production. A linear articulatory model was developed; the six tongue degrees of freedom were used as the articulatory control parameters of the midsagittal tongue contours and explained 96% of the tongue data variance. These control parameters were also used to specify the frontal lip width dimension derived from the labio-film front views. Finally, this model was complemented by a conversion model going from the midsagittal to the area function, based on a fitting of the midsagittal distances and the formant frequencies for both vowels and consonants. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1361090]

PACS numbers: 43.70.Bk [AL]

I. INTRODUCTION

''Speech is rather a set of movements made audible than a set of sounds produced by movement,'' posited Stetson in 1928 (p. 29). This statement could be more properly rephrased as ''Speech can be regarded as the *audible* and *visible* signals resulting from articulatory movement,'' as stated, for instance, in the *speech robotics* approach fostered in the collaborative European project *Speech Maps* (Abry *et al.*, 1994). In this framework, the speech apparatus is viewed as a *plant* driven by a *controller* so as to recruit articulators and coordinate their movements, which have simultaneous acoustic and visual consequences.

The concept of plant and controller implies the notion of a relatively small number of *independent degrees of freedom* for the articulatory plant, i.e., the specification, for each articulator, of a limited set of movements that can be executed independently of each other by the articulator. As emphasized by Kelso *et al.* (1986), however, the speech production apparatus is made of a large number of neuromuscular components that offer a potentially huge dimensionality and which must be functionally coupled in order to produce relatively simple gestures [this view forms the basis of the concept of coordinative structures in speech, cf. Fowler and Saltzman (1993)]. Maeda (1991) refers to a similar concept in terms of ''elementary articulators.''

One *independent degree of freedom* may be more precisely defined for a given speech articulator as one variable that can completely control a specific variation of shape and position of this articulator, and that is statistically independent of the other degrees of freedom over a set of tasks. These degrees of freedom can be determined by observing the correlations between the various parameters that constitute the accurate geometrical description of the articulators shapes and positions, and retaining only independent parameters. These correlations stem from mainly three levels of implicit or explicit constraints: (1) physical continuity of the articulators (the tongue cannot have a jigsaw shape for instance); (2) biomechanical constraints (the range of possible articulators shapes and positions is limited by the physiological properties of the bony structures and of the muscles); and (3) the nature of the task in relation with control (chewing involves lateral translations of the jaw, but speech does not, and thus jaw has different degrees of freedom depending on the task observed). The correlations observed on articulatory measurements thus cannot always be ascribed with certainty to either biomechanical constraints or to strategies related to the task. For instance, Hoole and Kroos (1998) observed that larynx height and lip protrusion are inversely correlated: this correlation obviously cannot be explained by biomechanical links between lips and larynx, but should be ascribed to control strategies related to the speech task. It thus appears understandable that determining which properties of speech can be attributed to the plant and which to the controller is a recurrent issue in speech motor control (cf., e.g., Perkell, 1991; Scully, 1991; Abry *et al.*, 1994).

It has long been known that midsagittal profiles constitute a privileged representation of speech articulation (cf., e.g., Boë *et al.*, 1995, for a review on vowel representations). Indeed, for most phonemes, the complete vocal tract shape can be fairly well inferred from the midsagittal plane, the most notable exception being lateral sounds. Moreover, midsagittal profiles allow linking of vocal tract articulation and

[a]Electronic mail: beautemps@icp.inpg.fr
[b]Electronic mail: badin@icp.inpg.fr
[c]Electronic mail: bailly@icp.inpg.fr

the resulting acoustics. Articulatory models can therefore be viewed as one of the most efficient means of manipulating vocal tract shapes, and midsagittal profiles as a privileged interface between *motor control* on the one hand and *acoustic and visual* modules of the speech production system on the other hand. Developing and evaluating such articulatory models finally constitutes a good means for identifying the degrees of freedom of speech articulators.

Among the large number of studies devoted to articulatory modeling since the seventies, two main approaches can be identified: *functional articulatory modeling*, where the position and shape of articulators are algebraic functions of a small number of articulatory parameters, and explicit *biomechanical modeling*, where the position and shape of articulators are computed from physical simulations of the forces generated by muscles and of their consequences on the articulators.

In linear articulatory models, the relations between articulator positions and shapes and the control parameters can either be defined in geometrical terms, in which case the degrees of freedom of the articulatory plant are decided *a priori* and fitted to the data *a posteriori* (cf., e.g., Coker and Fujimura, 1966; Liljencrants, 1971; Mermelstein, 1973), or based on articulatory data measured on one or several subjects, in which case the degrees of freedom of the plant emerge from the data (cf., e.g., Lindblom and Sundberg, 1971; Maeda, 1990; Stark *et al.*, 1996).

The general approach of biomechanical articulatory models consists in modeling muscular forces and articulator structure by means of methods inspired from mechanical analysis and numerical simulation (cf., e.g., Perkell, 1974; Wilhelms-Tricarico, 1995; Laboissière *et al.*, 1996; Payan and Perrier, 1997). These models present the advantage of being physical models with intrinsic dynamics, although necessarily extremely simplified, but their control remains very complex, in particular due to the high number of degrees of freedom represented by each individual muscle command. Sanguineti *et al.*'s (1998) work constitutes a good illustration of this. They fitted, with their model, the articulator shapes and positions measured from the x-ray database already used by Maeda (1990) and determined, by optimization, the commands of the 17 muscles involved in their model. They identified then, by linear component analysis applied in the so-called λ-space corresponding to the biomechanical tongue control parameter space, the synergies between these commands, and showed that six independent components could account for most of the data variance of the midsagittal tongue shape. These first six components are closely related to the degrees of freedom that could be extracted directly from the original x-ray contours (Maeda, 1990). It thus appears that, from the point of view of the degrees of freedom, such complex biomechanical models are not a prerequisite to the accurate description of static speech articulation.

Rather than developing *a priori* complex biomechanical models with degrees of freedom in large excess and then reducing this high dimensionality based on articulatory data, we have adopted a dual approach in the present work. More precisely, our objectives were to determine the linear degrees of freedom of one subject's articulators in the midsagittal plane, and to build an *articulatory-acoustic plant* that could be considered a faithful and coherent representation of this subject's articulatory and acoustic capabilities.

The present paper describes our approach to this problem: (1) design of the corpus and collection of articulatory-acoustic data for one subject, (2) analysis of the data and extraction of the independent linear degrees of freedom of the articulators, and (3) the development of a linear articulatory-acoustic model based on these degrees of freedom.

II. ARTICULATORY AND ACOUSTIC DATA ACQUISITION METHODOLOGY

A. The experimental setup: Synchronized cineradio- and labio-film

Cineradiography was chosen as the best compromise between good spatial and temporal resolutions for the articulatory data. This technique, which has been used successfully for speech studies at the Strasbourg Phonetic Institute (cf., e.g., Bothorel *et al.*, 1986), was used in synchrony with the video labiometric method developed at ICP by Lallouache (1990) (cf. also Badin *et al.* 1994a). The recordings were performed at the Strasbourg Schiltigheim Hospital, France. The subject's head was positioned at a distance of 50 cm from the x-ray emitter and 20 cm from the radiance amplifier. An aluminum filter was placed in the lip region to avoid overexposure of the lips, thus improving the contrasts in this region (cf. Bothorel *et al.*, 1986). The vocal tract images produced by the radiance amplifier were captured and recorded by a 35-mm film camera. The subject's lips, painted in blue to allow the lip contours to be extracted by an image processing procedure, were recorded using a video camera. Both cameras were operating at a rate of 50 frames per second. The speech signal, captured by a directional microphone placed at a distance of 10 cm from the subject's mouth, was synchronously recorded.

B. The subject

The choice of the set of subjects always poses a dilemma: a single subject study surely reduces the generality of the work but allows us to gather rich and detailed data, whereas a study with a larger panel of subjects may permit us to draw some general conclusions but limits the extent of the data that can be practically acquired and processed.

Under the auspices of the European collaborative project *Speech Maps*, Abry *et al.* (1994) aimed to gather a variety of converging and complementary articulatory/acoustic data for one subject uttering the same speech material in a controlled manner in different experimental setups. This policy resulted in a large set of data of potential use in speech production modeling, such as vocal tract acoustic transfer functions (Djéradi *et al.*, 1991), acoustic and aerodynamic pressure and flow in the tract (Stromberg *et al.*, 1994; Badin *et al.*, 1995; Shadle and Scully, 1995), electropalatography data (Badin *et al.*, 1994b), and more recently 3D MRI vocal tract images (Badin *et al.*, 1998, 2000) and video face data (Badin

TABLE I. Corpus used for the cineradiographic recordings.

[aɛeiyuoø]
[pavapavipavupivipivupivy]
[pazapazipazupizipizupizy]
[paʒapaʒipaʒupiʒipiʒupiʒy]
[abaabiabuibiibuiby]
[adaadiaduidiiduidy]
[agaagiaguigiiguigy]

et al., 2000). As the present study was conducted in the framework of this project, the same subject was therefore chosen as a *reference* subject.

C. The corpus

As mentioned in the previous section, one of the major aims of the present study was to determine the degrees of freedom of a subject's articulators for speech, excluding any other nonspeech movements. Attaining this goal would ideally require recording a large corpus of speech material containing all possible combinations of phonemes. This is obviously not practical in general, and particularly inappropriate in the case of cineradiography, due to health hazards related to this method. The corpus was thus designed to include as many combinations of Vowel Consonant Vowel (VCV) sequences as possible in a very limited amount of time.

The voiced French plosive and fricative consonants C=[vzʒbdg] were chosen in six vocalic contexts involving the four French extreme vowels V=[aiuy]:aCa, aCi, aCu, iCi, iCu, iCy. It was assumed that the voiceless cognates of these consonants correspond approximately to the same articulation. The presence of voicing was expected to simplify the tracking of formants during fricative consonants. The French [l] was excluded because of the impossibility of getting information on the lateral channels from midsagittal x-ray pictures. Nasals were also excluded because velar movements do not directly influence other articulators movement (although nasal vowels in French seems to imply some additional tongue backing compared to other vowels, cf., e.g., Zerling, 1984), and will thus be studied in the near future. Finally the French [ʁ] was also not included because it was not hypothesized to require extra degrees of freedom for the midsagittal profile. The extreme vowels were expected to represent the most extreme vocalic articulations in French. Moreover, the fricative items were interspersed with [p]'s in order to allow the estimation of subglottal pressure during the fricatives as the intraoral pressure during the closure of [p] (Demolin *et al.*, 1997). In addition, a series of connected vowels [aɛeiyuoø] was recorded in order to test formant/cavity affiliation hypotheses (Bailly, 1993). Finally, the corpus duration could be reduced to about 24.5 s of signal (actually leading to 1222 pictures), with the following distribution of phonemes: 30 [ai], 12 [u], 6 [y], 6 [vzʒbdg], 18 [p]. Table I presents the complete corpus.

D. Processing of the x-ray and video images

For each picture, the sagittal contours were first drawn by hand from a projection of the picture onto a piece of paper, digitized by a scanner, and finally 11 subcontours were hand-edited by means of an interactive software. These subcontours correspond to different articulators or vocal tract regions: the upper and lower lips, the hard palate, the velum, the different components of the pharynx, the larynx, the tongue, the jaw, and the hyoid bone. Figure 1 presents an example of the resulting midsagittal contour. Note that rigid structures (hard palate, jaw, hyoid bone) were not drawn again for each image, but were given reference contours that best fitted most of the shapes observed on the whole set of images; for these structures, the operator's task was then only to optimally position the reference shapes for each image by roto-translation. This procedure presented four advantages: (1) the operator's task was made easier and faster; (2) it reduced the noise due to manual drawing; (3) it avoided the difficulty of precisely determining reference landmarks such as incisor edges from images where the contrast is not always very high; (4) it offered the possibility of determining in a straightforward manner the positions of these rigid bodies in the midsagittal plane (see the discussion on jaw analysis below).

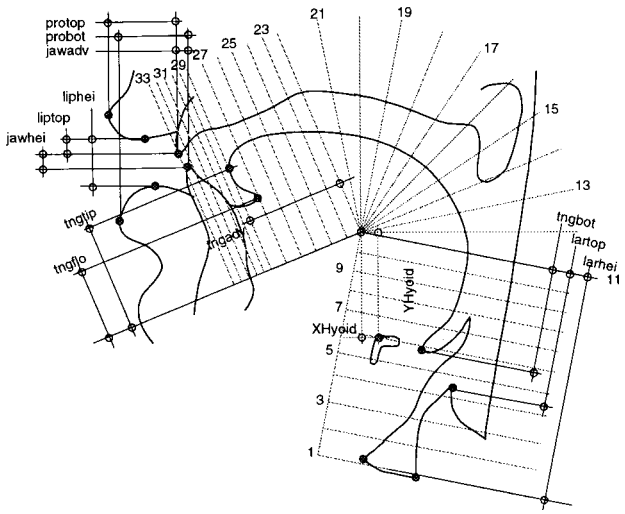


FIG. 1. Example of manually drawn *VT* contours and associated articulatory measures: upper lip protrusion *ProTop*, lower lip protrusion *ProBot*, upper lip elevation *LipTop*, jaw height *JawHei*, jaw advancement *JawAdv*, tongue tip advancement *TngAdv*, tongue tip height *TngTip*, tongue floor height *TngFlo*, tongue bottom *TngBot*, height of the larynx top *LarTop*, height of the larynx bottom *LarHei*, *x/y* coordinates of the hyoid bone *XHyoid,YHyoid*.

Concerning the lips, the blue of the video front lip images was converted into absolute black by means of an analogue Kroma-key, and the inner contour was then automatically determined by simple adapted thresholding (cf. Lallouache, 1990, or Badin *et al.*, 1994a, for more details).

E. Articulatory measurements

Before going into some details, it is useful to specify the midsagittal coordinate system attached to the skull structure and used in this study. The lower edge of the upper incisors is given arbitrary *x/y* coordinates (5,10) in cm. The *x*-axis is positive in the *posterior* direction of the head (toward the back), and negative in the *anterior* direction (toward the nose); the *y*-axis is positive in the *superior* direction (toward the brain), and negative in the *inferior* direction (toward the

feet). Finally, it happens that the direction of the *maxillary occlusal plane* (defined as the plane ''given by the tips of the central incisors and at least two other maxillary teeth on opposite sides of the mouth,'' Westbury, 1994), is oriented at an angle of 4.6° from the $y=0$ axis.

The values of a number of geometrical parameters (see Fig. 1) have been determined from the midsagittal contours: upper *ProTop* and lower *ProBot* lip protrusions, upper lip elevation *LipTop*, jaw height *JawHei*, jaw advancement *JawAdv*, tongue tip advancement *TngAdv* and height *TngTip*, tongue floor height *TngFlo*, tongue bottom *TngBot*, height of the larynx top *LarTop* and bottom *LarHei*, x/y coordinates of the hyoid bone *XHyoid*, *YHyoid*. In addition, three parameters were extracted from the video front views of the lips: lip height *B*, lip width *A*, and the intra-labial lip area *S*.

Note that the distance between the upper and lower lips can be determined either from the midsagittal profile *LipHei* or from the front view *B*: as expected, the two measures are very close to each other, *B* being less accurate when the lip opening is close to zero, particularly for the rounded vowels [uy], due to the fact that the subject's upper lip tends to mask the intra-labial orifice in such cases. The correlation coefficient between the measures is 0.99, while the rms error is 0.1 cm.

The articulators are expected to follow relatively smooth trajectories due to their long time responses (cf., e.g., the jaw characteristic resonance frequency of 5–6 Hz mentioned by Sorokin *et al.*, 1980). Deviations of geometric measures from their smooth trajectories revealed that the noise added by the whole chain of acquisition was in the range of 1-mm peak-to-peak.

F. Midsagittal contours

A semi-polar grid has been used to describe the midsagittal contours, as has traditionally been done since Heinz and Stevens (1965) or Maeda (1979). However, as proposed by Gabioud (1994), two parts of this grid have been made adjustable in order to follow the movements of the larynx (gridlines 1 to 6, line 1 being the lowest one near the glottis) and the movement of the tongue tip (in fact, the *tongue blade*, defined as the linguistic class *coronal articulation*; grid lines 24 to 28). This grid presents a double advantage: (1) the number of intersection points between the grid and the tongue contour is constant, whatever the extension of the tongue tip or of the larynx, which is a crucial feature for further statistical analysis: (2) the fact that the measurement grid follows tongue tip movements implies that all the points in the vicinity of the tongue tip present a behavior close to that of *flesh-points*, i.e., points mechanically attached to the tongue surface, which is more than just a description of tongue shape (this is useful when recovering tongue contours from flesh-points coordinates measured by electromagnetic articulometry; see Badin *et al.*, 1997). Finally, a third part of the grid has been introduced to describe the alveolar dental cavity with adjustable grid lines (grid lines 29 to 33) equally spaced between the tongue tip and the lower edge of the upper incisor. The inner and outer vocal tract midsagittal contours intersect thus the grid lines at 2×33 points; each contour can thus be represented by the 33-element vector (referred to as *Int* and *Ext*) of the abscissa of these intersection points along the grid lines. Note that, as mentioned by Westbury (1994), the dimensionality of the articulators may change depending on the coordinate system used. For instance, a point running on a fixed circle appears to have two *linearly* independent degrees of freedom in a Cartesian system, but only one single degree of freedom in a polar coordinate system. The choice of the dynamically adjustable semi-polar grid system seems a good solution to avoid artificial overdimensionality.

The velum is terminated in the midsagittal plane by the uvula. It can be in contact with the upper surface of the tongue, however, without creating a real constriction in the vocal tract, since air remains free to flow on each side. This fact has been approximately taken into account by making the velum artificially thinner by a factor linearly increasing from 0 at its extremity to about 40% at its base, and by shifting the result so as to align its posterior wall with the pharyngeal wall.

G. Formants

The speech signal was digitized at 16 kHz, and the first four formant values were estimated using LPC analysis with a 20-ms window centered on the times where the midsagittal views were acquired, leading to formant trajectories sampled at 50 Hz. Because of the background noise due to the x-ray emitter, the signal-to-noise ratio was rather poor (about 25 dB), and thus some of the formants had to be hand-edited. This was done in reference to another version of the same corpus recorded by the subject in good recording conditions. The $F1/F2$ and $F1/F3$ spaces for the pooled vowels and consonants are shown in Fig. 2.

III. ANALYSIS OF THE INDEPENDENT LINEAR DEGREES OF FREEDOM OF THE MIDSAGITTAL CONTOURS

A. Principles

1. Identifying degrees of freedom

As mentioned in Sec. I, the approach taken in the present study to determine the degrees of freedom of the various speech articulators is based on articulatory data obtained from one *subject* producing a given *corpus* in a given *language*.

In general, speech articulators possess excess degrees of freedom, i.e., a given articulation can be achieved by means of different combinations of the available degrees of freedom of the articulators (cf. bite-block experiments performed by Lindblom *et al.*, 1979). Control strategies finally aim at recruiting these degrees of freedom when they are needed to attain given articulatory/acoustic/visual goals, and leaving them free to anticipate other goals whenever possible (this is one basic principle of *coarticulation*). In the present data driven approach, the problem is to decide the repartition of the variance of the measured articulatory variables between the different variables associated with the degrees of freedom. The present work rests on a common consideration in speech motor control modeling: what is explained by the biomechanics of the speech plant does not need to be worked

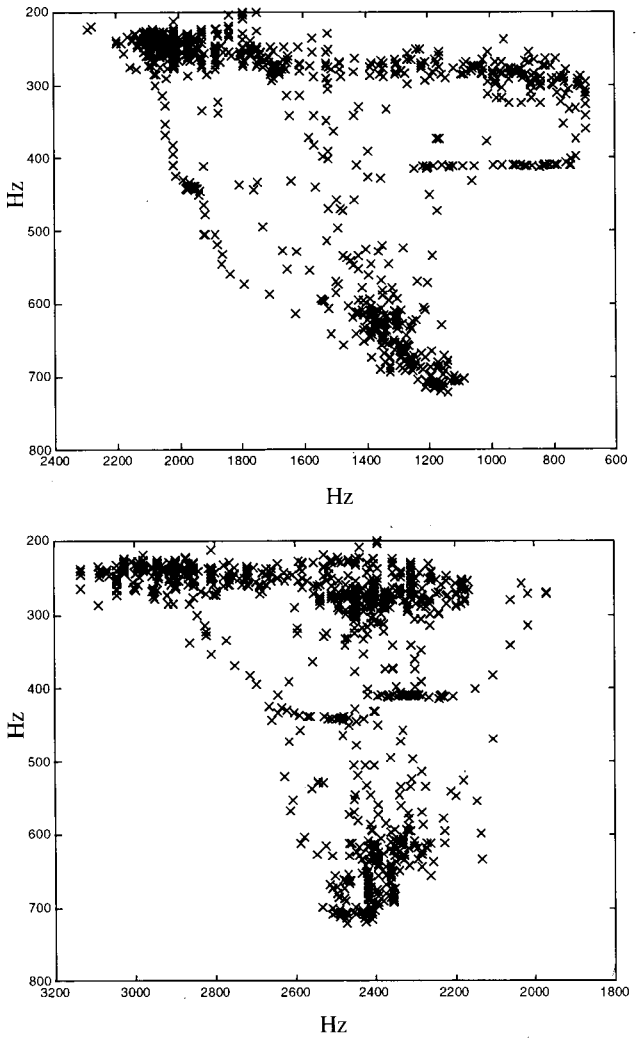


FIG. 2. Measured $F2/F1$ (top) and $F3/F1$ (bottom) formant spaces (in Hz) for the vowels and consonants pooled.

out by the controller (Abry *et al.*, 1994; Perrier *et al.*, 1996, 2000). In other words, any correlation observed between the articulatory variables should be used to reduce the number of degrees of freedom of the articulators. However, this approach must be carefully balanced by another criterion, the *biomechanical likelihood*. For instance, if larynx height and lip protrusion are inversely correlated due to the subject's articulatory control strategy (cf., e.g., Hoole and Kroos, 1998), two separate degrees of freedom should nevertheless be considered, even at the price of some residual correlation between the corresponding parameters.

2. Linear component analysis

Another important assumption in the present work is the *linearity* of the analysis and of the associated model: the shape data vectors *DT* are decomposed into linear combinations of a set of basic shape vectors *BV* weighted by loading factors *LF*, in addition to their average *neutral* shape $\overline{DT}$:

$$DT=\overline{DT}+LF\cdot BV.$$

Each loading factor *LFi* corresponds to an independent linear component, if its cross correlation with the other loadings is zero over the corpus of data. The dimensionality of the articulators' shapes and positions can thus be explored by classical linear analysis techniques such as principal component analysis (PCA) and linear regression analysis, as carried out by Maeda (1990, 1991), whose approach largely inspired the present work.

Maeda's approach to this decomposition was to iteratively determine each linear component in the following way: (1) the loading factor *LFi* is determined from the data as described below; (2) the associated basis shape vector *BVi* is determined by the linear regression of the current residual data for the whole corpus over *LFi*; (3) the corresponding contribution of the component is computed as the product of the loadings by the basis shape vector, and is finally subtracted from the current residue in order to provide the next residue for determining the next component.

For some of the linear components, the loading factors were arbitrarily chosen as the centered and normalized values of specific geometric measurements extracted from the contours, such as jaw or larynx height. For the other linear components, loading factors were derived by standard PCA applied to specific regions of the tongue contour.

Note that the solution of this type of linear decomposition is not unique in general: PCA delivers optimal components explaining the maximum data variance with a minimum number of components, but Maeda's linear component analysis allows a certain room of maneuver to control the nature and repartition of the variance explained by the components (for instance to make them more interpretable in terms of control), at the cost of a suboptimal variance explanation.

In this rest of this section, the various geometric measures are studied using statistical linear analysis in order to determine the correlations between these articulatory variables and to determine the degrees of freedom of the articulatory plant.

B. Jaw

The tongue is naturally identified as an important articulator in speech production, and its midsagittal contours, obtained from x-ray profile views of the vocal tract, have been the focus of most modeling efforts. The jaw has long been recognized as one of the main speech articulators, because it carries both the tongue and the lips. Its specific contribution to tongue shape has been clearly identified and related to the phonetic features of vowels (Lindblom and Sundberg, 1971). The dimensionality of jaw motion has been studied by many researchers (cf., e.g., Westbury, 1988; Edwards and Harris, 1990; Ostry *et al.*, 1997). The jaw, a rigid body, possesses six geometrical degrees of freedom (three rotations and three translations); however, it appears that for speech, movements are mostly restricted to the midsagittal plane if the rotation around the jaw axis is neglected (cf., e.g., Ostry *et al.*, 1997), which reduces the degrees of freedom of the jaw to three. From simple geometrical considerations, it is clear that the position of the jaw as a rigid body in a plane is uniquely defined by one rotation (defined here as *JawRot*) and by the two *x/y* translations of a reference point attached to the body, chosen as the upper edge of the lower incisors (defined here

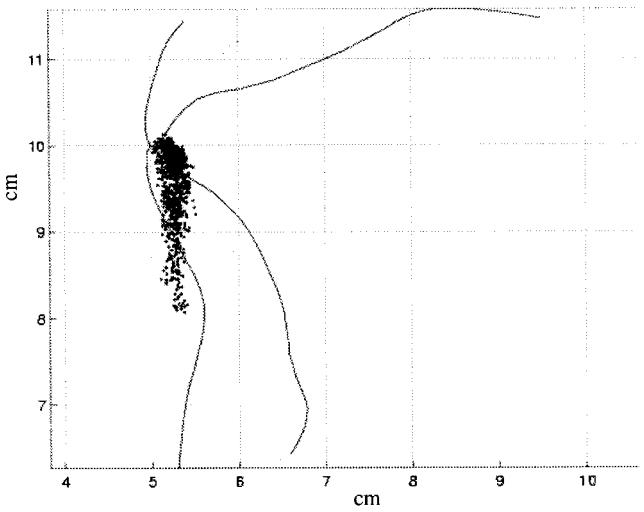


FIG. 3. Dispersion of the lower incisor upper edge superimposed on the contours of the hard palate and of the jaw.

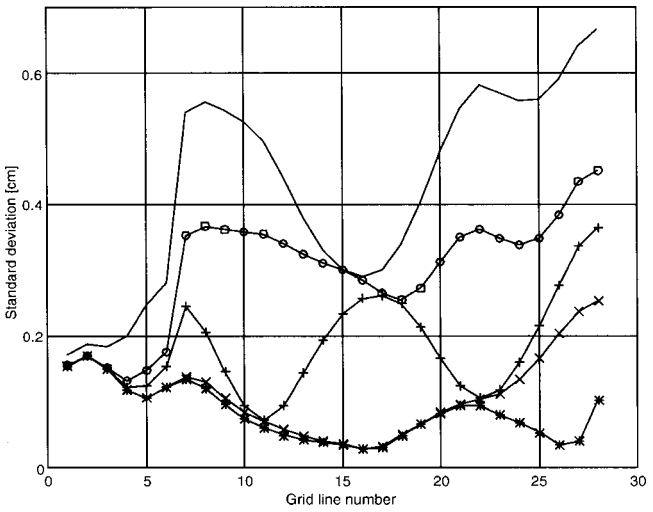


FIG. 4. Standard deviation (in cm) against gridline number for the inner contour of the vocal tract for the successive residues when the effect of the parameters are removed one by one (Raw data: solid line; data after suppression of the contribution of *JH* ''○,'' then of *TB* ''+,'' then of *TD* ''×,'' and finally of *TT* ''*'').

as *JawHei* and *JawAdv*; see Fig. 1). Figure 3 displays the resulting distribution of lower incisor position for the whole corpus.

PCA was applied to the centered—but nonnormalized—jaw position data *JawAdv* and *JawHei*. The first component explains 97.0% of the *JawAdv* and *JawHei* data variance. The overwhelming importance of this component could be predicted intuitively from the fact that the standard deviations of *JawAdv* and *JawHei* are, respectively, about 0.075 and 0.419 cm.

The jaw height component *JH* corresponding to the first degree of freedom of the jaw data was thus defined as the *JawHei* variable centered on its mean and normalized by its standard deviation. A second component, corresponding to jaw advance, *JA* was defined as the residue of *JawAdv* centered and normalized once the linear contribution of *JH* was removed. It can be concluded that for the present subject and corpus, the jaw possesses two independent degrees of freedom in the midsagittal plane, although the second component would have a rather limited influence on the tongue and the lips, as will be discussed further. The jaw was observed to be most retracted for labio-dentals: indeed this retraction allows the lower incisors and the upper lip to get in contact. Maximum jaw protrusion was observed for the coronal fricative [z]: this facilitates the creation of a constriction between the anterior region of the tongue blade and the front region of the alveolar ridge.

C. Tongue

The tongue shape is defined by the vector *Int* of the abscissa of its intersections with the grid lines. Since the jaw carries the tongue, the contribution of its movements should first be subtracted from tongue movements to maintain some biomechanical likelihood. However, due to the complexity of the muscular links between tongue and jaw (cf., e.g., Sanguineti *et al.*, 1998), it is very difficult to separate tongue movements induced by jaw movements from those due to active actions of tongue muscles themselves. In a study involving three subjects (including the subject of the present study), Bailly *et al.* (1998) showed that the slope of the regression line that links jaw height and tongue abscissa can be substantially greater than unity (by almost 100%). This means that tongue movements are apparently larger than the associated jaw movements, indicating that the subjects tend to actively move both jaw and tongue in synergy. In such a case, the passive tongue movement due to jaw movement needs to be determined. However, for the present subject, this synergy was rather weak compared to that of the other subjects (regression slopes lower than 1.15), and did not need to be taken into account: the *JH* parameter was directly considered as the first linear loading factor for each element of *Int*, and the corresponding prediction coefficients were obtained as the coefficients of the linear regression between *Int* and *JH* computed over all the items. Finally, the residual vector *Int_JH*, computed as the difference between predicted and measured values, for all the items, represents the tongue shape from which the contribution of the jaw has been removed.

The variance of the original *Int* data and the variance of the *Int_JH* residual data can be examined in Fig. 4 in terms of standard deviation (i.e., as the square root of the variance), as a function of grid line number. Table II gives, in addition, the global percentage of the total *Int* data variance explained by *JH* (numerical column 1).

The influence of the second jaw parameter *JA* upon tongue contours will be addressed later in this section.

The next step of the analysis consisted of extracting the degrees of freedom of the residual vector *Int_JH*. Gabioud (1994) showed that PCA applied to the whole tongue contour led to poor modeling of the tongue tip, even using three components. It was thus decided to apply PCA separately to the tongue body (gridline 7 to 24) and to the tongue tip (lines 24 to 28).

A first PCA procedure was thus applied to the residues of the 18 points considered for the tongue body, *Int_JH*(7:24). The first two components were retained. The

TABLE II. Summary of data variance explanation for the tongue contours. Column *Design* indicates how the factor was extracted. First column *Var* shows the ratio of data variance explained by the factor for the case the influence of jaw movements is taken into account by one parameter only. The second and third columns *Var* show the ratio of data variance explained when jaw is taken into account by two factors, the second factor being imposed at two different stages of the analysis.

Param.	*Design*	*Var.*	*Var.*	*Var.*
JH	Jaw height	52.2%	52.2%	52.2%
JA	Jaw advance		1.4%	
TB	PCA/tongue body	28.8%	28.6%	28.8%
TD	PCA/tongue body	11.4%	10.4%	11.4%
TT	PCA/tongue tip	3.6%	3.6%	3.6%
JA	Jaw advance			0.2%
	Total	**96.0%**	**96.1%**	**96.1%**

corresponding principal axes are characterized by the eigenvectors associated with the highest two eigenvalues of the cross-correlation matrix computed from these residues. The projections of the centered and normalized residues on these two principal axes give the values of the two associated components: *tongue body* component *TB*, and *tongue dorsum* component *TD*, which describe, respectively, the *front–back* and *flattening–arching* movements of the tongue (see also the nomograms in Fig. 8). These components were then used as predictors for the whole tongue contour. Table II presents a summary of the proportion of the total tongue data variance explained by each component, while Fig. 4 shows the details of the variance of the residues.

The tongue tip was found to possess two independent degrees of freedom [its coordinates, measured as *TngAdv* and *TngTip*[1] (see Fig. 1) are plotted in Fig. 5]: indeed, the residues of *TngAdv* and *TngTip*, after subtraction of the contributions of *JH*, *TB*, and *TD* (determined by the linear regression of *TngAdv* and *TngTip* for the whole corpus over *JH*, *TB*, and *TD*) are clearly not correlated. A first component, more generally dedicated to the representation of the apical region of the tongue, was then extracted: the *tongue tip* component *TT* is defined as the first component determined by the PCA of the residues of the tongue tip region (lines 24 to 28), from which the contributions of *JH*, *TB*, and *TD* have been removed. Its effects can be observed in Fig. 4 and Table II.

The *tonge advance* parameter *TA* was defined as the centered and normalized residue of the measured tongue advance *TngAdv* from which the contributions of *JH*, *TB*, *TD*, and *TT* were subtracted. Since it was, as expected, found to have a negligible predictive power on the tongue abscissa, it was not used as a loading factor for *Int*, but just to control the longitudinal extension of the grid in the front mouth region.

In order to test the influence of the *JA* parameter upon tongue contours, two experiments were carried out: in a procedure similar to that applied to *JH*, *JA* was used as the second imposed loading factor for the tongue analysis in one experiment, and as the loading factor imposed after *JH*, *TB*, *TD*, and *TT* in the other experiment. It was found that *JA* explained only 1.3% of the tongue data variance in the first

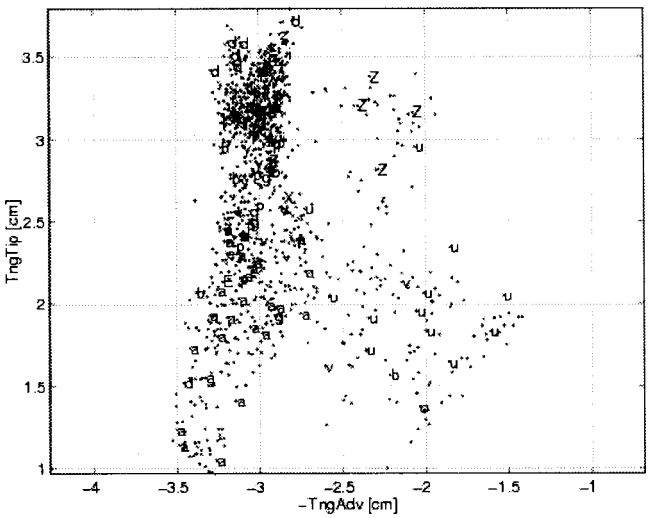


FIG. 5. Plot of the *TngAdv/TngTip* coordinates of the tongue tip (note that these are expressed in the rotated coordinate system attached to the gridline for the front part of the vocal tract; see Fig. 1). Phonemes /ø/, /ʒ/, /ɛ/ are referred to by symbols X, Z, and E, respectively.

case (see Table II, numerical *Var* column 2), and 0.2% in the second case (numerical *Var* column 3). A comparison of the associated nomograms in Fig. 6 suggests that the data variance explained by *JA* in the first case is actually explained by the other components *TB*, *TD*, and *TT*, in the second case. This hypothesis is also supported by results in Table II (and by a more detailed analysis of tongue shape data). *JA* was therefore not used as a control parameter of tongue shape.

In summary, the tongue contours in the grid line system possess four degrees of freedom, controlled by components *JH*, *TB*, *TD*, and *TT*. These four components account for 96% of the tongue variance data, which is only 1.5% less than the variance explained by the first four independent components (but with no direct articulatory interpretation) of a principal component analysis. The standard deviation of the residual error (normally distributed around zero on each gridline) reaches a maximum of 0.15 cm in the vicinity of the pharynx and of 0.1 cm at the tongue tip. The rms reconstruction error for the tongue, i.e., the root mean square error between the measured tongue data and the data calculated with the linear decomposition, amounts to a global value of 0.09 cm, while reaching maxima of 0.15 cm in the vicinity of the pharynx and of 0.1 cm at the tongue tip. The relatively poor modeling of the tongue tip extremity (which results in

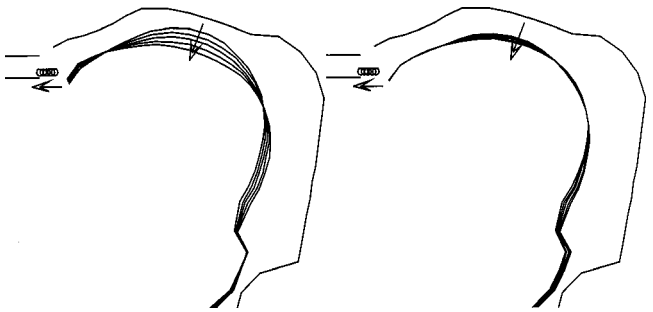


FIG. 6. Articulatory nomograms for *JA*. Left: case where *JA* is the second loading factor in the analysis; right: case where *JA* is the fifth loading factor after *JH*, *TB*, *TD*, and *TT*.

TABLE III. Correlation coefficients of some articulatory measurements. Absolute values higher than 0.6 are in bold face.

	JawHei	*JawAdv*	*LipHei*	*LipTop*	*ProTop*	*ProBot*	*TngTip*	*TngAdv*	*TngFlo*	*TngBot*	*LarHei*	*XHyoid*	*YHyoid*
JawHei	1.000												
JawAdv	−0.226	1.000											
LipHei	**0.641**	−0.039	1.000										
LipTop	−0.448	0.127	0.176	1.000									
ProTop	−0.463	0.148	−0.443	0.060	1.000								
ProBot	**−0.608**	0.062	−0.673	0.024	**0.912**	1.000							
TngTip	**−0.734**	0.207	−0.325	0.375	0.027	0.150	1.000						
TngAdv	0.412	−0.108	0.366	−0.242	−0.568	−0.571	0.142	1.000					
TngFlo	**−0.593**	0.053	−0.290	0.219	−0.083	0.056	**0.877**	0.454	1.000				
TngBot	−0.038	−0.109	−0.347	−0.325	0.352	0.411	−0.219	−0.205	−0.160	1.000			
LarHei	−0.358	−0.068	**−0.550**	−0.136	**0.574**	**0.664**	0.005	−0.380	−0.008	0.797	1.000		
XHyoid	**0.733**	−0.128	0.544	−0.259	−0.416	−0.533	−0.556	0.288	−0.469	−0.169	−0.421	1.000	
YHyoid	0.355	0.100	0.522	0.099	−0.460	−0.572	−0.047	0.351	−0.033	−0.815	**−0.859**	0.574	1.000

only minor acoustical effect) is mainly due to measurement inaccuracies related to the difficulty of precisely defining this tongue tip extremity. A supplementary articulatory control parameter could be extracted to more precisely control the pharyngeal region as implied in the ±Advanced Tongue Root languages (cf., e.g., Tiede, 1996).

Recall finally that the grid system is controlled, in addition, by two parameters, i.e., *TA*, and a parameter related to *LarHei* that will be defined in Sec. III E.

D. Lips

A PCA analysis revealed that 98.4% of the variance of the lip measures *LipHei*, *LipTop*, *ProTop*, and *ProBot* can be explained by three independent components, in addition to the natural contribution of jaw height to lip shape. This is expected, as lip protrusions *ProBot* and *ProTop* are strongly correlated (cf. Table III). Note also that in another study on the same subject, where the lip shape was more accurately described as a three-dimensional mesh of points controlled by the 3-D coordinates of 30 control points (Revéret and Benoît, 1998), Badin *et al.* (2000) also found that three degrees of freedom were sufficient to describe the position of the lips on a corpus of 34 sustained articulations (French vowels and consonants), in addition to the *JH* contribution (the *JA* contribution explained only 1% of the lip data variance). These degrees of freedom are related to three gestures: lip protrusion/rounding, lip closure, and a sort of simultaneous vertical movement of both lips as needed for the subject to realize labio-dentals. In order to simplify the model and its relations to simple articulatory measurements and acoustic interpretations of the lip horn, we decided to use an equivalent set of components: (1) a component related to *LipHei*, taken into account by *LH*, the centered and normalized residue of *LipHei* after removing the *JH* contribution; (2) a component related to *ProTop*, *LP*, the centered and normalized value of the residue of *ProTop* after removing the *JH* contribution; and (3) a component related to a mere vertical, roughly synchronous, movement of both upper and lower lips relative to upper incisors lower edge, taken into account by the *lip vertical position* parameter *LV*, the centered and normalized residue of *LipTop* after removing *JH*, *LH*, and *LP* contributions. Note that this approach results in a slight correlation between *LP* and *LV*. Note also that the horizontal jaw retraction aiming at producing labio-dental constrictions is not taken into account as such, but that its acoustical consequences are dealt with in an indirect way (cf. Sec. IV B 3).

E. Other articulatory measurements

Finally, a number of other articulatory measurements were analyzed. Table III provides the linear correlation coefficients between these measurements.

Table III shows that *LarHei* is partially correlated with *ProBot*, *ProTop*, and *LipHei*. These correlations cannot be explained *a priori* by obvious biomechanical effects, and will thus be ascribed to the speaker control strategies. Indeed, it is clearly established that lip rounding and larynx lowering constitutes, for some subjects, a synergetic strategy for high rounded vowels [uy] (Hoole and Kroos, 1998). Larynx height was thus represented by its centered and normalized value *LY*, and further used to control the grid system (cf. Sec. IV A 2).

The horizontal position of the hyoid bone, *XHyoid*, is very highly correlated to jaw height, while its vertical position, *YHyoid*, is even more strongly correlated to larynx height (see Table III and Fig. 7). These two components are

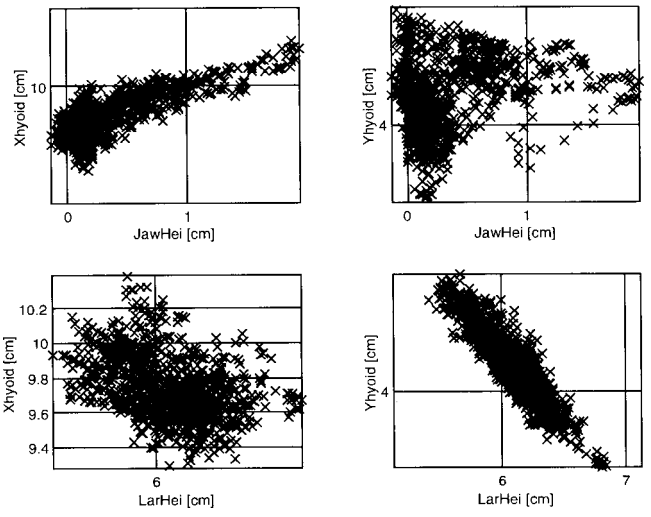


FIG. 7. Plots showing the relations between hyoid bone coordinates and jaw and larynx heights.

TABLE IV. Correlation coefficients of the articulatory control parameters of the model.

	JH	*TB*	*TD*	*TT*	*TA*	*LY*	*LP*	*LH*	*LV*
JH	1.000								
TB		1.000							
TD			1.000						
TT				1.000					
TA					1.000				
LY	−0.358	−0.041	0.272	−0.499	−0.103	1.000			
LP		0.215	0.472	−0.125	−0.118	0.461	1.000		
LH		−0.256	−0.039	0.236	−0.075	−0.417	−0.215	1.000	
LV		0.215	0.065	−0.006	−0.164	−0.045			1.000

clearly less correlated with each other (correlation coefficient $R=0.574$), than found by Westbury (1988) using a more restricted corpus for a single subject ($R=0.871$).

Note that the position of the highest connection point between tongue and epiglottis, referred to as *TngBot* (see Fig. 1), is highly correlated with *LarHei*, as expected. The elevation of tongue floor *TngFlo* is correlated with *TngTip* and *JawHei*.

IV. BERGAME: AN ARTICULATORY-ACOUSTIC MODEL

As stated above, the main function of an articulatory model is to offer a compact representation of articulation, i.e., a representation that needs as few control parameters as possible and is nevertheless accurate enough to be meaningful for speech. The analysis presented in the previous section prepared the ground for establishing such a model, which is necessarily the result of a compromise between a minimum number of control parameters and a maximal explanation of the data variance (or minimal data reconstruction error). The present section describes *Bergame*, an articulatory-acoustic model developed at ICP with the aim of mimicking as closely as possible experimental data gathered on the reference subject.

Bergame consists of: (1) a physiologically oriented linear articulatory model, based on the articulatory data measured from the cineradiofilm and the video labiofilm made on the reference subject; (2) a model of midsagittal-to-area function conversion based on the same subject; (3) an acoustic model.

A. The linear articulatory model

The principle of a linear articulatory model is to *calculate the position and shape of the various articulators as linear combinations of the articulatory control parameters.* The development of the model thus amounts to defining the control parameters and to determining the coefficients of these linear combinations. The nine parameters chosen for controlling the articulatory model stem directly from the previous component analysis: *JH*, *TB*, *TD*, *TT*, *TA*, *LY*, *LH*, *LP*, and *LV*, which are dimensionless, centered, and normalized. These parameters are, in most cases, orthogonal to each other, the exceptions (see Table IV) being due to the subject and language specific control strategies. The model equations are described in some detail in the following. The model behavior is illustrated in Fig. 8 by *articulatory nomograms*, i.e., the variations of the midsagittal contours resulting from variations of the articulatory control parameters from −3 to +3 with +1 steps.

1. Jaw

The jaw has been shown above to possess essentially one degree of freedom for this subject and the corpus analyzed. Jaw position is therefore controlled by the single parameter *JH* that defines *JawHei_mod* by the simple linear relation:

$$JawHei\_mod = JawHei\_mean + JawHei\_std \cdot JH,$$

where *JawHei_std* is the standard deviation of *JawHei* and *JawHei_mean* its mean over the corpus.

2. Tongue, midsagittal distances, and vocal tract outer contours

Since the tongue contours are attached to the grid lines, the next necessary step is to determine the position of the mobile parts of the grid system, namely *TngAdv* and *LarHei*. The modeled tongue advance, *TngAdv_mod*, was found to be almost linearly related to *TA*, *JH*, *TB*, and *TD*, and was therefore controlled by:

$$TngAdv\_mod = TngAdv\_mean + pred\_TngAdv\_JH\_TB\_TD\_TA \cdot [JH,TB,TD,TA],$$

where [*JH*,*TB*,*TD*,*TA*] is the matrix of control parameters, and *pred_TngAdv_JH_TB_TD_TA* are the associated coefficients determined by multiple linear regression. As seen in Sec. II E, *LarHei* is controlled only by *LY*, and not *LP* and *LH*, despite a slight correlation between lips and larynx, in order to ensure an independent control of lips and larynx in the model. *LY* is therefore partially correlated with a number of other control parameters, as seen in Table IV. Note that *TngAdv* and *LarHei* are reconstructed without error.

Finally, the abscissa of the whole tongue contour *Int_mod* (lines 1 to 28) is determined as linear combinations of the parameters *JH*, *TB*, *TD*, and *TT*:

$$Int\_mod = Int\_mean + pred\_Int\_JH\_TB\_TD\_TT \cdot [JH,TB,TD,TT].$$

Figure 8 displays articulatory nomograms for *JH*, *TB*, *TD*, *TT*, and *TA* as well.

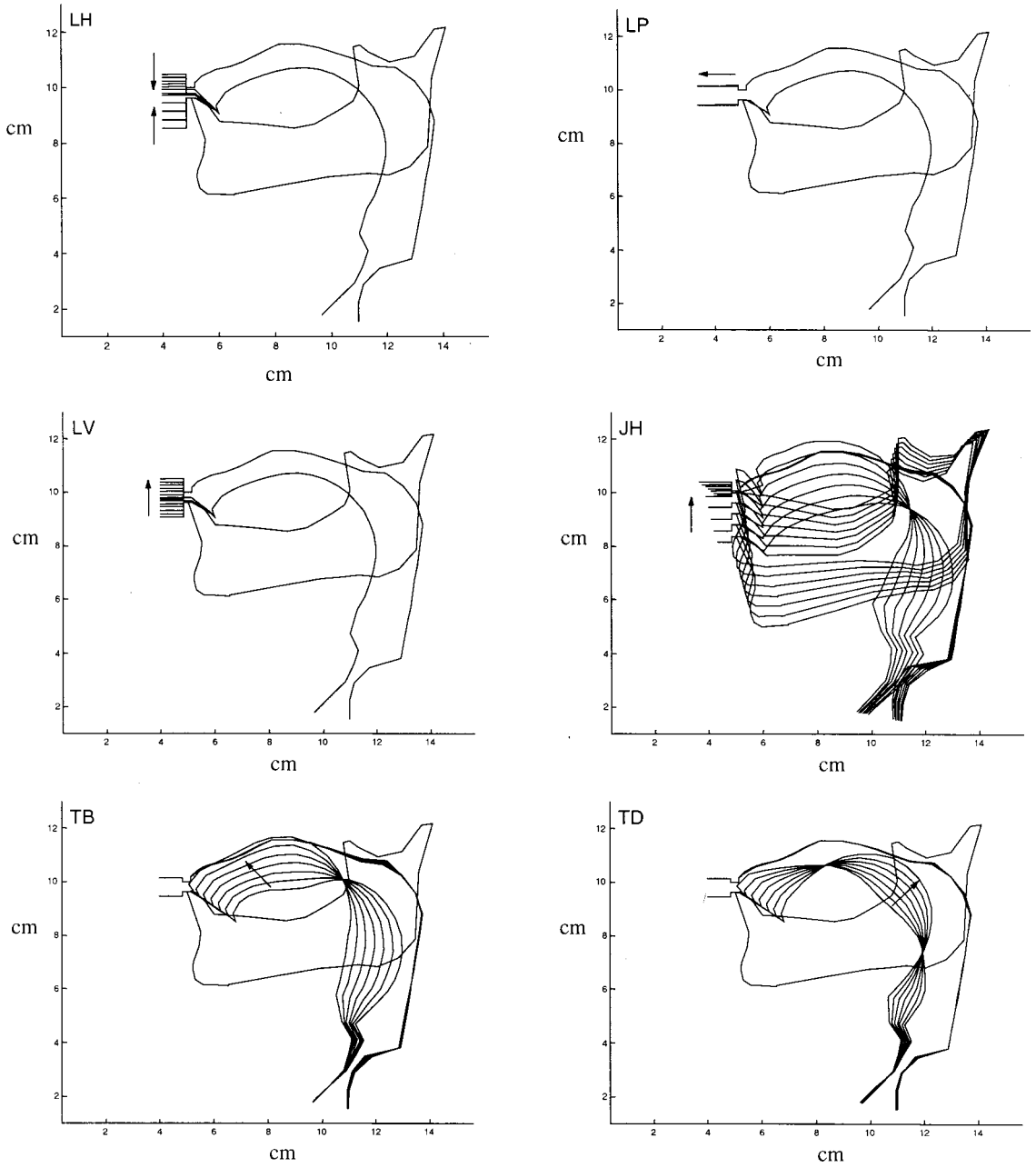


FIG. 8. Articulatory nomograms: variations of the midsagittal contours resulting from variations of the articulatory control parameters from −3 to +3 with +1 steps. Note that the movements of upper and lower lips are *opposite* for the *LH* nomogram, but *parallel* for the *LV* nomogram.

The midsagittal distances have been similarly handled. The abscissa of the vocal tract outer contours are computed as the sum of the abscissa of the tongue and the corresponding midsagittal distances, except for the hard palate region that is considered as a fixed contour.

3. Lips

It has been shown above that the lip geometry of the subject is best described with three degrees of freedom, represented by *LH*, *LP*, and *LV*, in addition to the contribution of *JH*. The lip horn, considered as the vocal tract region anterior to the upper incisor plane, is represented by a single tube section, with a length *ProLip_mod* proportional to the prediction of the *LipTop* dimension, with a proportionality factor of 0.6. This length reduction aims at approximately taking into account the fact that the lip corner position is not known, and that the effective acoustical end of the lip horn is located between the lip corner and the extremities of the lips measured by *ProTop* and *ProBot*. The parameter *ProLip_mod* is thus defined by:

$$ProLip\_mod = 0.6 \cdot (ProTop\_mean + pred\_ProTop\_JH\_LH \cdot [JH, LH]),$$

where the coefficients *pred_ProTop_JH_LH* are obtained by multiple linear regression.

Lip height *LipHei_mod* is similarly modeled as a linear combination of *JH*, *LH*, and *LP*, as well as lip width *A_mod*. Lip vertical position *LipTop_Mod* is also a linear combina-

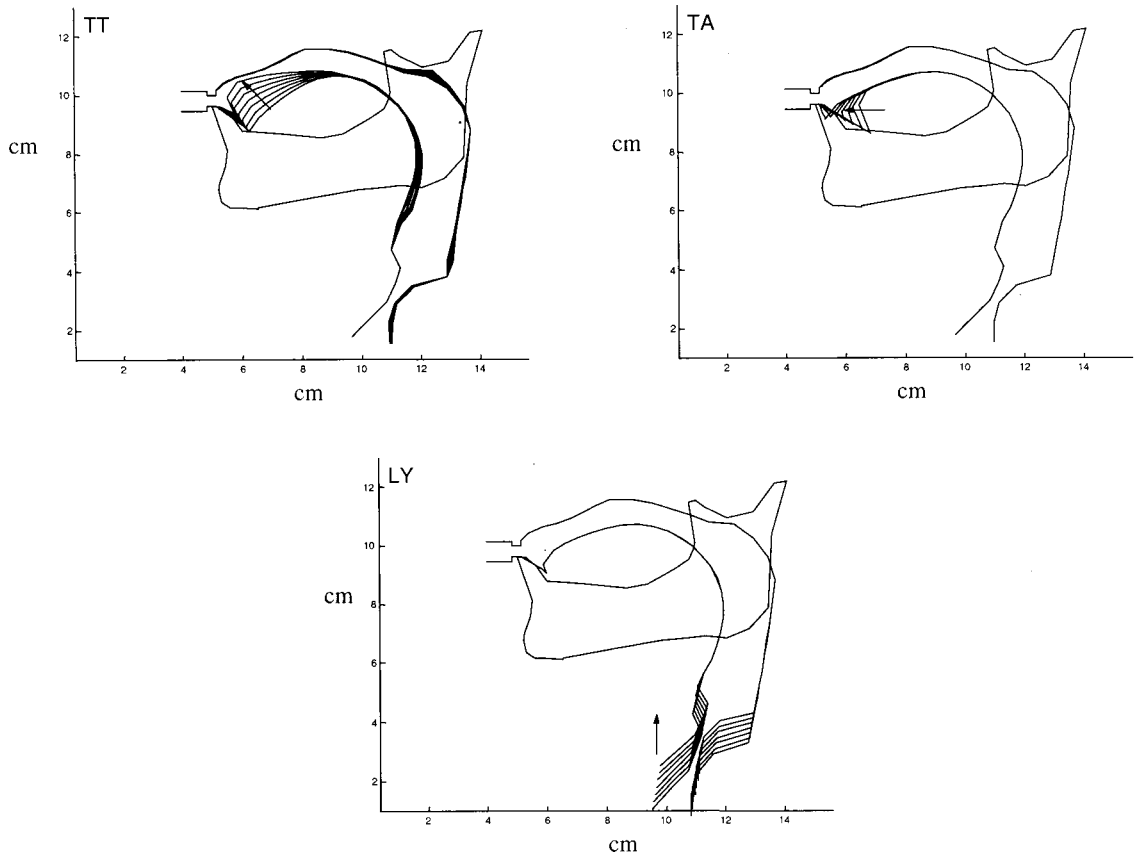


FIG. 8. (*Continued.*)

tion of *JH*, *LH*, *LP*, and *LV*. Figure 8 shows also the articulatory nomograms relating these lip parameters.

B. Midsagittal and area functions—Acoustic models

The midsagittal contour alone is not sufficient to derive the corresponding vocal tract acoustic features. Acoustic vocal tract models are indeed based on tube acoustics, and thus need a description of the vocal tract in terms of *area function*. Most studies devoted to the problem of converting midsagittal distances d to area functions S resulted in solutions based on the ''α,β model'' proposed by Heinz and Stevens (1965), where $S=\alpha\cdot d^{\beta}$: the principle consists in calculating the area of each vocal tract section as a power function of the corresponding midsagittal distance (cf., e.g., Beautemps *et al.*, 1995, for more details). Finally, vocal tract aeroacoustic simulations in the time or in the frequency domain allow the computation of the speech signal or speech acoustic characteristics from the area function (cf., e.g., Maeda, 1982; Badin and Fant, 1984; Mawass *et al.*, 2000).

1. Vocal tract

The midsagittal function represents the sagittal distances between the tongue contour and the outer vocal tract contour along the vocal tract midline, estimated for each section enclosed between two consecutive measurement grid lines. For each section, a quadrilateral can be defined in the midsagittal plane by the intersection points of the tongue contour and of the vocal tract outer contours with the corresponding two lines of the grid. The midsagittal distance for this section is calculated as the surface of the quadrilateral divided by the length of the section. Following Heinz and Stevens (1965), the vocal tract area function is estimated from the midsagittal function. It uses an extended version of a conversion model (Beautemps *et al.*, 1996) optimized for both vowels and consonants: the area function is then derived from the midsagittal function using a polynomial expression where the cross-sectional area S depends on both the midsagittal distance d and the x distance from the glottis measured along the vocal tract midline:

$$S(x,d)=\alpha_1(x)\cdot d+\alpha_2(x)\cdot d^{1.5}+\alpha_3(x)\cdot d^{2}+\alpha_4(x)\cdot d^{2.5}.$$

The $\alpha_i(x)$ functions are expressed as Fourier series, up to the third order, of $\pi.x/l_{\text{tot}}$, where the l_{tot} is the vocal tract length (including the lips):

$$\alpha_i(x)=a_{i0}(x)+\sum_{n=1}^{3} a_{in}\cdot\cos\left(n\frac{\pi}{l_{\text{tot}}}x\right)+\sum_{n=1}^{3} b_{in}\cdot\sin\left(n\frac{\pi}{l_{\text{tot}}}x\right).$$

The values of the Fourier coefficients (altogether, 28 parameters) were optimized so as to minimize, for the N selected configurations, the χ^2 distance between the four formants F_{ik} computed from the area function derived from the synthesized contours and the formants F_{ik}^{c} measured on the acoustic signal of the original data:

$$\chi^2=\sum_{i=1}^{N}\sum_{k=1}^{4}\frac{(F_{ik}^{c}-F_{ik})^2}{F_{ik}^{c}}.$$

At first, the optimization procedure was applied to a restricted set of eight vowels [aɛeiyuoø], in order to ensure an easy convergence. The results were then refined by applying

the same optimization procedure to the whole corpus, excluding only the data for which measurement of the four formants was not possible, i.e., excluding the 347 configurations mainly associated with occlusive consonants [pbdg].

Dang and Honda (1998) developed a similar polynomial decomposition where the coefficients, a function of the distance from the glottis, are determined by minimizing the difference between the estimated and MRI-based area functions for five Japanese vowels. The present procedure, developed before any 3-D vocal tract data were available for the subject, does not make use of 3-D data. However, Badin *et al.* (1998) subsequently acquired 3-D MRI data allowing the direct determination of both midsagittal contours and area functions for a set of vowel articulations for the same subject. These data have therefore been used to assess the quality of this algorithm: for the ten French vowels [aɛeiyuoɔøœ] the comparison between the areas directly estimated from the 3-D measurements (excluding the larynx region and the lips for which no MRI data were available) and those computed by the present procedure from the midsagittal contours estimated from the 3-D measurements has revealed a global root mean square (rms) error value lower than 0.6 cm^2. The relatively important rms error (more than 1 cm^2) observed in the low pharyngeal region for [aɔøœ] is probably due to the whispered production mode used to maintain the articulation during the long 3-D MRI data recording duration whose main consequence is a more constricted tongue in the back region (cf. Matsuda and Kasuya, 1999). This implies a decrease of the cross-sectional areas between the glottis and the epiglottis and probably modifies the relation between the midsagittal distances and the related area functions. Finally, in the uvular region, the small midsagittal distance measured is not representative of the entire cross-section, due to the fact that the main part of the velum body is concentrated in the midsagittal plane with free air flow on both sides. The consequence is an underestimation of the area inherent to the conversion model.

The fit between measured and reconstructed data was also assessed at the level of midsagittal functions. The midsagittal functions of the synthesized vocal tract contours have thus been compared to the midsagittal functions of the original data calculated with the measurement gridline system implemented in the model. The rms errors of the length and of the midsagittal distances are almost zero except for the larynx region where the errors can respectively reach 0.1 and 0.3 cm, probably due to the poor modeling of the tongue in this region (see Fig. 4). A maximum of 0.2 cm for the rms error on the midsagittal distances is also noted in the front part of the vocal tract between the last tongue point and the teeth, due to the fact that the sublingual cavity is not taken into account in the model.

2. Lip area

A possibility for computing lip area *S_mod* is the relation established for the first time by Fromkin (1964):

$$S\_mod = pred\_S\_A\_B \cdot A \cdot B,$$

where *A* and *B* are, respectively, the intra-oral lip width and height measured from the video labio-film. For the present subject, a value of 0.80 was found for *pred_S_A_B* by linear regression applied to the whole corpus. However, the lack of accuracy of *A_mod* resulted in a poor modeling of small areas. This method was therefore abandoned, and lip area was in practice calculated as a second order multilinear regression of the *JH*, *LH*, and *LP* components, for which the coefficients were optimized as to obtain the best fit to the lip area measured on video front pictures. With this modeling, we obtained 0.2 cm^2 for the rms error.

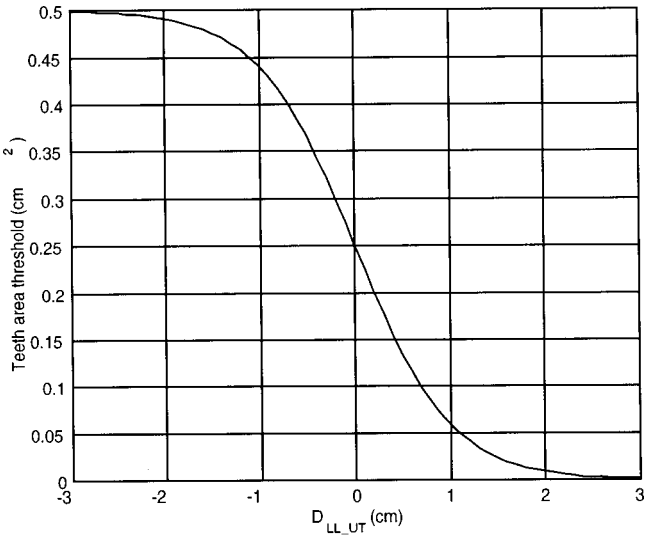


FIG. 9. Minimum threshold for the area at the incisors as a function of difference between lower lip position and upper incisor edge.

3. Acoustic effect of LV

In the absence of the horizontal jaw control parameter *JA*, only vertical movements are taken into account for the lower incisors, i.e., through *JH*. Therefore, there is no straightforward provision in the model for producing labio-dental constrictions by a combination of jaw retraction and lower lip elevation movements, which is the standard articulatory strategy for producing labio-dental fricatives.

This problem is overcome by a mechanism that uses the *LV* parameter for the production of the labio-dental constriction at the incisor section. The incisor section area is made an indirect function of lower lip vertical position, and thus of *LV*, by limiting it to a minimum threshold value function of the difference between lower lip position and upper incisor edge (see Fig. 9). This allows *LV* to be audible, i.e., to have acoustic consequences, at least in circumstances typical of labio-dentals where the lower lip has to be higher than the upper incisor edge in order to produce the proper constriction. This feature was particularly useful for the inversion of the articulatory-to-acoustic relation for fricatives (Mawass *et al.*, 2000).

4. Acoustic model

Acoustic transfer functions as well as formants and bandwidths were determined from these area functions by means of a frequency domain vocal tract acoustic model (Badin and Fant, 1984). A time domain reflection-type line analogue (Bailly *et al.*, 1994), extended to include improved

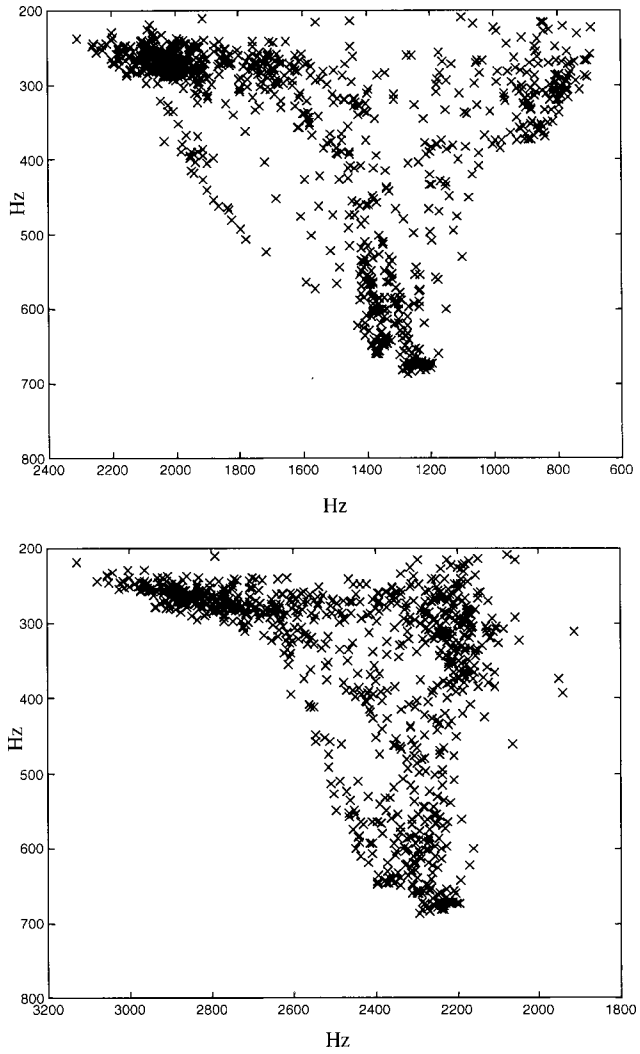


FIG. 10. Predicted $F2/F1$ (top) and $F3/F1$ (bottom) formant spaces (in Hz) for the vowels and consonants pooled.

voice (Pelorson *et al.*, 1996) and noise source models (Badin *et al.*, 1995), can also be driven by these area functions, in association with lung pressure and vocal cords parameters, to produce high quality articulatory synthesis (cf Mawass *et al.*, 2000).

The area functions and the derived formants have been computed for the assessment of the midsagittal to area function conversion, starting from the synthesized midsagittal contours. The rms error and the rms relative error on formants have both been calculated for the whole corpus (excluding 347 configurations for which the measurement of the four formants was not possible): 45 Hz (12.86%), 100 Hz (7.32%), 162 Hz, (6.47%), and 173 Hz (5.21%), respectively, for $F1$, $F2$, $F3$, and $F4$. The mean differences between the formants obtained by the model and those measured are 14 Hz, -65 Hz, and 26 Hz, respectively, for $F2$, $F3$, $F4$ (no significant difference was found for $F1$, except 27 Hz for the vowels).

The predicted maximal formant spaces are comparable to the measured ones (cf. Fig. 2 and Fig. 10). However, the computed formant $F1$ of [a] is about 34 Hz too low.

When modeling the area function of the four point vowels of their two American subjects, Baer *et al.* (1991) reported a deviation (in terms of the rms of the relative error) of 13%, 31%, and 13% on the measures for, respectively, $F1$, $F2$, and $F3$; these deviations are noticeably higher than those in the present study. From a set of 16 disyllabic utterances [hə'CV] and one sentence, Mermelstein (1973) obtained 10.3%, 4.9%, and 5.5% for the average absolute error on $F1$, $F2$, and $F3$; in terms of rms, these errors are still lower. Mermelstein's fits are thus clearly better than ours; however, it should be recalled that they were obtained on a much more restricted set of data.

V. DISCUSSION AND PERSPECTIVES

A. Summary

A linear component analysis of tongue contours and articulatory measures extracted from cineradio- and labio-films made on a reference subject revealed a relatively small number of degrees of freedom. The jaw appears, for the subject studied, to have mainly two degrees of freedom, related to the lower incisors vertical and horizontal movements. However, only the vertical component exerts a significant effect on tongue shape. The residue of tongue shape, once the contribution of the jaw has been removed, possesses four degrees of freedom: tongue body, tongue dorsum, tongue tip, and tongue advance. An extra parameter takes into account the larynx height variance. Similarly, the lip shape possesses, in addition to the jaw contribution, three degrees of freedom: lip protrusion, lip height, and lip vertical elevation. These nine parameters are mostly independent of each other (cf. Table IV), except for *LY* that is correlated with lips and tongue parameters and the correlation between *LP* and *TD*. These degrees of freedom are specific to the vocal tract and articulators of one subject uttering one specific corpus in one language. The corpus was designed to include as many French vowels and consonants as possible. A linear articulatory model was developed based on these data; it explains 96% of the tongue data variance, with an rms reconstruction error of about 0.09 cm. It was complemented by a model converting the midsagittal contours to an area function based on a fitting of midsagittal functions and formant frequencies. Finally this model allows the calculation of formants with rms errors of 45 Hz for $F1$, 100 Hz for $F2$, and 162 Hz for $F3$ over the corpus. To the knowledge of the authors, no such comprehensive model has been developed so far; most of the available models deal with vowels only, while others do not include acoustics.

B. Choice of subject and corpus

The development of such an articulatory-acoustic model based on a specific reference subject was motivated by the need for a model that could fit a real subject's midsagittal profiles of French fricative consonants, plosives, and vowels, as well as formants, with a fairly high degree of accuracy for a large number of configurations. The possibility now exists to investigate in detail the articulatory strategies employed by the subject, and in particular coarticulatory strategies (cf. Mawass *et al.*, 2000, or Vilain *et al.*, 1998). One may argue that no general conclusions may be drawn from such studies, as they are supported by one single subject's data. However,

we were very much aware of the risk of blurring out clear individual articulatory strategies employed by individual subjects when merging together several subjects' data, and therefore made the choice of a single subject for the present study. Similar analyses are under way for other subjects, in order to determine which features may be considered as general and which ones as more subject-specific (Bailly *et al.*, 1998; Vilain *et al.*, 1998; Engwall and Badin, 1999). These studies will also allow us to investigate the number of degrees of freedom of the jaw involved in speech.

The influence of the number of items used for the linear analysis was studied for the present subject by Badin *et al.* (1998); they found that, by choosing the contour samples, i.e., by selecting only vowel and consonant targets in the initial corpus, an articulatory model was produced that represented the whole corpus data with an accuracy close to that obtained when the full model based on the whole corpus was used. More specifically, they showed that the data reconstruction error, computed as the rms error of the abscissa of the tongue contour along each grid line for the 1222 images of the available corpus of midsagittal contours, was 0.09 cm, 0.11 cm, and 0.17 cm when the model was elaborated using, respectively, 1222, 20, and 8 configurations. This justifies the elaboration of models from a much lower number of articulations, and thus, in particular, the use of MRI images instead of x-ray images.

C. Comparison of the degrees of freedom found in other studies

Degrees of freedom are clearly subject and corpus-dependent but their number and their definition are closely related to the method used to explain the whole data set variance. The linear component analysis used by Maeda (1979, 1990) is sometimes referred to as a two-way factor analysis of the variance, where one mode corresponds to the predictors and the other one to the matrix of coefficients of the linear combinations. Using this principle, Maeda (1979) extracted one loading factor for the jaw, and three for the residual midsagittal tongue data to explain 98% of the variance for a corpus made of 400 frames of $[pV_1CV_2]$ ([aiu] and [dg]) sequences uttered by one subject. In an extended corpus of 519 frames corresponding to 10 French sentences, three supplementary components were obtained for the lips including the frontal lip-opening shapes, and four tongue degrees of freedom explaining 88% of the variance (Maeda, 1990). Finally, for these data, Sanguineti *et al.* (1998) imposed two degrees of freedom for the jaw (protrusion and rotation), and one for the larynx, and obtained three other degrees of freedom for the tongue residue from a similar analysis in the so-called λ-space.

The PARAFAC method (Harshman *et al.*, 1977; Nix *et al.*, 1996; Hoole, 1999) is a three-way factor generalization where the third mode corresponds to linear coefficients that account for differences between subjects. Harshman *et al.* (1977) derived two components for the description of representative midsagittal tongue and lip shapes of ten English vowels uttered by five subjects. Hoole (1999) proposed a two-factor PARAFAC model of the German vowels in a symmetrical stop consonant context, plus an additional PCA component to capture the subject-specific nonvocalic behavior of the tongue. More recently, Hoole *et al.* (2000) extracted a two-factor PARAFAC solution that explained 90% of the variance from a set of MRI midsagittal tongue contours measured during the production of seven German vowels by nine speakers. The first component captured the dimension low-back to high-front, and the second was associated with the mid-front to high-back motion. The complex effect of the first component can be decomposed in a co-variation of *JH* and *TB*, the second component being related to *TD*. Ultimately, and to the knowledge of the authors, no analysis based on the PARAFAC method has been realized with an imposed jaw component. To conclude, no analysis based on PARAFAC principles has shown success in explaining the large and phonetically varied data obtained from multiple speakers.

Bailly *et al.* (1998) studied the synergy between tongue and jaw for three subjects, including the present reference subject. They found that the two other subjects used a fairly strong synergy: the amplitude of the tongue movements measured at the tip and at the root that were correlated with jaw movements were about twice as large as might be expected from the simple mechanical carrying effect of the jaw. In other words, the jaw and the tongue shared the execution of the tongue movements. The present subject does not use this synergy: the tongue is not so active, and appears to be passively carried by the jaw. However, all three subjects' articulators had qualitatively the same degrees of freedom. This synergy is still a crucial issue for understanding coarticulation strategies.

D. Perspectives

The principles of this work have been duplicated for the modeling of Swedish midsagittal tongue shapes (Engwall and Badin, 1999). Over 90% of the variance is explained by the four tongue degrees of freedom *JH*, *TB*, *TD*, and *TT*.

One of the main issues in the analysis of speech degrees of freedom is the possibility to build a linear articulatory model that takes into account the redundant feature of the articulators shapes. For instance, it can help to reconstruct complete tongue shapes from a reduced number of articulatory measurement points, such as those provided by electromagnetic articulometry. Badin *et al.* (1997) used the present model to retrieve, from one coil on the lower incisor and three coils on the tongue of the reference subject, the tongue shape as well as the articulatory control parameters *JH*, *TB*, *TD*, *TT*, and *TA* with a fairly good accuracy. This may be useful for investigating speech coarticulation and synergetic strategies (cf., e.g., Vilain *et al.*, 1998), for testing hypotheses of the Frame/Content concept in the child's language development (Vilain *et al.*, 1999), or evaluating the adaptability of speech articulation to various linguistic tasks and environmental conditions such as changes illustrated by the Lombard reflect (Beautemps *et al.*, 1999).

The present articulatory-acoustic model can also be used to derive, by inversion, articulatory control parameters from formants measured in other utterances produced by the same subject (Mawass *et al.*, 2000). These data, in conjunction

with aerodynamic data obtained for the same subject, have been used for the articulatory synthesis of French fricatives (Mawass *et al.*, 2000).

Another extension of the present study is the third dimension. 3-D MRI images have been recorded for the same subject, and a 3-D linear articulatory model is being developed according to the same approach (Badin *et al.*, 1998, 2000); the new model has been elaborated in such a way that part of its control parameters are identical with those of the present midsagittal model, which opens the possibility of inheriting knowledge already acquired for the midsagittal plane, while acquiring new features such as the capability of producing lateral consonants. Finally, the modeling of the velum from MRI midsagittal data should complement the present model.

ACKNOWLEDGMENTS

This work has been partially funded by the European Community (ESPRIT/BR project *Speech Maps* No. 6975), and by the Rhône-Alpes Agency for Social and Human Sciences (ARASSH) (project ''A Virtual Talking Head: Data and models in speech production''). It has benefited from the valuable help of many people to whom the authors are very much indebted: Bernard Gabioud (who initiated a part of this work in the framework of the *Speech Maps* project), Tahar Lallouache (for the lip measurements), Shinji Maeda (for the initial version of the contour edition program, and more importantly for having largely inspired this work), Gilbert Brock, Péla Simon, and Jean-Pierre Zerling (for their expertise on cineradiography), Agnes Hennel (for access to the cineradiography equipment at the Strasbourg Schiltigheim Hospital), Thierry Guiard-Marigny and the late Christian Benoît (for their help on data gathering and processing), Christian Abry (for many stimulating discussions), Marija Tabain (for polishing our French English), as well as many other colleagues at ICP, Grenoble. We have also greatly appreciated the pertinent comments and careful editorial advice of Anders Löfqvist and two anonymous reviewers.

[1]Note that *TngTip* is identical to the last point of the tongue contour abscissa vector.

Abry, C., Badin, P., and Scully, C. (**1994**). ''Sound-to-gesture inversion in speech: The *Speech Maps* approach,'' in *Advanced Speech Applications*, edited by K. Varghese, S. Pfleger, and J. P. Lefèvre (Springer, Berlin), pp. 182–196.

Badin, P., Baricchi, E., and Vilain, A. (**1997**). ''Determining tongue articulation: from discrete fleshpoints to continuous shadow,'' in *Proceedings of the 5th EuroSpeech Conference* (University of Patras, Wire Communication Laboratory, Patras, Greece), Vol. 1, pp. 47–50.

Badin, P., Mawass, K., and Castelli, E. (**1995**). ''A model of frication noise source based on data from fricative consonants in vowel context,'' in *Proceedings of the 13th International Congress of Phonetic Sciences*, edited by K. Elenius and P. Branderud (Arne Strömbergs Grafiska Press, Stockholm, Sweden), Vol. 2, pp. 202–205.

Badin, P., Bailly, G., Raybaudi, M., and Segebarth, C. (**1998**). ''A three-dimensional linear articulatory model based on MRI data,'' in *Proceedings of the Third ESCA/COCOSDA International Workshop on Speech Synthesis* (Jenolan Caves, Australia), pp. 249–254.

Badin P., Motoki, K., Miki, N., Ritterhaus, D., and Lallouache, T. M. (**1994a**). ''Some geometric and acoustic properties of the lip horn,'' J. Acoust. Soc. Jpn. (E) **15**, 243–253.

Badin, P., Borel, P., Bailly, G., Revéret, L., Baciu, M., and Segebarth, C. (**2000**). ''Towards an audiovisual virtual talking head: 3D articulatory modeling of tongue, lips and face based on MRI and video images,'' in *Proceedings of the 5th Seminar on Speech Production: Models and Data & CREST Workshop on Models of Speech Production: Motor Planning and Articulatory Modelling* (Kloster Seeon, Germany), pp. 261–264.

Badin, P., and Fant, G. (**1984**). ''Notes on vocal tract computation,'' Speech Transmission Laboratory—Quarterly Progress Status Report Vol. 2-3/1984, pp. 53–108.

Badin, P., Shadle, C. H., Pham Thi Ngoc, Y., Carter, J. N., Chiu, W., Scully, C., and Stromberg, K. (**1994b**). ''Frication and aspiration noise sources: contribution of experimental data to articulatory synthesis,'' in *Proceedings of the 3rd International Conference on Spoken Language Processing* edited by Mike Edington (Yokohama, Japan), Vol. 1, pp. 163–166.

Baer, T., Gore, J. C., Gracco, L. C., and Nye, P. W. (**1991**). ''Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels,'' J. Acoust. Soc. Am. **90**, 799–828.

Bailly, G. (**1993**). ''Resonances as possible representations of speech in the auditory-to-articulatory transform,'' in *Proceedings of the 3rd Eurospeech Conference on Speech Communication and Technology* (Berlin), Vol. 3, pp. 1511–1514.

Bailly, G., Badin, P., and Vilain, A. (**1998**). ''Synergy between jaw and lips/tongue movements: Consequences in articulatory modelling,'' in *Proceedings of the 5th International Conference on Spoken Language Processing*, edited by R. H. Mannell, J. Robert-Ribes, and E. Vatikiotis-Bateson (Australian Speech Science and Technology Association, Inc., Sydney, Australia), Vol. 5, pp. 1859–1862.

Bailly, G., Castelli, E., and Gabioud, B. (**1994**). ''Building prototypes for articulatory speech synthesis,'' in *Proceedings of the 2nd ESCA/IEEE Workshop on Speech Synthesis* (New York), pp. 9–12.

Beautemps, D., Badin, P., and Laboissière, R. (**1995**). ''Deriving vocal-tract area functions from midsagittal profiles and formant frequencies: A new model for vowels and fricative consonants based on experimental data,'' Speech Commun. **16**, 27–47.

Beautemps, D., Borel, P., and Manolios, S. (**1999**). ''Hyper-articulated speech: Auditory and visual intelligibility,'' in *Proceedings of the 6th European Conference on Speech Communication and Technology*, Vol. 1 (Budapest, Hungary), pp. 109–112, September 1999.

Beautemps, D., Badin, P., Bailly, G., Galván, A., and Laboissière, R. (**1996**). ''Evaluation of an articulatory-acoustic model based on a reference subject,'' in *Proceedings of the 4th Speech Production Seminar* (Autrans, France), pp. 45–48.

Boë, L. J., Gabioud, B., Schwartz, J. L., and Vallée, N. (**1995**). ''Towards the unification of vowel spaces,'' in *Proceedings of the XIIIth International Congress of Phonetic Sciences*, edited by K. Elenius and P. Branderud (Arne Strombergs Grafiska Press, Stockholm, Sweden), Vol. 4, pp. 582–585.

Bothorel, A., Simon, P., Wioland, F., and Zerling, J. P. (**1986**). ''Cinéradiographie des voyelles et consonnes du francais [Cineradiography of vowels and consonants in French],'' Trav. de l'Inst. de Phonétique de Strasbourg, 296 pp.

Coker, C., and Fujimura, O. (**1966**). ''Model for specification of the vocal-tract area function,'' J. Acoust. Soc. Am. **40**, 1271.

Dang, J., and Honda, K. (**1998**). ''Speech production of vowel sequences using a physiological articulatory model,'' in *Proceedings of the 5th International Conference on Spoken Language Processing*, edited by Robert H. Mannell and Jordi Robert-Ribes, Vol. 5 (Sydney, Australia R.H.), pp. 1767–1770.

Demolin, D., Giovanni, A., Hassid, S., Heim, C., Lecuit, V., and Soquet, A. (**1997**). ''Direct and indirect measurements of subglottic pressure,'' *Proceedings of Larynx 97* (Marseille, France), pp. 69–72.

Djéradi, A., Guérin, B., Badin, P., and Perrier, P. (**1991**). ''Measurement of the acoustic transfer function of the vocal tract: a fast and accurate method,'' J. Phonetics **19**, 387–395.

Edwards, J., and Harris, K. S. (**1990**). ''Rotation and translation of the jaw during speech,'' J. Speech Hear. Res. **33**, 550–562.

Engwall, O., and Badin, P. (**1999**). ''Collecting and analyzing two- and three-dimensional MRI data for Swedish,'' Tal Musik Hörsel, Quarterly Progress Status Report, Stockholm Vol. 3-4, pp. 11–38.

Fowler, C. A., and Saltzman, E. (**1993**). ''Coordination and coarticulation in speech production,'' Language and Speech **36**, 171–195.

Fromkin, V. A. (**1964**). ''Lip positions in American English vowels,'' Language and Speech **7**, 215–225 .

Gabioud, B. (**1994**). ''Articulatory models in speech synthesis,'' in *Funda-*

mentals of Speech Synthesis and Speech Recognition, edited by E. Keller (Wiley, Chichester), pp. 215–230.
Harshman, R., Ladefoged, P., and Goldstein, L. (**1977**). ''Factor analysis of tongue shape,'' J. Acoust. Soc. Am. **62**, 693–707.
Heinz, J. M., and Stevens, K. N. (**1965**). ''On the relations between lateral cineradiographs, area functions, and acoustic spectra of speech,'' *Proceedings of the Fifth International Congress of Acoustics* (Liége, Belgium), Paper A44.
Hoole, P. (**1999**). ''On the lingual organization of the German vowel system,'' J. Acoust. Soc. Am. **106**, 1020–1032.
Hoole, P., and Kroos, C. (**1998**). ''Control of larynx height in vowel production,'' in *Proceedings of the 5th International Conference on Spoken Language Processing,* edited by R. H. Mannell, J. Robert-Ribes, and E. Vatikiotis-Bateson (Australian Speech Science and Technology Association, Inc., Sydney, Australia), Vol. 2, pp. 531–534.
Hoole, P., Wismüller, A., Leisinger, G., Kroos, C., Geumann, A., and Inoue, M. (**2000**). ''Analysis of tongue configuration in multi-speaker, multi-volume MRI data,'' in *Proceedings of the 5th Seminar on Speech Production: Models and Data & CREST Workshop on Models of Speech Production: Motor Planning and Articulatory Modelling* (Kloster Seeon, Germany), pp. 157–160.
Kelso, J. A. S., Saltzman, E. L., and Tuller, B. (**1986**). ''The dynamical theory of speech production: Data and theory,'' J. Phonetics **14**, 29–60.
Laboissière, R., Ostry, D. J., and Feldman, A. G. (**1996**). ''Control of multi-muscle systems: Human jaw and hyoid movements,'' Biol. Cybern. **74**, 373–384.
Lallouache, M. T. (**1990**). ''Un poste Visage-Parole. Acquisition et traitement de contours labiaux [A ''Face-Speech'' workstation. Acquisition and processing of labial contours],'' *Proceedings of the 18th Journées d'Etude sur la Parole* (Montréal, Canada), pp. 282–286.
Liljencrants, J. (**1971**). ''A Fourier series description of the tongue profile,'' Speech Transmission Laboratory—Quarterly Progress Status Report Vol. 4/1971, pp. 9–18.
Lindblom, B. E. F. and Sundberg, J. E. F. (**1971**). ''Acoustical consequences of lip, tongue and jaw movements,'' J. Acoust. Soc. Am. **50**, 1166–1179.
Lindblom, B. E. F., Lubker, J., and Gay, T. (**1979**). ''Formant frequencies of some fixed-mandible vowels and a model of speech-motor programming by predictive simulation,'' J. Phonetics **7**, 141–161.
Maeda, S. (**1979**). ''Un modèle articulatoire de la langue avec des composantes linéaires,'' *Proceedings of the 10th Journées d'Etude sur la Parole* (Grenoble, France), pp. 152–163.
Maeda, S. (**1982**). ''A digital simulation method of the vocal tract system,'' Speech Commun. **1**, 199–299.
Maeda, S. (**1990**). ''Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model,'' in *Speech Production and Speech Modeling,* edited by W. J. Hardcastle and A. Marchal (Kluwer Academic, The Netherlands), pp. 131–149.
Maeda, S. (**1991**). ''On articulatory and acoustic variabilities,'' J. Phonetics **19**, 321–331.
Matsuda, M., and Kasuya, H. (**1999**). ''Acoustic nature of the whisper,'' in *Proceedings of the 6th European Conference on Speech Communication and Technology,* Vol. 1 (Budapest, Hungary), pp. 133–136, September 1999.
Mawass, K., Badin, P., and Bailly, G. (**2000**). ''Synthesis of French fricatives by audio-video to articulatory inversion,'' Acta Acoustica **86**, 136–146.
Mermelstein, P. (**1973**). ''Articulatory model for the study of speech production'' J. Acoust. Soc. Am. **53**, 1070–1082.
Nix, D. A., Papcun, G., Hogden, J., and Zlokarnik, I. (**1996**). ''Two cross-linguistic factors underlying tongue shapes for vowels,'' J. Acoust. Soc. Am. **99**, 3707–3717.
Ostry, D., Vatikiotis-Bateson, E., and Gribble, P. (**1997**). ''An examination of the degrees of freedom of human jaw motion in speech and mastication,'' J. Acoust. Soc. Am. **40**, 1341–1351.
Payan, Y., and Perrier, P. (**1997**). ''Synthesis of V-V sequences with a 2D biomechanical tongue model controlled by the equilibrium point hypothesis,'' Speech Commun. **22**, 185–205.
Pelorson, X., Hirschberg, A., Wijnands, A. P. J., Bailliet, H., Vescovi, C., and Castelli, E. (**1996**). ''Description of the flow through the vocal cords during phonation. Application to voiced sounds synthesis,'' Acta Acoustica **82**, 358–361.
Perkell, J. S. (**1974**). ''A physiological-oriented model of the tongue activity during speech production,'' Ph.D. dissertation, MIT, Cambridge.
Perkell, J. S. (**1991**). ''Models, theory and data in speech production,'' *Proceedings of the XIIth International Congress of Phonetic Sciences* (Université de Provence, Aix-en-Provence, France), Vol. 1, pp. 182–191.
Perrier, P., Ostry, D. J., and Laboissière, R. (**1996**). ''The equilibrium point hypothesis and its application to speech motor control,'' J. Speech, Language, and Hear. Res. **39**, 365–578.
Perrier, P., Payan, P., Perkell, J. S., Zandipour, M., Pelorson, X., Coisy, V., and Matthies, M. (**2000**). ''An attempt to simulate fluid-walls interactions during velar stops,'' in *Proceedings of the 5th Seminar on Speech Production: Models and Data & CREST Workshop on Models of Speech Production: Motor Planning and Articulatory Modelling* (Kloster Seeon, Germany), pp. 149–152.
Revéret, L., and Benoît, C. (**1998**). ''A new 3D lip model for analysis and synthesis of lip motion in speech production,'' in *Proceedings of the International Conference on Auditory-Visual Speech Processing/Second ESCA ETRW on Auditory-Visual Speech,* edited by D. Burnham, J. Robert-Ribes, and E. Vatikiotis-Bateson (Terrigal-Sydney, Australia), pp. 207–212.
Sanguineti, V., Laboissière, R., and Ostry, D. J. (**1998**). ''A dynamic biomechanical model for neural control of speech production,'' J. Acoust. Soc. Am. **103**, 1615–1627.
Scully, C. (**1991**). ''The representation in models of what speakers know,'' in *Proceedings of the XIIth International Congress of Phonetic Sciences* (Université de Provence, Aix-en-Provence, France), Vol. 1, pp. 192–197.
Shadle, C. H., and Scully, C. (**1995**). ''An articulatory-acoustic-aerodynamic analysis of [s] in VCV sequences,'' J. Phonetics **23**, 53–66.
Sorokin, V. N., Gay, T., and Ewan, W. G. (**1980**). ''Some biomechanical correlates of jaw movements,'' J. Acoust. Soc. Am. Suppl. 1 **68**, S32.
Stark, J., Lindblom, B., and Sundberg, J. (**1996**). ''APEX: An articulatory synthesis model for experimental and computational studies of speech production,'' TMH-QPSR 2/1996, pp. 45–48.
Stetson, R. H. (**1928**). ''Motor phonetics. A study of speech movements in action,'' Archives Néerlandaises de Phonétique Expérimentale **3**, 216.
Stromberg, K, Scully, C., Badin, P., and Shadle, C. H. (**1994**). ''Aerodynamic patterns as indicators of articulation and acoustic sources for fricatives produced by different speakers,'' Institute of Acoustics **16**, 325–333.
Tiede, M. K. (**1996**). ''An MRI-based study of pharyngeal volume contrasts in Akan and English,'' J. Phonetics **24**, 399–421.
Vilain, A., Abry, C., and Badin, P. (**1998**). ''Coarticulation and degrees of freedom in the elaboration of a new articulatory plant: Gentiane,'' in *Proceedings of the 5th International Conference on Spoken Language Processing,* edited by R. H. Mannell, J. Robert-Ribes, and, E. Vatikiotis-Bateson (Australian Speech Science and Technology Association, Inc., Sydney, Australia), Vol. 7, pp. 3147–3150.
Vilain, A., Abry, C., Badin P., and Brosda, S. (**1999**). ''From idiosyncratic pure frames to variegated babbling: Evidence from articulatory modeling,'' in *Proceedings of the 14th International Congress of Phonetic Sciences,* edited by J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, and A. C. Bailey (Congress organizers at the Linguistics Department, University of California at Berkeley, San Francisco, CA), Vol. 3, pp. 2497–2500.
Westbury, J. R. (**1988**). ''Mandible and hyoid bone movements during speech,'' J. Speech Hear. Res. **31**, 405–416.
Westbury, J. R. (**1994**). ''On coordinate systems and the representation of articulatory movements,'' J. Acoust. Soc. Am. **95**, 2271–2273.
Wilhelms-Tricarico, R. (**1995**). ''Physiological modeling of speech production: Methods for modeling soft-tissue articulators,'' J. Acoust. Soc. Am. **97**, 3085–3098.
Zerling, J. P. (**1984**). ''Phénomènes de nasalité et de nasalisation vocaliques: Étude cinéradiographique pour deux locuteurs [in French],'' Trav. de l'Inst. de Phonétique de Strasbourg **16**, 241–266.

The relationship between spectral characteristics and perceived hypernasality in children

Ryuta Kataoka,[a] Donald W. Warren, David J. Zajac, Robert Mayo, and Richard W. Lutz
UNC Craniofacial Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7450

(Received 6 October 1999; accepted for publication 5 February 2001)

The purpose of this study was to quantify perceived hypernasality in children. One-third octave spectra of the isolated vowel [i] were obtained from 32 children with cleft palate and 5 children without cleft palate. Four experienced listeners rated the severity of hypernasality of the 37 speech samples using a 6-point equal-appearing interval scale. When the average 1/3-octave spectra from the hypernasal group and the normal resonance group were compared, spectral characteristics of hypernasality were identified as increased amplitudes between $F1$ and $F2$ and decreased amplitudes in the region of $F2$. Based on the findings of the children's speech, 36 speech samples with manipulated spectral characteristics were used to minimize the influences of voice source characteristics on perceived hypernasality. Multiple regression analysis revealed a high correlation ($R=0.84$) between the amplitudes of 1/3-octave bands (1 k, 1.6 k, and 2.5 kHz) and the perceptual ratings. Increased amplitudes of bands between $F1$ and $F2$ (1 k, 1.6 kHz) and decreased amplitude of the band of $F2$ (2.5 kHz) was associated with an increasing perceived hypernasality. These results suggest that the amplitudes of the three 1/3-octave bands are appropriate acoustic parameters to quantify hypernasality in the isolated vowel [i]. © *2001 Acoustical Society of America.*
[DOI: 10.1121/1.1360717]

PACS numbers: 43.70.Gr, 43.70.Dn, 43.71.Bp [AL]

I. INTRODUCTION

Quantitative evaluation of hypernasality is important to monitor speech development and to determine whether surgical intervention or speech therapy is necessary. Several instrumental techniques are available for clinicians to measure velopharyngeal inadequacy or the severity of hypernasality (e.g., aerodynamic assessment and nasometry, respectively). The acoustic analysis of speech samples recorded by a microphone seems to have an advantage for evaluating the severity of hypernasality because it analyzes the same signals that are perceived by listeners. No acoustic method, however, has been used to evaluate hypernasality in a clinical setting. The goal of this study was to develop an acoustic method to quantify perceived severity of hypernasality in children with velopharyngeal inadequacy.

Using a comparison of nasalized and non-nasalized vowels in the same subject or model, articulatory synthesis and formant analysis have been used to investigate the acoustic characteristics of nasalization. In an attempt to quantify spectral properties of nasalized vowels, Hawkins and Stevens (1985) acoustically synthesized nasalized and non-nasalized vowels by manipulating both the frequencies and the spacing of the first nasal pole–zero pair. They speculated that the acoustic property corresponding to perceived nasality was the degree of spectral prominence in the $F1$ region.

To evaluate this hypothesis, Chen (1995) proposed the acoustic parameter $A1-P1$ in dB. $A1$ is the amplitude of $F1$, $P1$ is the amplitude of the second nasal peak in the vicinity of $F1$. The author reported several concerns related to the acoustic parameter $A1-P1$. The nasal peak was not easily detected when it appeared in the vicinity of a formant. The frequency and the amplitude of the nasal peak may show some variability when the fundamental frequency is high, which often occurs in children's speech. In addition, the parameter $A1-P1$ was vowel dependent. The reported correlation coefficient of nasality judgments and normalized $A1-P1$ for different vowels was -0.76.

Chen (1997) proposed another acoustic parameter $A1-P0$. $P0$ is the amplitude of the first nasal peak. The parameter $A1-P0$ was not applicable for the nasalized vowel [i] because the first nasal peak sometimes appears in the frequency region of $F1$. This parameter was also influenced by vocal breathiness.

The findings of these studies suggest that the severity of nasalization cannot be measured accurately by formant analysis when the fundamental frequency is high, which is common in children's speech. In addition to these, Bakkum, Plomp, and Pols (1995) reported that formant analysis was not suitable for real-time application. Therefore, its use is not suitable for a clinical setting.

We have used a 1/3-octave analysis to represent the acoustic characteristics of hypernasality in children as an auditory spectrum (Kataoka, 1988; Kataoka *et al.*, 1996, 2001). A 1/3-octave analysis appears to be compatible with the concept of formant analysis as an auditory model (Bakkum *et al.*, 1995). This bandwidth was selected because it compared well with the critical bandwidth of the analyzing mechanism utilized by the ear (Pols *et al.*, 1969). The differences in perceived quality of vowels, including the severity of hypernasality, should be represented as the differences in the 1/3-octave spectrum. Therefore, a 1/3-octave analysis is considered to be suitable to quantify perceived severity of

[a]Author to whom correspondence should be addressed; also at First Department of Oral and Maxillofacial Surgery, Showa University 2-1-1 kitasenzoku, ohta-ku, Tokyo 145-8515, Japan. Electronic mail: kataoka@senzoku.showa-u.ac.jp

hypernasality. Furthermore, a 1/3-octave analysis is suitable for the clinical setting because it can be performed in real time.

The average 1/3-octave spectra of a normal resonance group and hypernasal group were compared. From the comparison, the spectral characteristics of hypernasality in children with cleft palate during the vowel [i] were identified.

Kataoka *et al.* (2001) reported that voice quality distortions and a particular type of spectral change that relate to the severity of hypernasality over the duration of vowel could be factors that influence perceived hypernasality. We hypothesized that a perceptual evaluation of hypernasality has a better relationship to the acoustic characteristics of hypernasality within a speech sample when acoustic factors that may produce perceptual variability among listeners are controlled. Speech samples with manipulated spectral characteristics, therefore, were used to control the acoustic parameters (e.g., fundamental frequency, duration, sound level, voice quality, and spectral change over time). The spectral characteristics of the speech samples were manipulated using a 1/3-octave equalizer based on the spectral characteristics of children's speech samples identified in Sec. II.

In Sec. III, the relationship between the spectral characteristics and perceived hypernasality was examined by multiple regression analysis using these manipulated speech samples. The predicted hypernasality scores obtained from the multiple regression equation using the acoustic parameters that represent the acoustic characteristics of hypernasality should correlate closely to reliable perceptual evaluations. Thus, both more reliable perceptual evaluations of hypernasality and the acoustic method to evaluate perceived hypernasality in the vowel [i] could be obtained if the acoustic factors are controlled.

II. ACOUSTIC CHARACTERISTICS OF HYPERNASALITY IN CHILDREN WITH CLEFT PALATE

A. Method

1. Subjects

Speech samples were obtained from 32 children who were seen for evaluation at the UNC Craniofacial Center. All children, 14 boys and 18 girls, had a diagnosis of cleft palate, velopharyngeal impairment of varying degrees, or both. They were clinically judged to have hypernasal speech by two speech-language pathologists. The mean age for this group was 9.0 years with a range from 5 to 13 years for boys and 5 to 15 years for girls.

Speech samples were also obtained from 32 children without cleft palate (control group). They consisted of 20 boys and 12 girls who ranged in age from 6 to 13 years, with a mean age of 9.5 years. They were free of any known craniofacial anomalies and were judged to exhibit normal phonation and resonance.

2. Speech sample and recording procedure

Subjects were instructed to sustain the vowel [i] with natural loudness and pitch. A model was provided to facilitate normal production. In order to facilitate ratings of hypernasality, the vowel was prolonged. An omnidirectional electret condenser microphone (Sony ECM 44S) was placed approximately 15 cm from the subjects' lips. The speech samples were recorded using the PERCI-SARS system (Microtronics Corp., Chapel Hill, NC). Each speech sample was low-pass filtered at 7500 Hz with a roll-off of -80 dB per octave and digitized at 20 kHz with 16-bit resolution. All subjects' speech samples were recorded in a sound-attenuated booth.

The examination of the relationship between the perceptual evaluation and spectral characteristics of hypernasality was based on 37 speech samples. Thirty-two samples were from the group with cleft palate. In an effort to maintain the number of samples at a manageable level, five speech samples randomly selected from the control group were used.

3. Acoustic analysis

The digitized speech samples were filtered using 16 1/3-octave digital bandpass filters with center frequencies from 250 Hz to 8 kHz. Chebyshev-type infinite impulse response filters with a ripple of ± 0.2 dB were used (CATCH system, Coretex Corp., Tokyo). The rms (root-mean-square) of each filter output was calculated every 1.44 ms and was averaged for a steady-state portion of the vowel to obtain an averaged 1/3-octave spectrum.

Since the speech samples were produced using natural loudness and pitch, the intensity levels were variable. Thus, to make comparisons possible, all spectra were normalized to have the average level of these 16 1/3-octave bands equal to 0 dB (Suomi, 1984).

4. Perceptual experiment

Three speech-language pathologists and one dentist, each of whom had at least 15 years experience in the assessment of speech associated with cleft palate, served as listeners. These listeners were considered to have better skills in isolating hypernasality from other speech characteristics than naive or less experienced listeners. All of the listeners reported that rating the isolated vowel was more difficult than rating connected or conversational speech because the number of cues was limited. When they rated connected or conversational speech, however, other speech characteristics associated with velopharyngeal inadequacy could influence the evaluation of hypernasality.

The listeners were instructed to use a 6-point equal-appearing interval scale to rate the severity of hypernasality with 1 representing ''no nasal resonance,'' 2, 3, 4, 5, and 6 representing ''mild,'' ''mild–moderate,'' ''moderate,'' ''moderate–severe,'' and ''severe hypernasal resonance,'' respectively. Written instructions were provided and reviewed orally at the beginning of the task. First, listeners were asked to listen to all the speech samples to acquaint themselves with the range of samples with respect to the severity of hypernasality. Then, they were asked to rate the 37 speech samples. All listeners had practice sessions before judging sessions. Three of them (listeners A, B, and C) had

three extra-practice sessions using the same set of the samples. The listeners were seated individually in a sound-treated room at a multimedia computer equipped with headphones (Sony MDR-V600).

A control program (SARS-RATEWAV, Microtronics Corp., Chapel Hill, NC) was used to present the speech samples binaurally by the headphones at a constant sound level (approximately 80 dB A-weighted SPL) through a 16-bit D/A converter. The program also recorded listener's responses. The presentation of samples was self-paced, and each speech sample could be repeated (at the discretion of the listener) a maximum of nine times before a decision had to be made. This session was then repeated three times for each listener. The control program randomized the order of sample presentation.

B. Results

In analyzing the rated scores of 37 speech samples obtained from four experienced listeners, intralistener reliability and interlistener reliability were calculated. Intralistener reliability was calculated for each listener using intraclass correlation coefficients for the scores of the three ratings. Intralistener reliability ranged from 0.67 to 0.81 with an average of 0.75. Average rating scores were obtained from the three sessions for each listener. The interlistener reliability, calculated using an intraclass correlation coefficient, was 0.68.

Each listener's average and standard deviation of the 37 rating scores were different from each other (Table I). In order to eliminate the individual difference in the dispersion of the rated scores, the averaged rating scores were normalized so as to have each listener's average and standard deviation equal to 3 and 1.25, respectively. Then, the mean rating scores of the four listeners for each sample were calculated.

Based upon the averaged normalized rating scores, the 37 samples were divided into three groups: normal resonance group ($n=7$) (average rating less than 2), mild to moderate hypernasal resonance group ($n=24$) (average rating equal to or greater than 3 and less than 5), moderate to severe hypernasal resonance group ($n=6$) (average rating equal to or greater than 5). Average 1/3-octave spectra were calculated for each of these three groups and for the control group, except for the five speakers whose speech samples were used for the perceptual experiment ($n=27$).

In Fig. 1, the average 1/3-octave spectra from the moderate to severe and the mild to moderate hypernasal resonance groups are compared to the spectra from the normal resonance group and the control group. Obvious differences were observed in the level of $F1$, the levels between $F1$ and $F2$, and in the levels of $F2$, $F3$ region. In the hypernasal groups, the level of $F1$ and the levels between $F1$ and $F2$ were higher and the levels in $F2$, $F3$ regions were lower than those of the normal resonance and the control groups.

An analysis of variance (ANOVA) for differences in mean level [dB] of each 1/3-octave band among the four groups showed significant differences in the levels of the bands whose central frequency were 630 Hz, 1, 1.6, 2, and 3.2 kHz. The obtained F values are shown in Table II. In the level of $F1$ (630 Hz), the moderate to severe and the mild to moderate hypernasal groups had a higher amplitude than the normal resonance and the control groups. In the levels of 1, 1.6, and 2 kHz, the moderate to severe hypernasal group had the highest amplitude and the normal resonance group had the lowest amplitude. The other significant differences were observed in the level of $F2$ region (3.2 kHz). In this band, the normal resonance group had the highest amplitude and the moderate to severe group had the lowest amplitude (Fig. 1).

Based on these results, the spectral characteristics of hypernasality of the vowel [i] in children with cleft palate were identified as an increase in the level of $F1$ and an increase in the levels between $F1$ and $F2$ and a decrease in the levels of $F2$, $F3$ region.

TABLE I. Four listeners' average rating scores of 37 speech samples without normalization.

Sample	Listener A	B	C	D
1	4.8	3.7	4.8	5.7
2	4.0	4.2	4.8	6.0
3	1.2	1.5	1.3	3.7
4	1.7	2.7	3.7	2.7
5	1.8	3.0	2.5	4.7
6	3.0	2.3	3.2	5.3
7	1.3	1.3	3.2	4.3
8	2.8	3.0	2.3	3.0
9	1.3	3.2	3.2	5.3
10	1.3	2.0	1.2	3.0
11	1.3	3.2	1.8	3.7
12	1.8	2.0	1.8	4.0
13	2.0	3.0	2.5	3.3
14	1.0	1.8	1.3	3.3
15	2.3	3.0	4.2	5.3
16	2.5	2.7	3.7	5.0
17	3.2	2.0	4.0	5.0
18	1.7	2.5	4.5	5.3
19	3.5	4.5	4.7	3.3
20	2.8	3.2	4.3	6.0
21	2.2	3.5	3.2	3.7
22	4.0	3.5	4.8	5.3
23	4.3	4.3	3.7	6.0
24	1.7	2.7	1.5	3.0
25	1.7	2.7	4.5	4.3
26	3.3	2.8	2.8	6.0
27	2.2	3.2	3.8	5.3
28	2.8	3.0	3.5	4.7
29	5.5	5.8	5.5	6.0
30	2.0	2.0	3.0	4.7
31	4.0	3.7	4.2	5.7
32	3.0	2.7	3.8	4.7
33	1.8	3.0	2.2	4.0
34	1.0	1.3	1.0	3.0
35	1.0	1.5	1.0	1.3
36	1.0	1.0	1.0	1.0
37	1.0	2.0	1.5	3.7
average	2.4	2.8	3.1	4.3
s.d.	1.2	1.0	1.3	1.3

III. HYPERNASALITY IN SPEECH SAMPLES WITH MANIPULATED SPECTRAL CHARACTERISTICS

The relationship between spectral characteristics and perceived hypernasality was examined using speech samples with manipulated 1/3-octave spectral characteristics.

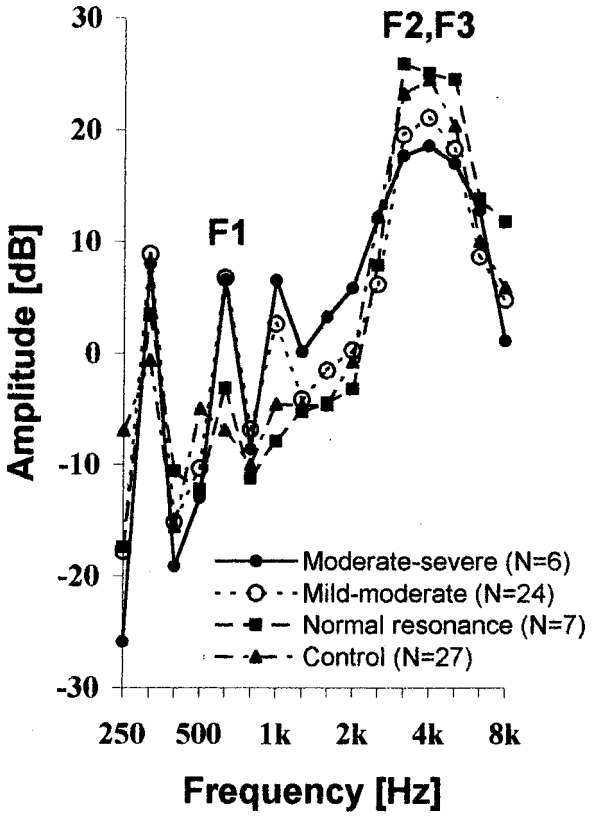


FIG. 1. Average 1/3-octave spectra from the four groups of subjects.

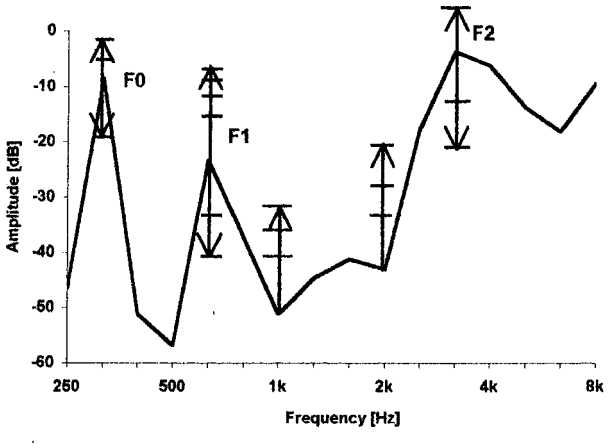


FIG. 2. A 1/3-octave spectrum of an original speech sample produced by a 13-year-old girl from the control group and a method of manipulating the relative amplitude of 1/3-octave bands which contain spectral components. The vertical lines represent different amplitude modifications.

A. Method

1. Stimuli

An original speech sample that did not have obvious spectral changes over the duration of the vowel was used. This speech sample was produced by a subject without cleft palate (female, age 13) and received a rating of 1 (no nasal resonance) from all four experienced listeners in the perceptual experiment of Sec. II.

Spectral properties of the speech sample were changed. A number of possible hypernasal speech samples of the vowel were created by manipulating the relative amplitudes of the 1/3-octave bands that were specific to the spectral characteristics of hypernasality identified in Sec. II. The manipulated spectral components were $F0$, $F1$, the spectral components between $F1$ and $F2$, and the $F2$, $F3$ region. The amplitude manipulation was performed for each spectral component and all possible combinations of these spectral components to create as many hypernasal speech samples as possible. The manipulated samples had the same duration (1 s), the same fundamental frequency (290 Hz), and the same sound-pressure level (approximately 80 dB A-weighted SPL).

Figure 2 shows a 1/3-octave spectrum of the original sample and the method of amplitude manipulation. The modifications of spectral characteristics were specified in terms of differences in the amplitudes of particular spectral components compared to the original sample. Forty-nine samples were created in this manner.

As a preliminary test, listeners A, B, and C rated these 49 samples at least twice in terms of both acceptability as the vowel [i] and the severity of hypernasality. If two listeners judged a sample as unacceptable as the vowel [i], then the sample was discarded. Fourteen samples were eliminated in this manner and the remaining 35 samples and the original sample were used for the perceptual experiment (Table III). For the purpose of familiarizing the listener with the manipulated speech samples, listener D, who did not participate in the preliminary test, rated these 49 samples before the judging sessions. Among the 35 samples, the amplitude(s) of only one spectral component was manipulated in 18 samples (sample 1–18). Manipulation of each spectral component was performed by modifying the amplitude of only one 1/3-octave band except the spectral component of $F2$, $F3$ region. The amplitudes of two bands (3.2, 4 kHz) were modified for manipulating the $F2$, $F3$ region. To examine the influence of each spectral component on perceived hypernasality, the 18 samples and the original sample were used.

TABLE II. Results of analysis of variance (ANOVA) for differences in mean level [dB] of each 1/3-octave band among the four groups. Band=the center frequency of the 1/3-octave band.

Band [Hz]	250	320	400	500	630	800	1k	1.25k	1.6k	2k	2.5k	3.2k	4k	5k	6.3k
$F(3,60)$	ns	ns	ns	ns	3.32[a]	ns	5.29[b]	ns	4.66[b]	3.71[a]	ns	3.75[a]	ns	ns	ns

[a] $p<0.05$.
[b] $p<0.01$.

TABLE III. 35 speech samples with manipulated 1/3-octave spectral characteristics used in Sec. III. The numbers indicate difference in the amplitude (dB) of particular spectral components (1/3-octave band) compared to the original sample. The minus sign represents a decrease in amplitude.

	Spectral component															
Sample	*F*0				*F*1				*F*2, *F*3 region							
Band (Hz)	250	320	400	500	630	800	1k	1.25k	1.6k	2k	2.5k	3.2k	4k	5k	6.3k	8k
1		5														
2		8														
3		−10														
4					7											
5					12											
6					15											
7					18											
8					−10											
9					−18											
10							10									
11							15									
12							20									
13									10							
14									15							
15									20							
16												7	7			
17												−10	−10			
18												−20	−20			
19										−5	−12	−20	−8			
20	13	8														
21		5	10	5												
22			8	15	10											
23			10	15	−15	−6										
24					−23	10	10									
25				10	28											
26							5	9	9							
27						10	11	15	14	5						
28						5	6	10	14	14						
29						7	8	14	14	15						
30						11	15	17	18	19						
31						14	16	23	16	12						
32					7	9	9	13	22	8	−5					
33				5	7				16	15	5					
34						10	11	8	5			7				
35						12	12	15	14	10	5	6	7			

2. Acoustic analysis

The 36 speech samples including the original sample were analyzed and normalized in the same manner as Sec. II. The levels of 16 1/3-octave bands (250 Hz to 8 kHz) were used as the acoustic parameters.

3. Perceptual experiment

The same four experienced listeners who participated in Sec. II also participated in Sec. III. They again used a 6-point equal-appearing interval scale to rate the severity of hypernasality. Written instructions were provided and reviewed orally at the beginning of the task. The same control program (SARS-RATEWAV) was used to present the speech samples. In this study, the original sample was used as a reference sample to represent category 1 (no nasal resonance). The listeners were permitted to review the reference sample at any time during the session and as many times as they wished.

First, for the purpose of familiarizing the listeners with the manipulated speech samples, a practice session consisting of six manipulated speech samples was performed. The practice session was repeated three times. These practice samples were not used during the judging session. After the practice session, a judging session was conducted consisting of the 36 speech samples. The presentation of samples was self-paced, and each speech sample could be repeated (at the discretion of the listener) a maximum of 20 times before making a final decision. The judging session was repeated twice for each listener. The control program randomized the order of sample presentation.

4. A relationship between spectral characteristics and perceptual evaluation

Multiple regression analysis was used to examine the relationship between the acoustic parameters and the averaged perceptual ratings of hypernasality obtained from the four experienced listeners.

TABLE IV. Four listeners' rating scores of 36 manipulated speech samples without the normalization.

Sample	Listener A	B	C	D
1	1.5	1.0	1.0	2.0
2	2.5	2.0	1.5	2.0
3	2.5	2.5	2.5	1.5
4	1.0	1.5	1.5	1.0
5	1.5	2.0	3.5	2.0
6	2.0	4.0	3.5	2.0
7	2.5	3.5	6.0	3.5
8	1.0	1.0	1.0	1.5
9	1.0	1.0	1.0	2.0
10	2.0	3.0	5.0	2.5
11	3.0	4.5	5.0	3.5
12	3.0	4.0	5.0	5.5
13	1.0	2.0	1.0	2.0
14	2.0	2.5	4.0	3.0
15	3.0	3.5	6.0	5.5
16	1.0	2.0	1.5	1.0
17	2.5	2.0	5.5	2.0
18	4.5	3.5	6.0	5.0
19	3.0	3.5	5.0	3.0
20	2.0	2.5	4.5	3.5
21	1.0	2.5	1.5	1.5
22	2.0	4.5	5.5	2.0
23	1.0	2.0	1.0	2.0
24	2.5	2.5	3.0	5.0
25	2.0	3.5	5.5	3.5
26	1.0	2.5	2.5	3.0
27	2.0	4.5	4.5	3.0
28	2.0	3.5	2.5	3.5
29	2.5	4.0	3.5	3.5
30	3.5	5.0	5.0	6.0
31	4.0	4.5	5.5	5.5
32	5.0	5.0	5.5	5.0
33	2.0	2.5	3.0	2.0
34	2.5	4.5	4.0	4.5
35	2.5	6.0	4.5	5.5
36	1.0	1.0	1.0	1.0
average	2.2	3.0	3.6	3.1
s.d.	1.0	1.3	1.8	1.5

B. Results

1. Perceptual experiment

In analyzing the rated scores of 36 manipulated speech samples obtained from four experienced listeners, intralistener reliability and interlistener reliability were calculated. Intralistener reliability was calculated for each listener using Pearson's r for the scores of the two ratings. Intralistener reliability ranged from 0.71 to 0.88 with the average of 0.77. The average rating scores were obtained from the two sessions for each listener. Interlistener reliability, calculated using intraclass correlation coefficient, was 0.71.

Average rating scores were obtained for each listener. Each listener's average and standard deviation of the 36 rating scores were different from each other (Table IV). To eliminate the individual differences in the variation of the rated scores, the averaged rating scores were normalized so as to have each listener's average and standard deviation equal to 3 and 1.25, respectively. The mean rating scores of the four listeners for each sample were then calculated.

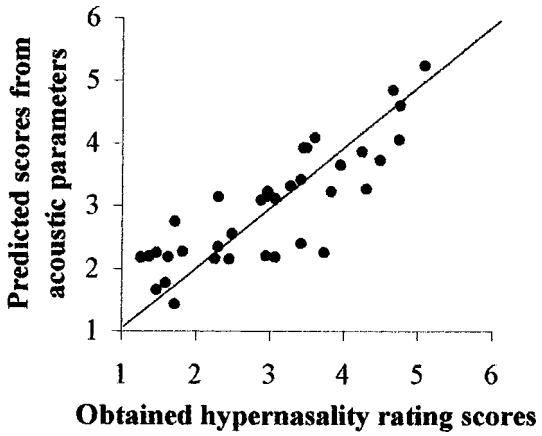


FIG. 3. Correlation between the perceptual rating scores and the predicted hypernasality rating scores obtained from the multiple regression equation using the acoustic parameters.

2. The relationship between spectral characteristics and perceived hypernasality

The predicted scores (y) obtained from the multiple regression equation (1) using the amplitudes of the three 1/3-octave bands (1, 1.6, and 2.5 kHz) had a high correlation ($R=0.84$) with the perceptual rating scores (Fig. 3). This multiple regression coefficient was significant, $F(3,32)=25.33$, $p<0.001$. The multiple regression equation was

$$y=7.42+0.11\times\mathrm{L1k}+0.06\times\mathrm{L1.6k}-0.17\times\mathrm{L2.5k}. \quad (1)$$

In this equation, L1k=level of 1 kHz, L1.6k=level of 1.6 kHz, and L2.5k=level of 2.5 kHz. An increase in the amplitude of the 1/3-octave band of 1 or 1.6 kHz and a decrease in the amplitude of the band of 2.5 kHz increased the predicted score (y). That is, the predicted scores from the acoustic parameters become more hypernasal. The standardized regression coefficients for the level of 1 kHz were almost identical among the listeners. The coefficient for the level of 2.5 kHz, however, was considerably different among the listeners (Table V). When the listeners rated the severity of hypernasality, they placed nearly the same weight on the levels of 1 and 1.6 kHz. They placed a considerably different weight, however, on the level of 2.5 kHz.

3. The influence of each spectral component on perceived hypernasality

To examine the influence of each spectral component on (a) perceived hypernasality, and (b) the individual difference of the perceptual ratings among the listeners, the 18 samples were divided into four groups based on the manipulated

TABLE V. Standardized regression coefficients for each acoustic parameter which were calculated using the average rating score of the four listeners and the rating score of each listener (A, B, C, and D).

Acoustic parameter	Listener 4 listeners	A	B	C	D
Level of 1 kHz (L1k)	0.59	0.47	0.53	0.55	0.54
Level of 1.6 kHz (L1.6k)	0.39	0.45	0.31	0.24	0.40
Level of 2.5 kHz (L2.5k)	−0.42	−0.59	−0.23	−0.38	−0.30

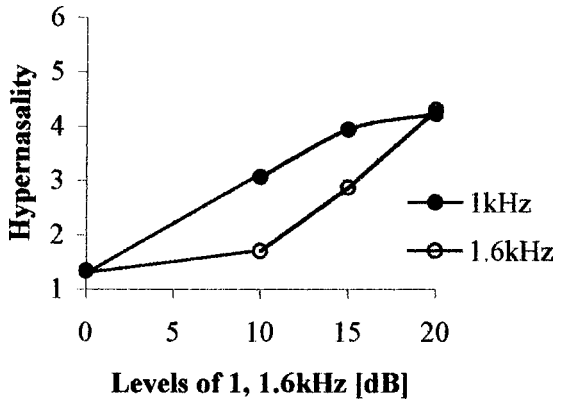


FIG. 4. The relationships between the levels of the bands between $F1$ and $F2$ (1 kHz, 1.6 kHz) and the perceptual ratings of hypernasality. Level of 1, 1.6 kHz=the level of the 1/3-octave band whose center frequency is 1 kHz, 1.6 kHz. Hypernasality=perceptual ratings of the severity of hypernasality.

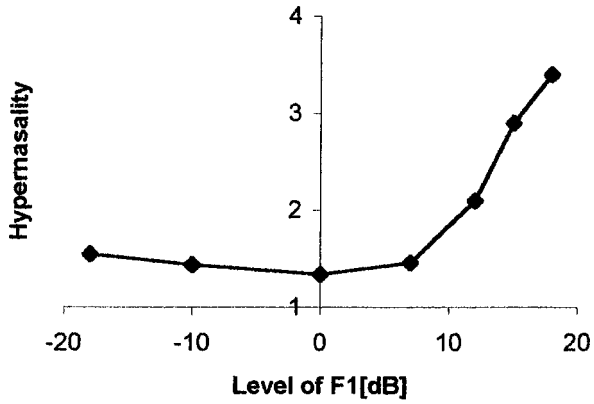


FIG. 6. The relationship between the level of the 1/3-octave band which contains $F1$ and the perceptual ratings of hypernasality. Level of $F1$=the level of the 1/3-octave band which contains $F1$. Hypernasality=perceptual ratings of the severity of hypernasality.

spectral component. The influence of each spectral component on perceived hypernasality was examined. As the amplitude of the band between $F1$ and $F2$ (1 and 1.6 kHz) increased, the hypernasality ratings increased (Fig. 4). If the amplitude of the band of 1 kHz was increased by more than 15 dB, or the amplitude of the band of 1.6 kHz was increased by less than 10 dB, no obvious change was observed in ratings. Decreasing the amplitude of $F2$, $F3$ region increased the hypernasality ratings (Fig. 5). As the amplitude of $F1$ increased, the hypernasality ratings increased. When the amplitude of $F1$ decreased, however, no obvious change was observed (Fig. 6). Modifying the amplitude of $F0$ in either direction increased the ratings of hypernasality (Fig. 7).

IV. DISCUSSION

A. Spectral characteristics of hypernasality

The results of multiple regression analysis showed that the severity of hypernasality can be evaluated by using the two levels of 1/3-octave bands between $F1$ and $F2$ and the level of $F2$. These results were in agreement with the spectral characteristics of hypernasality in children with cleft palate identified in Sec. II. These results were also in agreement with the reported spectral characteristics of nasalization of the vowel [i] such as an appearance of additional spectral peak(s) between $F1$ and $F2$ and a decrease in the amplitude of $F2$ and $F3$ (House and Stevens, 1956; Hattori, Yamamoto, and Fujimura, 1958; Fant, 1970). Thus, this acoustic method for quantitative evaluation of hypernasality using the three levels of 1/3-octave bands is compatible with the other reports.

Decreased amplitude of $F1$ was also reported as a spectral characteristic of nasalization (House and Stevens, 1956; Fant, 1970). The spectral characteristics of nasalization, however, differ from one vowel to another (House and Stevens, 1956; Hattori *et al.*, 1958; Fant, 1970; Hawkins and Stevens, 1985; Maeda, 1993; Beddor, 1993). From theoretical considerations, coupling of the nasal tract to the main vocal tract introduces pole–zero pairs in the transfer function. When low vowels such as [a] and [æ] are nasalized, the amplitude of $F1$ decreases because the first nasal zero appears in the frequency region of $F1$. When high vowels such as [i] and [u] are nasalized, however, the first nasal zero appears in a higher frequency region than $F1$; therefore, the amplitude of $F1$ is not attenuated.

Hypernasal vowels, in general, have broadened peaks

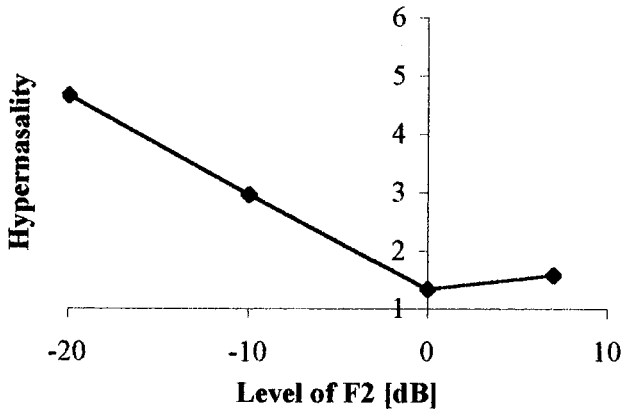


FIG. 5. The relationship between the levels of the 1/3-octave band which contains $F2$, $F3$ region and the perceptual ratings of hypernasality. Level of $F2$=the level of the 1/3-octave band containing $F2$, $F3$ region (3.2, 4 kHz). Hypernasality=perceptual ratings of the severity of hypernasality.

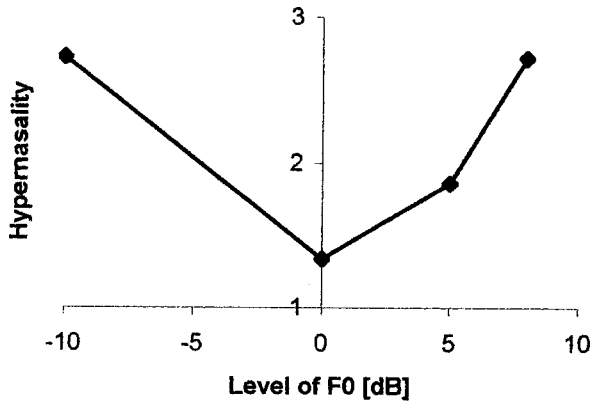


FIG. 7. The relationship between the level of the 1/3-octave band which contains $F0$ and the perceptual ratings of hypernasality. Level of $F0$=the level of the 1/3-octave band which contains $F0$. Hypernasality=perceptual ratings of the severity of hypernasality.

and flattened spectra. When the spectral peaks (formants) are not prominent, the shape of the entire region of the spectral envelope is important for vowel perception rather than the frequency and the amplitude of the spectral peaks (Beddor and Hawkins, 1990). Therefore, 1/3-octave spectral analysis that evaluates the overall spectral envelope may have a theoretical advantage in analyzing hypernasal vowels. Furthermore, the static properties of the vowel spectra have been examined by formant analysis, whereas 1/3-octave analysis can utilize both static and dynamic approaches.

An increase in the amplitude between $F1$ and $F2$ consistently increased the ratings of hypernasality with relatively similar ratings among the listeners. When only the amplitudes of 1 and 1.6 kHz were used as the acoustic parameters, multiple regression analysis showed a relatively high correlation ($R=0.74$) between the perceptual rating scores and the predicted scores. An increase in the amplitude between $F1$ and $F2$ is considered to have the greatest influence on perceived hypernasality.

The amplitudes of 1 kHz had a different influence on perceived hypernasality than the amplitude of 1.6 kHz (Fig. 4). Studies using an articulatory synthesizer have shown that the distance in frequency between $F1$ and the spectral peak between $F1$ and $F2$ increased according to the size of velopharyngeal opening (Maeda, 1993; Stevens, 1998). A similar result was obtained from an adult with cleft palate using a specially designed prosthetic appliance (Watterson and Emanuel, 1981). It is suggested that the location of the spectral peak in the frequency axis, such as 1 and 1.6 kHz, could be related to the size of velopharyngeal opening as well as the size and shape of the nasal tract.

Furthermore, based on the comparison between 1/3-octave analysis and formant analysis in children's speech, Bakkum *et al.* (1995) reported that formant analysis was not suitable for real-time application. Because 1/3-octave analysis can be performed in real time, it is more suitable for a clinical setting. This method is noninvasive, reproducible, does not interfere with phonation, and can be performed in real time. Even very young children can produce the test speech sample of the vowel [i]. This acoustic method, therefore, could be considered suitable for use in a clinical setting.

B. Reliable perceptual evaluation of hypernasality

In Sec. II, listeners A, B, and C rated the 37 speech samples six times. Intralistener reliability was calculated using intraclass correlation coefficient for the scores of the first three ratings and the last three ratings (Listener A 0.82, 0.81; Listener B 0.62, 0.67; Listener C 0.84, 0.77). No obvious difference was observed in the first three and the last three ratings in terms of intralistener reliability. Then average ratings of the first three ratings and the last three ratings for each listener were compared using Pearson r correlation coefficients (Listener A 0.92, Listener B 0.72, Listener C 0.87). There was no obvious improvement in the reliability of rating after the experienced listeners were familiar with the rating task.

Kataoka *et al.* (2001) examined how the acoustic characteristics of the speech samples influenced the variability of the perceptual ratings among listeners. The authors reported that voice quality distortions and a particular type of spectral change that related to the severity of hypernasality over the duration of the vowel could contribute to the variability of perceptual ratings in the isolated vowel [i].

A greater degree of increase in intra- and interlistener reliability was expected in the perceptual experiment using the manipulated speech samples. Contrary to our expectation, inter- and intralistener reliability increased slightly. As shown in Table V, each listener placed a different weight on the level of 2.5 kHz when he or she evaluated the severity of hypernasality.

The change of the amplitude of $F2$, $F3$ region shifts the overall spectral balance of a high-frequency region in relation to a low-frequency region. This could change the perception in voice quality that listeners might associate with the degree of nasal resonance when spectral peaks are not prominent (Hawkins and Stevens, 1985). The change in voice quality as well as in the severity of hypernasality could confuse listeners and result in a large individual difference in ratings of the vowel [i].

Reliable perceptual evaluation of hypernasality is considered to have a stronger relationship to the acoustic characteristics of hypernasality in the same speech sample of the vowel [i]. The predicted hypernasal scores obtained from the multiple regression equation using the acoustic parameters of hypernasality should correlate closely to the reliable perceptual evaluation. Stated in other terms, we believe that more reliable perceptual evaluation could be obtained by examining the relationship between the perceptual evaluation and the acoustic characteristics of the same speech signal controlling the acoustic parameters that may cause the perceptual variability among the listeners.

ACKNOWLEDGMENTS

This research was supported by NIDR Grants Nos. DE07105 and DE06957. The authors would like to thank Kaoru Okabe, who spent many hours developing programs and provided advice throughout the study. The authors would also like to thank Betty J. Philips for her participation in the perceptual experiments and for making a number of helpful suggestions. The authors would also like to thank to Anders Löfqvist and two anonymous reviewers for their comments and suggestions on the manuscript.

Bakkum, M. J., Plomp, R., and Pols, L. C. W. (**1995**). ''Objective analysis versus subjective assessment of vowels pronounced by deaf and normal-hearing children,'' J. Acoust. Soc. Am. **98**, 745–762.

Beddor, P. S., and Hawkins, S. (**1990**). ''The influence of spectral prominence on perceived vowel quality,'' J. Acoust. Soc. Am. **87**, 2684–2704.

Beddor, P. S. (**1993**). ''The perception of nasal vowels,'' in *Phonetics and Phonology, Nasals; Nasalization, and the Velum*, edited by M. K. Huffman and R. A. Krakow (Academic, San Diego), Vol. 5, pp. 171–196.

Chen, M. Y. (**1995**). ''Acoustic parameters of nasalized vowels in hearing-impaired and normal-hearing speakers,'' J. Acoust. Soc. Am. **98**, 2443–2453.

Chen, M. Y. (**1997**). ''Acoustic correlates of English and French nasalized vowels,'' J. Acoust. Soc. Am. **102**, 2360–2370.

Fant, G. (**1970**). *Acoustic Theory of Speech Production*, 2nd ed. (Mouton, The Hague, The Netherlands).

Hattori, S., Yamamoto, K., and Fujimura, O. (**1958**). ''Nasalization of vowels in relation to nasals,'' J. Acoust. Soc. Am. **30**, 267–274.

Hawkins, S., and Stevens, K. N. (**1985**). ''Acoustic and perceptual correlates of the non-nasal–nasal distinction for vowels,'' J. Acoust. Soc. Am. **77**, 1560–1575.

House, A. S., and Stevens, K. N. (**1956**). ''Analog studies of the nasalization of vowels,'' J. Speech Hear. Disord. **21**, 218–232.

Kataoka, R. (**1988**). ''Quantitative evaluation of hypernasality.—relation between spectral characteristics and perception of hypernasality,'' J. Jpn. Cleft Palate Assoc. **13**, 204–216.

Kataoka, R., Michi, K., Okabe, K., Miura, T., and Yoshida, H. (**1996**). ''Spectral properties and quantitative evaluation of hypernasality in vowels,'' Cleft Palate-Craniofacial J. **33**, 43–50.

Kataoka, R., Zajac, D. J., Mayo, R., Lutz, R. W., and Warren, D. W. (**2001**). ''The influence of acoustic and perceptual factors on perceived hypernasality in the vowel [i]: A preliminary study,'' Folia Phoniatr. (in press).

Maeda, S. (**1993**). ''Acoustics of vowel nasalization and articulatory shifts in French nasal vowels,'' in *Phonetics and Phonology: Nasals, Nasalization, and the Velum*, edited by M. K. Huffman and R. A. Krakow (Academic, San Diego), Vol. 5, pp. 147–167.

Pols, L. C. W., Tromp, H. R. C., and Plomp, R. (**1969**). ''Perceptual and physical space of vowel sounds,'' J. Acoust. Soc. Am. **46**, 458–467.

Stevens, K. N. (**1998**). *Acoustic Phonetics* (The MIT Press, Cambridge, MA).

Suomi, K. (**1984**). ''On talker and phoneme information conveyed by vowels: A whole spectrum approach to the normalization problem,'' Speech Commun. **3**, 199–209.

Watterson, T., and Emanuel, F. (**1981**). ''Observed effects of velopharyngeal orifice size on vowel identification and vowel nasality,'' Cleft Palate J. **18**, 271–278.

A cross-language comparison of /d/–/ð/ perception: Evidence for a new developmental pattern

Linda Polka,[a] Connie Colantonio, and Megha Sundara
School of Communication Sciences and Disorders, McGill University, 1266 Pine Avenue West, Montreal, Quebec H3G 1A8, Canada

(Received 16 August 2000; accepted for publication 2 February 2001)

Previous studies have shown that infants perceptually differentiate certain non-native contrasts at 6–8 months but not at 10–12 months of age, whereas differentiation is evident at both ages in infants for whom the test contrasts are native. These findings reveal a language-specific bias to be emerging during the first year of life. A developmental decline is not observed for all non-native contrasts, but it has been consistently reported for every contrast in which language effects are observed in adults. In the present study differentiation of English /d–ð/ by English- and French-speaking adults and English- and French-learning infants at two ages (6–8 and 10–12 months) was compared using the conditioned headturn procedure. Two findings emerged. First, perceptual differentiation was unaffected by language experience in the first year of life, despite robust evidence of language effects in adulthood. Second, language experience had a facilitative effect on performance after 12 months, whereas performance remained unchanged in the absence of specific language experience. These data are clearly inconsistent with previous studies as well as predictions based on a conceptual framework proposed by Burnham [Appl. Psycholing. **7**, 201–240 (1986)]. Factors contributing to these developmental patterns include the acoustic properties of /d–ð/, the phonotactic uniqueness of English /ð/, and the influence of lexical knowledge on phonetic processing. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1362689]

PACS numbers: 43.71.Ft, 43.71.An, 43.71.Hw [KRK]

I. INTRODUCTION

One of the major accomplishments of early childhood is learning to understand one or more languages. As a first step, the infant faces the challenge of learning to recognize the meaningful patterns in the speech stream, including the segmental units that distinguish words in their native language. We now know that infants come well equipped for this task. Numerous studies show that infants as young as 1 month of age are already able to differentiate[1] a wide range of phonetic contrasts including native contrasts and non-native contrasts that are difficult for adults to distinguish (e.g., Eimas *et al.*, 1971; for a review see Best, 1994a). Research has also firmly established that these initial perceptual abilities appear to shift from an unconstrained, language-general form towards a more restricted, language-specific pattern by the end of the first year of life. This conclusion is supported by evidence that infants can differentiate certain non-native contrasts at 6–8 months of age but not at 10–12 months, whereas a decline is not observed in infants for whom the test contrasts are native. This finding was first reported by Werker and colleagues. They investigated the effects of age and language experience on cross-language phonetic perception in a series of studies using the conditioned headturn procedure (Werker *et al.*, 1981; Werker and Tees, 1983, 1984a). They tested English-learning infants at several ages on several non-native contrasts including the Hindi retroflex versus dental stop contrast, /t̪–ʈ/, the Hindi voiceless aspirated versus breathy voiced contrast, /t^h–d^h/ and the Salish (Nthlakampx) glottalized velar versus uvular ejective stop contrast, /k′–q′/. The results clearly showed that English-learning infants at 6–8 months of age were able to differentiate these non-native phonetic contrasts even though English adults could not, and that 10- to 12-month-old English infants exhibited a decline in perception of these non-native contrasts when compared to infants in the younger age group. They subsequently replicated this pattern using a longitudinal design and also tested three 11- to 12-month-olds from a Hindi-speaking family and three 11- to 12-month-olds from a Salish-speaking family. The results confirmed the language-specific nature of the decline in perceptual abilities between 6 and 12 months of age. More recently, the same pattern of decline has also been shown using a habituation/dishabituation procedure for English infants tested on the Salish contrast (Best *et al.*, 1995) and for Japanese infants tested on English /r–l/ (Tsushima *et al.*, 1994), a contrast which is notoriously difficult for Japanese adults.

A decline in perceptual differentiation in the first year has also been shown in English-learning infants for several non-native vowel contrasts, including Norwegian /y–ʉ/ (Best *et al.*, 1997) and the German /u–y/ and /U–Y/ (Polka and Werker, 1994). In the latter study, language effects emerged between 4 and 6 months of age, suggesting that language effects begin to emerge earlier in development for vowel contrasts than for consonant contrasts (see also Kuhl *et al.*, 1992). Best (1995; Best *et al.*, 1990) has also reported poor perceptual differentiation by English-learning infants at 10–12 months of age for several additional non-native con-

[a]Electronic mail: lpolka@po-box.mcgill.ca

sonant contrasts from Zulu, including a lateral fricative voicing contrast, a velar voiceless aspirated versus ejective stop contrast, and a plosive versus explosive bilabial stop contrast. For the Norwegian and the German vowel contrasts and for two of these Zulu contrasts, English adults' differentiation of the non-native contrasts was very good, indicating that infant perception may decline even when language effects are not evident in adults.

Several studies have reported there to be no evidence of a perceptual decline for certain other non-native contrasts in the first year of life. One study using a visual habituation procedure (Best *et al.*, 1988) showed that English adults and English infants between 6 and 14 months of age were able to differentiate the non-native Zulu apical/lateral click consonant contrast. These sounds are unlike any English phonetic category, and English adults did not even perceive them as speech. Best (1991; Best *et al.*, 1990) reported that both English adults and English 10- to 12-month-olds were able to differentiate the Ethiopian ejective /p′–t′/ contrast. Likewise, using the conditioned headturn procedure, Polka and Bohn (1996) failed to find any effects of age or language experience when English-learning and German-learning infants at both 6–8 and 10–12 months of age were tested on the English /æ–ɛ/ contrast, which is not phonemic in German, and the German /u–y/ contrast, which is not phonemic in English.[2] German and English infants showed good differentiation on both contrasts; adult perception of the non-native contrasts in these studies was also very good. These studies confirm that a simple loss-due-to-lack-of-exposure or lack-of-attention explanation is inadequate to account for developmental changes in phonetic perception.

Overall, the findings to date have been taken as evidence that for some phonetic contrasts language experience serves to maintain perceptual differentiation in the first year of life, which otherwise declines when relevant language experience is lacking. However, it is important to note that absence of language experience does not lead to a complete loss in perception. As mentioned above, infants have shown declines for non-native contrasts that non-native adults can easily differentiate. As well, training studies with adults have shown that differentiation of difficult and rarely heard non-native contrasts can be achieved with sufficient training (e.g., Jamieson and Morosan, 1986; Logan *et al.*, 1991; Morosan and Jamieson, 1989; Pisoni *et al.*, 1982; Tees and Werker, 1984) or when sensitive testing procedures are employed (e.g., Werker and Tees, 1984b). Several other findings also argue against an interpretation of developmental decline as a simple effect of exposure. Pegg and Werker (1997) found that perceptual differentiation also declines between 6–8 and 10–12 months for a contrast that has an allophonic status in the native language. This finding highlights the importance of functional status of the phoneme contrast rather than mere allophonic exposure to maintain perception. Lalonde and Werker (1995) have also shown that developmental decline in phonetic perception occurs in synchrony with age-related advances in general cognitive abilities such as visual categorization and object permanence. Taken together these findings argue for an interpretation of age-related changes in cross-language speech perception as a functional reorganization of phonetic perception that reflects advances in linguistic processing (Werker and Pegg, 1992; Werker *et al.*, 1995).

The finding that differentiation declines between 6 and 12 months for some non-native contrasts but not for others is clearly established in the literature, but is not yet well explained. The fact that decline is observed during infancy for non-native contrasts that are easily differentiated by adults reveals that, although infants are beginning to recognize some segmental aspects of their native language, their perceptual structures are still immature. It is not clear whether infants are simply able to recognize some phonetic categories and not others, or whether infants have formed any categories that are comparable to mature listeners. A better understanding of the facts that underlie divergent development patterns in cross-language phonetic perception can help address this issue.

Best (1995) has proposed that, for adults, language effects are evident for some contrasts but not others because there are differences in the way in which the contrasting non-native phones relate to the native phonology. She has developed a perceptual assimilation model (PAM) that is able to account for some differences in adult differentiation based on assimilation patterns (Best, 1990, 1993, 1994a, 1994b). Although it is difficult to assess assimilation patterns in infants, it may be possible to gain some understanding of how the native phonology is beginning to shape perception during infancy by comparing adult and infant differentiation for non-native contrasts that conform to different assimilation patterns in adults. To date, this line of investigation has shown that perceptual declines occur with non-native contrasts that adults assimilate equally well to the same native phonological category, referred to as a single category assimilation (Best *et al.*, 1995). In comparison, perceptual differentiation is maintained for controls that cannot be assimilated within the native phonological space, referred to as a nonassimilated (NA) pattern, such as the Zulu click contrast mentioned above (Best *et al.*, 1988). However, results are inconsistent for two other assimilation patterns. Different outcomes have been reported for contrasts in which each phone is a good match to a different native category, referred to as a two-category (TC) pattern. Specifically, Best found that English infants at 10–12 months were able to discriminate the Ethiopian /p′–t′/ contrast but not the Norwegian /y–ʉ/ contrast or the Zulu lateral fricative voicing contrast; all three of these contrasts conformed to a TC assimilation pattern for English adults. Likewise, results are also inconsistent for contrasts in which both phones are assimilated to a single native category but differ in goodness of fit, referred to as a category-goodness (CG) pattern. Specifically, Polka and Werker (1994) found a decline for two CG vowel contrasts, whereas Polka and Bohn (1996) found that differentiation was maintained for two other CG vowel contrasts. Best also found poor performance at 10–12 months for the Zulu voiceless aspirated versus ejective stop contrast; English adults assimilated this contrast as a CG pattern (Best, 1995). Clearly, more research involving a wider range of non-native contrasts is needed to extend and refine PAM to explain phonetic perception in infancy.

Burnham (1986) has provided a more specific prediction

regarding developmental patterns for different phoneme contrasts which is based on their relative position on a ''robust–fragile'' continuum. This continuum is defined with respect to acoustic salience and frequency of occurrence across languages, such that robust contrasts are high in acoustic salience and occur frequently in the world's languages and fragile contrasts are low in acoustic salience and occur rarely across languages. Although databases exist to establish the latter values, no metric is available to quantify acoustic salience. According to Burnham, perceptual reorganization occurs at two points in development. The first reorganization occurs early in development, within the first year, and affects perception of fragile contrasts; the second reorganization occurs later, around 5–6 years of age, and may affect differentiation of robust non-native contrasts.

Clearly, developmental data on additional non-native contrasts are needed to better understand the basis for divergent patterns in cross-language phonetic perception. The present study was designed to address this need by examining the effects of age and language experience on the perception of a stop–fricative contrast, /d–ð/, by French- and English-speaking adults, and French- and English-learning infants. This contrast is phonemic in English but not in French, which has a dental voiced stop, /d/, but lacks either a /ɵ/ or /ð/ phoneme.

There is abundant anecdotal evidence supporting the claim that French Canadians typically substitute /d/ for /ð/ and /t/ for /ɵ/ when they attempt to produce the English interdental fricatives. Several studies also indicate that this production error is due, at least in part, to problems in perceptual differentiation of these phones (Jamieson and Morosan, 1986, 1989; Morosan and Jamieson, 1989; Werker *et al.*, 1992). These researchers report that French Canadian adults who are learning English perceive English /ð/ as an instance of French /d/ and English /ɵ/ as an instance of French /t/, and thus describing a single category assimilation pattern for English /d–ð/. To date, Morosan and Jamieson (1989) have reported the most detailed assessment of Francophone perception of this stop–fricative contrast. They examined the perception of /ð/ versus /d/ by Francophones in the context of a study designed to examine the effects of training on perception of English /ɵ–ð/ by Quebec French-speaking adults. They used both natural /d/ and /ð/ speech tokens and a synthesized /ð/ to /d/ continuum. Their results indicate that the Francophones had difficulty identifying the natural tokens correctly. In particular, Francophones had problems identifying the /ð/ items with the shortest frication durations, which suggests that they were relying only on the frication duration due to differentiate this contrast. Their identification of the /d–ð/ continuum also failed to show a clear categorization of the series.

In the present study, we examined perceptual differentiation of the English /d–ð/ contrast by English-speaking and Canadian French-speaking adults and English-learning and Canadian French-learning infants at 6–8 and 10–12 months of age using the conditioned headturn procedure. We expect to replicate the effects of age and language experience first reported by Werker and colleagues. Specifically, we expect performance to be poorer for French compared to English subjects for the adults and 10- to 12-month-olds but no difference to be evident for 6- to 8-month-olds. Thus, with respect to developmental patterns, we expected to find evidence of maintenance (i.e., no change with age) for the English group and developmental decline beginning at 10–12 months in the French group.

These results are predicted for three reasons. First, this outcome follows clearly from previous findings given that, in every study to date, when a language effect is evident in adult perception a decline in differentiation has been observed in the first year of life for infants from the non-native group. Second, as outlined above, previous adult studies indicate that Francophone adults assimilate English /d/ and /ð/ to a single category in their native language. Thus, perceptual decline during infancy is consistent with Best's PAM model. Third, English /d/ and /ð/ are short, low-amplitude phones, /ð/ is among the weakest sounds that occur in English (Fletcher, 1953; Ling and Ling, 1978), and the /d–ð/ contrast is also quite rare across languages of the world (Maddieson, 1984). According to Burnham, these facts define this contrast as fragile, and thus, perception of this contrast is expected to decline in the first year for non-native listeners.

II. METHODS

A. Stimuli

The stimuli consisted of two sets of naturally produced English minimal work pairs, /bot–vot/ (''boat''–''vote'') and /doz–ðoz/ (''doze''–''those''). The /b–v/ pair was selected as a control for task performance because it is present in both English and French and it contrasts the same manner classes and similar place differences as /d–ð/. Specifically, each contrast consists of a voiced stop versus a nonsibilant fricative with a small difference in place of articulation. It was considered important to use a comparable control contrast given previous findings suggesting that perception of fricatives may be more difficult for infants even in their native language (Eilers and Minifie, 1975; Eilers *et al.*, 1977; Holmberg *et al.*, 1977). Using this control, we could be confident that differences in performance with /d–ð/ can be attributed to effects of age or language experience over and above any effects of stimulus salience.

Multiple instances of boat, vote, doze, and those were recorded as produced in a short carrier phrase (''Number 1, those'') by a 26-year-old male monolingual native Canadian English speaker. These stimuli were digitized at a sampling rate of 10 kHz (12-bit resolution; low-pass filtered at 4.8 kHz) using a Macintosh IIfx computer with MACSPEECH software. Four /bot/, four /vot/, five /doz/, and five /ðoz/ tokens were selected so that the items within each contrast had matching values with respect to fundamental frequency, vowel amplitude, and five durational measures (prevoicing, total syllable, vowel, closure, and final burst/frication). this ensured that neither contrast could be differentiated by attending to extraneous acoustic differences that are not reliable cues to the phonetic contrast. In the final set, the /bot/ and /vot/ tokens had a mean duration of 325 ms (range=285.5–327.4 ms) and the /doz/ and /ðoz/ tokens had a mean duration of 443.4 ms (range=405.7–479.1 ms).

TABLE I. Acoustic cues distinguishing the initial consonants for the /bot–vot/ stimuli and /doz–ðoz/ stimuli. The * denotes a cue with completely nonoverlapping values; the + denotes a cue in which average values differ but there is some degree of overlap.

	Mean	Range	Mean	Range	Cue status	Difference between means
	/b/		/v/			
Noise duration (ms)	7.8	5.3–10.4	35.9	27.3–43	*	28.1
Noise amplitude (dB)	53.5	49.6–55.8	59	56.9–62.2	*	5.5
*F*2 at onset (Hz)	1172	1129–1210	1241	1210–1272	+	69
*F*2 change (Hz)	45.5	10–81	58.7	10–142		13.3
	/d/		/ð/			
Noise duration (ms)	16.2	7.7–20.4	18.6[a]	8.5–39.6[a]		2.4
Noise amplitude (dB)	53.7	50.1–56.7	56.2[a]	51.4–60[a]		2.5
*F*2 at onset (Hz)	1746	1699–1790	1536	1455–1699	+	210
*F*2 change (Hz)	604	550–681	456	336–631	+	148

[a]These values reflect only four of the five items because aperiodic noise energy could not be measured for one item.

The final stimulus set was redigitized to a 386 PC using BLISS (Mertus, 1990) and SWIFT software (sampling rate of 10 kHz, 12-bit resolution) to conduct adult and infant perceptual testing. Four monolingual English-speaking adults were asked to identify the initial consonant in each item and to comment on any unclear items. All four adults reported that all 18 items were clear, unambiguous examples of the intended words, confirming that the stimuli were high intelligible for native adults. These adult judgments were obtained in the same room and using the same setup, delivery system, and intensity level that was used in the infant and adult testing.

Acoustic analyses of the selected tokens were performed to identify the cues distinguishing each stop/nonsibilant fricative pair. As shown in Table I, these analyses included several measures of the initial aperiodic noise portion of each token as well as formant frequency measures. In this table, an asterisk (*) in the cue status column indicates that there were completely nonoverlapping distributions of the measure for the contrasting phones. These cues provide highly reliable information to distinguish the contrast. A plus (+) in this column denotes a measure for which there is a substantial difference in average values for the contrasting phones but some degree of overlap is present. Although these cues are less systematic they may also be useful in distinguishing the contrast. Measures with no asterisk or plus are similar across the contrasting phones and provided little or no information to help distinguish the contrast.

The duration of the initial aperiodic noise (i.e., stop burst or fricative energy) was measured by simultaneously referring to waveform and spectrographic displays. The average amplitude of the initial noise portion of the syllable was measured for the segment beginning at the onset of any prevoicing to the onset of periodicity of the vowel. For the /b–v/ contrast, the initial noise segment was shorter and less intense for /b/ than /v/ and there was no overlap of values for either measure. For /d–ð/, the duration and average amplitude of the initial noise segment differed only slightly between the /d/ and /ð/ items and there was considerable overlap in both values.[3]

Formant frequencies were measured for each item using MACSPEECH LAB II software via an LPC algorithm (13 poles) using the procedure described in Polka (1995). A 24-ms Hamming window was centered at successive 12.5-ms locations within each syllable, and measures were taken at the onset of voicing until voicing offset. Table I contains formant frequency measures corresponding to *F*2 at the onset of voicing and the *F*2 frequency change. The latter measure was defined as the extent of *F*2 frequency change between the onset and midpoint of the syllable, that is, during the initial *F*2 formant transition. The midpoint measure was taken at 50 ms from the onset for /bot/ and /vot/ tokens and at 137.5 ms from the onset for /doz/ and /ðoz/ tokens. These locations correspond to a point within each syllable which is close to where *F*2 reaches a frequency maximum for /bot/ and /vot/ tokens and to where *F*2 reaches a frequency minimum for /doz/ and /ðoz/ tokens.

As expected, given the place differences in these contrast, we found that on average the *F*2 onset frequencies were higher for /v/ than for /b/ and were considerably higher for /d/ than for /ð/ (Fant, 1970). For /b–v/, the mean *F*2 difference was small and there was one pair of /b/ and /v/ items with identical values. For /d–ð/, the mean difference was much larger than for /b–v/ and again there was one pair of /d/ and /ð/ stimuli with the same value. Given the manner differences in these contrasts, we expected to find evidence for more rapid *F*2 transitions for the stops relative to the fricatives (Pickett, 1999). Because our measures of *F*2 changes are computed over the same syllable durations, a greater *F*2 change indicates more rapid *F*2 change. For /b–v/ there was no clear difference or trend in the extent of *F*2 formant change. With respect to the /d–ð/ contrast, *F*2 change showed a clear tendency to be larger for /d/ than for /ð/, but there was some overlap in values for /d/ and /ð/ tokens. Thus, overall, for /bot/–/vot/, there are clear differences in aperiodic noise whereas *F*2 differences are present but less systematic. In comparison, for /doz/–/ðoz/, aperiodic noise cues are absent whereas *F*2 cues are prominent.

B. Subjects

Twenty-nine adults (19 females, 10 males) and 42 healthy, full-term infants (21 females, 21 males) served as subjects. The adult subjects included 15 monolingual French speakers (mean=24, range 19 to 35 years) and 14 monolingual English speakers (mean=21, range=18 to 28 years). Two additional adults, one French speaker and one English speaker, were excluded as they failed to meet the criterion (described below) with the /b–v/ control contrast. All adults reported having normal hearing and were paid for their participation. The English adults had resided in the province of Quebec for less than 4 years, and had only classroom instruction in French at the introductory level. None of the English adults listened to French radio or TV and none could speak a second language fluently. Francophone subjects were native speakers of Canadian French and had no more than basic high school English instruction. All Francophone adults resided in Quebec most of their lives. Some of the Francophone adults occasionally listened to English radio or TV but none was a fluent speaker of a second language.

The infant sample included 23 babies between 6 and 8 months of age: 10 French-learning (mean=7:15, range=6:02 to 8:16) and 13 English-learning infants (mean=7:04, range =6:00 to 8:10); and 19 babies between 10 and 13 months of age: 9 French-learning (mean=10:24, range 10:00 to 11:25) and 10 English-learning infants (mean=10:07, range=10:00 to 10:25). To get 42 infant subjects to complete the entire procedure, an additional 68 babies (32 English; 34 French) were tested but their data were excluded for the following reasons: failed to meet criterion with /b–v/ on day 1 (n =37);[4] fussed during testing ($n=7$); failed to meet retest criterion on day 2 [described below] ($n=9$); unable to return for second test session ($n=10$); technical problems or experimenter error ($n=5$). Parents reported that infants had normal hearing and none had been treated for an ear infection in the month prior to the study. Parents were provided with two small gifts as tokens of our appreciation.

Language experience of each infant was assessed via parent interview. For the English infants, English was the native language of both parents; however, for five infants one or both of the parents were bilingual.[5] However, in every case both parents had learned English very early in life, spoke it fluently, and preferred to speak English. The language used by the parents and directed to the baby was English; other relatives, caregivers, and visitors interacted with the baby in English. For each of these infants, the language in the home with respect to TV and radio was English.

For the French infants, Canadian French was the native tongue of both parents except for one infant, whose mother's native language was Romanian, a language which does not contain interdental fricatives. Some of the parents could also speak some English but all of them preferred to speak in French. Thus, for every infant in the French group the language used by the parents and directed to the baby was French. Other relatives, caregivers, and visitors also interacted with the baby in French. The language in the home with respect to TV and radio was predominantly French but including some English for 8 of the 19 French infants.

C. Procedure

Infants were tested in the headturn procedure using the protocol described in Polka and Werker (1994). In this procedure a syllable is played from a loudspeaker every 1500 ms and at random intervals this background syllable changes to a target syllable for a brief interval. Testing began by first conditioning the infant to turn his/her head toward a visual reinforcer when a change in the background syllable occurred. Correct headturns are reinforced by the activation of a visual reinforcer (an electromechanical animal that moves) and the verbal praise of an assistant. We implemented this procedure as a category change paradigm in which the background and the target consist of multiple tokens of each syllable type.

Perception of each stop/fricative contrast was tested in a single session. The infant was seated on the parent's lap across a small table from an experimenter (E1). The loudspeaker and an array of four visual reinforcers, located behind a smoked Plexiglas panel, were located to the left of the parent and infant. Both the parent and E1 listened to vocal music over headphones to prevent them from influencing the infant's behavior. A second experimenter (E2) located outside the test room observed the infant through a one-way window and operated the computer.

The session began with a conditioning stage in which the infant was given an opportunity to learn the contingency between the consonant change and the visual reinforcer. Only a single exemplar of each syllable type was used in the conditioning stage. Once the infant had made at least three consecutive correct anticipatory headturns, the testing stage began. Conditioning continued up to a maximum of 20 trials. During the testing stage multiple tokens of each syllable type were presented as the background and target stimuli. Also, both change and control (no-change) trials were presented according to a semirandom schedule in which no more than three consecutive control or change trials could occur. In the testing stage, E2 initiated trials when the infant was in a ''state of readiness'' (facing E1, not fussing, etc.). E2 was blind to the trial type and pushed a button when she observed a headturn during the trial interval. The visual reinforcer was activated automatically on a change trial when E2 recorded a headturn within a 4.5-s window; three syllables are presented during this interval. A retraining protocol was also implemented during the test stage to provide up to six retraining trials (change trials that are automatically reinforced if no headturn occurs). Performance on retraining trials was excluded from all data analyses. Twenty-five trials were presented during the testing stage; approximately 55 percent of the test trials were change trials and 45 percent were control trials.[6]

Infant testing was conducted in a sound-treated chamber. The stimuli presentation was controlled on-line from an IBM format computer via a data Translation DT2801 D/A board and stimuli were routed through a Yamaha AX-350 amplifier for delivery at an intensity of 68 dBA (at approximate ear level of the baby) via a Cyrus 780 loudspeaker. Computer software controlled the stimulus delivery, activation of the reinforcers, and trial selection, and recorded hits, misses, correct rejections, and false alarms.

Adult subjects were tested using the same basic procedure as the infants. Each adult was seated in the same chair as the parents had used and was read a short set of instructions that instructed them to raise their hand after hearing a sound change. The reinforcers were activated just as with the infants. Thus, adults experienced the same masking effects created by the noise (if any) and also received feedback for their correct responses. In the conditioning stage, adults were given a minimum of three and a maximum of six conditioning trials. In testing only 25 test trials were presented; no retraining trials were presented.

D. Design

Each subject was tested on both contrasts. With the infants, testing was completed in two sessions conducted on different days, with /bot/–/vot/ on day 1 and /doz/–/ðoz/ on day 2. Infant pilot testing with /bot/–/vot/ revealed that differentiation was more difficult when /bot/ was the background than when /vot/ was played as the background. Following this, it was decided that all infants would be tested with /vot/ as the background (and /bot/ as the target). This would allow a greater number of babies to succeed in the control condition and would also prevent differences in performance with the control condition from influencing performance in the test condition. Only infants who met a minimum criterion of eight out of ten consecutive correct responses in the testing phase on day 1 were invited to continue the experiment on day 2.

On day 2, infants were tested on the /doz/–/ðoz/ contrast. For this condition, the direction was counterbalanced between subjects such that half were tested with /doz/ as the background stimulus and half were tested for /ðoz/ as the background. If a minimum criterion of 8/10 correct was not met on day 2, a retest was immediately implemented using the /vot/–/bot/ stimuli from day 1. The retesting was conducted until the 8/10 minimum criterion had been met or a maximum of 15 test trials was administered. The retest was conducted to determine whether the infant was still on task and thus to assess whether poor performance was due specifically to difficulty differentiating the test stimuli or due to general inattentiveness and/or fatigue.

The adult test protocol differed from that of the infants in three ways. First, French- and English-speaking adults were tested in a single session in which both stimulus pairs were tested: /vot/–/bot/ followed by /doz/–/ðoz/. Second, the direction of change was counterbalanced across subjects for both contrasts. Finally, adults were not given a retest following poor performance on /doz/–/ðoz/.

III. RESULTS

Data obtained from 42 babies were included in our analyses. Each of these babies provided evidence that she/he could perform the headturn procedure by meeting one or two criteria. To meet criterion one, the infant had to respond correctly on eight out of ten consecutive trials on /b–v/ on day 1 and either on /d–ð/ or on the /b–v/ retest on day 2. Thirty-three babies met this criterion. Infants were also included if they met a second criterion. A second criterion was

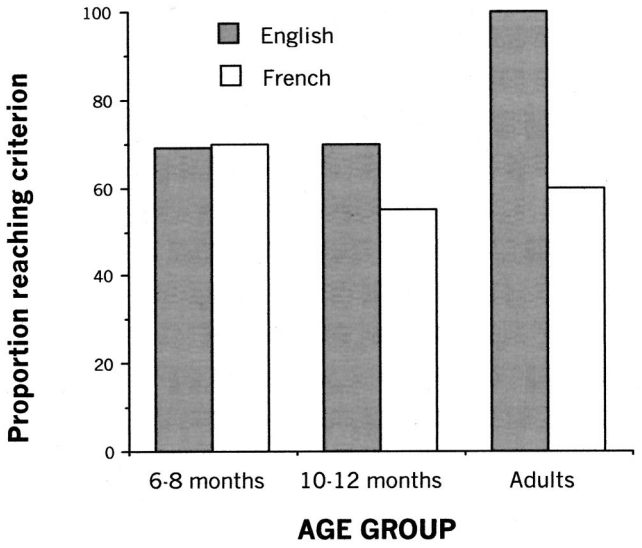


FIG. 1. Proportion of subjects reaching criterion on differentiation of /d–ð/ in each age and language group.

defined because we found that very few infants could complete the /b–v/ retest. These data and our observations indicated that presenting the retest immediately following the /d–ð/ test was too taxing for many infants. To meet criterion 2, the infant had to respond correctly on seven out of eight consecutive trials and show an overall accuracy greater than 60% on day 1. Thus, these infants could fall below the 8/10 criterion on both /d–ð/ and the /b–v/ retest on day 2, but they were required to show a higher level of performance on day 1 compared to criterion 1 infants.[7] Nine infants met criterion 2; five French infants (two 6- to 8-month-olds; three 10- to 12-month-olds) and four English infants (three 6- to 8-month-olds; one 10- to 12-month-old). Each analysis was repeated with criterion 2 babies removed from the sample and in every case the pattern of results obtained was identical.

The proportion of subjects reaching an 8/10 criterion on /d–ð/ is shown for each age and language group in Fig. 1. Over 55% of the subjects in every group reached this criterion. Chi-squares were conducted to determine whether the proportion of subjects reaching criterion in each group was different for those subjects tested on a /d/ to /ð/ change versus those tested on a /ð/ to /d/ change. No direction effects were evident in any of the groups; thus, all subsequent analyses were conducted on proportions collapsed across direction of change. Chi-squares were computed to assess the effect of language in each age group. The results showed that significantly more adults reached the criterion in the English group than in the French group, $X^2=6.64$, $p<0.01$. The proportion of French and English infants reaching criterion did not differ for either the 6- to 8-month-olds or the 10- to 12-month-olds. Separate analyses of proportions (ANPROs) as described by Marascuilo (1966) were conduced for each language group to analyze effects of age (6–8 months versus 10–12 months versus adults). The ANPRO analyses failed to reach significance for either language group. Thus, there were no age differences in either language group with respect to the proportion of individual subjects reaching criterion.

To analyze for more subtle effects of age and language

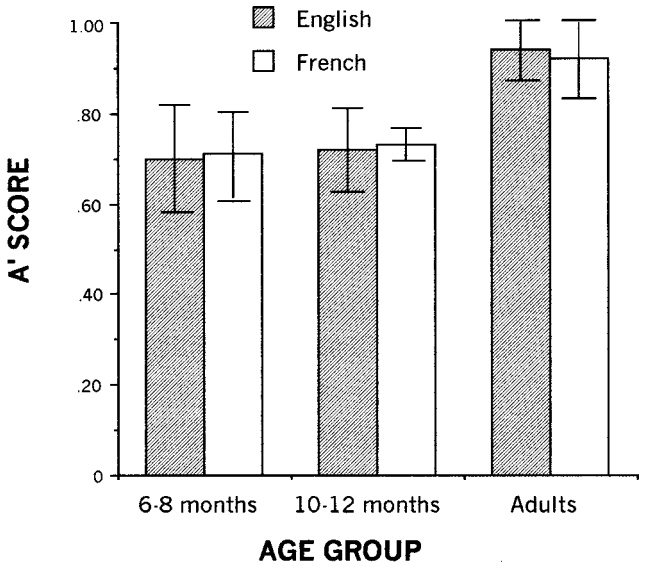


FIG. 2. Mean A-prime scores on differentiation of /b–v/ in each age and language group. Error bars show +1 to −1 SD.

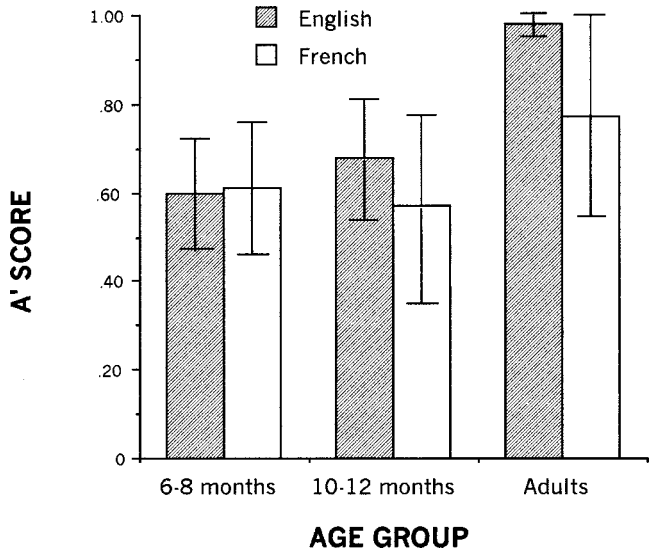


FIG. 3. Mean A-prime scores on differentiation of /d–ð/ in each age and language group. Error bars show +1 to −1 SD.

experience on perceptual performance, we computed A-prime scores for each subject on each contrast.[8] A-prime is a nonparametric index of sensitivity (similar to d-prime), which ranges from 0 to 1 (with 1 being a perfect score, 0.5 representing chance). The A-prime scores reflect each individual subject's hit rate corrected for their rate of false alams (Grier, 1971).[9] Mean A-prime values are plotted for differentiation of the /b–v/ and /d–ð/ contrasts in Figs. 2 and 3, respectively; error bars show the range corresponding to plus and minus 1 standard deviation. For the /b–v/ contrast (Fig. 2), the mean A-prime scores for the English and the French groups are almost identical at each age. For /b–v/, the variability in performance is similar for French and English subjects in the 6–8 month group and in the adult group, whereas the older French infants show less variability compared to the older English infants. To assess whether there were differences among the groups on differentiation of the /b–v/ control contrast, the A-prime scores were submitted to a two-way age (6–8 months, 10–12 months, adults) by language (English, French) analysis of variance (ANOVA). As expected, the main effect of language and the age by language interaction failed to reach significance. There was a main effect of age [$F(2,59)=29.36$, $p<0.01$]. Subsequent Tukey pairwise comparisons ($p<0.01$) showed that the adults performed better than both infant groups, but the two infant groups did not differ.

For the /d–ð/ contrast (Fig. 3), A-prime scores show divergent patterns of performance across the two language groups. For the English subjects, mean A-prime scores increase with age with a very large increase between infants and adults. For the French subjects, mean A-prime scores for the 6- to 8-month-olds and the 10- to 12-month-olds are almost identical, whereas A-prime scores for the French adults are higher. With age, intersubject variability increased for the French subjects and decreased for the English subjects. Intersubject variability is similar for French and English subjects at 6–8 months, whereas at 10–12 months and in the adults intersubject variability is clearly higher in the French group compared to the English group. The A-prime scores on the /d–ð/ contrast were analyzed in a three-way age (6–8 months, 10–12 months, adults) by language (English, French) by direction (d to ð; ð to d) ANOVA. Neither the main effect of direction nor any interactions of direction with the other factors were significant. As expected, the main effect of language [$F(1,59)=6.85$, $p<0.01$] was significant, showing that overall the English subjects performed better than the French subjects. The main effect of age was also significant [$F(2,59)=18.90$, $p<0.01$]. Subsequent Tukey pairwise comparisons ($p<0.01$) revealed that the adults did significantly better than both groups of infants, whereas the infant groups did not differ from each other. The only interaction to approach significance was age by language [$F(2,59)=2.72$, $p<0.07$].[10] Simple effects analyses were conducted to probe the age by language interaction. The effect of language was significant only for the adults [$F(1,59)=13$, $p<0.01$], consistent with the ANPRO results. The effect of age was significant for the English subjects [$F(2,34)=47.29$, $p<0.0001$], but not for the French subjects. Subsequent Tukey pairwise comparisons showed that the English adults performed better than both infant groups ($p<0.01$) but there was no significant difference between the two infant groups.[11]

IV. DISCUSSION

The present experiment provided a cross-language comparison of perceptual differentiation of the stop–fricative contrast /d–ð/ by English- and French-learning infants, and English- and French-speaking adults. Overall, the findings were inconsistent with our predictions regarding language effects and development patterns. With respect to language effects, we observed a robust effect of language experience in the adults, as expected. French adults were consistently less accurate and showed greater intersubject variability

compared to English adults. With respect to the 6- to 8-month-olds, there was no evidence of a language effect in any of our analyses, also as expected. However, contrary to our predictions, the 10- to 12-month-olds failed to show an effect of language experience. At this age, English and French infants did not differ either in the proportion reaching a preset criterion or in A-prime scores. There was only one subtle sign of any difference in the older infants, i.e., intersubject variability was observed to be higher in the older French infants compared to the older English infants. Thus, a robust effect of language experience was observed in adults but not in infants.

With respect to developmental patterns, the English and French listeners differed, but the specific development patterns did not match our predictions. For English listeners, comparable performance levels were maintained across both infant ages, whereas there was clear evidence of an increase in performance between 10–12 months and adulthood. This developmental pattern is statistically supported in analysis of A-prime scores and there is a clear trend towards the same pattern in the proportion of infants reaching a preset criterion. These results indicate that for native listeners, perceptual differentiation of /d–ð/ improves with increasing age and language experience after 12 months of age.

For French listeners, there were no significant changes in differentiation of the /d–ð/ with increasing age. No age effects were found in analyses of A-prime scores, or proportion reaching criterion. Greater intersubject variability was noted in the French adults and the older French infants compared to the younger French infants. These findings suggest that, for French listeners, the level of perceptual differentiation observed in infancy is maintained across development.

It is important to note that analysis of A-prime scores for the /b–v/ contrast showed the same pattern of age differences (adults $>$6–8$=$10–12) for both English and French subjects. This finding further confirms that the divergent patterns observed in French and English subjects' differentiation of /d–ð/ are attributable to differences in language experience.

Overall, our data fail to confirm our initial predictions regarding effects of age and language experience. Instead, two new findings emerged in the present study. First, perceptual differentiation of /d–ð/ in the headturn paradigm appears to be unaffected by language experience in the first year of life, even though very clear language effects are evident in adults. Second, it appears after 12 months of age language experience has a facilitative effect on perception of this contrast, in that significant age-related increase in differentiation of /d–ð/ is evident in the English groups but no change is evident in the French groups.

How do we explain our finding that perceptual differentiation of this contrast does not appear to be strongly affected by language experience within the first year of life? We have identified several plausible explanations; they are not mutually exclusive. First, within the framework provided by PAM (Best, 1995), it is conceivable that effects of language experience for this contrast are subtle because of the way in which this contrast relates to the phonology of French. Recall that previous studies have consistently shown language effects at 10 to 12 months of age for contrasts that fit the SC assimilation pattern for non-native adults and that earlier findings describe Francophone perception of English /d–ð/ as fitting this pattern. The high variability in performance of our French adults led us to question the latter conclusion. Therefore, we also gathered some data on assimilation of this contrast from the last six French adults that we tested using procedures outlines by Best (Best *et al.*, 1996) for making such assessments.[12] Four different assimilation patterns were supported in these data. One subject's data were consistent with a two-category (TC) pattern in that he identified English /d/ and /ð/ as good matches to ''d'' and ''th,'' respectively. Three of the subjects selected ''d'' to label both /d/ and /ð/; one of these subjects showed a single-category (SC) pattern and the other two subjects supported a category-goodness (CG) pattern. The two remaining subjects did not fit a clear assimilation pattern in that they selected both ''d'' and ''th'' to identify each phone.[13] Overall, this effort revealed that most French adults confuse /d/ and /ð/ to varying degrees, whereas a small subset of French adults perceived English /d/ as similar to French /d/ and recognized English /ð/ as being a non-native phone.[14] The order of performance among these six subjects also appears predictable from their assimilation data, with the TC subject showing the best performance, the two confused subjects and the SC subject doing the poorest, and the two CG subjects falling in between. Thus, the high intersubject variability in the French adults' performance is consistent with the observed differences in assimilation pattern. Given these findings, perhaps the greater intersubject variability found in the older French infants relative to the older English infants is indeed tied to differences in language experience.

It is also possible that the absence of clear language effects for this contrast at 10–12 months is due—at least in part—to the acoustic properties of these contrasting phones. As mentioned earlier, /d/ and /ð/ are short, low-amplitude phones (Pickett, 1999) and /ð/ has been reported to be among the weakest phones present in English (Fletcher, 1953; Ling and Ling, 1978). Thus, it could be that it is more difficult for infants to attend to such acoustically weak units. In fact, several earlier studies have suggested that infants have greater difficulty differentiating contrasts involving nonsibilant fricatives even in their native language (Eilers and Minifie, 1975; Eilers *et al.*, 1977; Holmberg *et al.*, 1977). Two aspects of our data further support this claim. First, our attrition due to inability to meet criterion with the /b–v/ control contrast was quite high in each infant group. Second, we found that 70 percent of the infants in each group were able to meet our criterion on /d–ð/. In comparison, successful differentiation has been observed in 80 to 90 percent of 6- to 8-month-olds tested on other consonant contrasts in studies using the same procedure and similar or even higher criteria (e.g., Werker *et al.*, 1981; Werker and Tees, 1984a).

Recall that Burnham (1986) has claimed that language effects emerge earlier in development (during the first year) for contrasts that meet two criteria that form the definition of a fragile contrast, i.e., low acoustic salience and rare occurrence among languages of the world. Given that the /d–ð/

contrast clearly fits this definition, our findings are clearly inconsistent with the hypothesis. Indeed, our findings appear to suggest the opposite—that fragile contrasts remain unaffected by language experience in the first year of life. Perhaps acoustically weak contrasts are less vulnerable to perceptual decline when relevant language experience is lacking because, although they can, infants do not readily attend to such differences. If so, then contrary to Burnham's proposal, an acoustically weak contrast appears to delay rather than to promote the onset of language specific tuning.

The unique phonotactic properties of English /ð/ may also contribute to the absence of language effects during infancy. The phoneme /ð/ occurs with a very high frequency in initial position in spoken English, but only in function words (Morgan, Shi, and Allopenna, 1996). Function words are less salient forms in natural discourse given that they are short, contain unstressed vowels, are typically not produced in isolation, and are not highlighted by intonation (Morgan, Shi, and Allopenna, 1996). Recent findings show that newborns are sensitive to the correlated acoustic and phonological properties that distinguish content and function words (Shi, Werker, and Morgan, 1999). Moreover, at 6 months of age infants prefer to listen to content words over function words, thus showing that infants are paying more attention to content words than function words (Shi and Werker, 2001). Accordingly, recognition of phonetic elements that appear exclusively in function words, such as /ð/, may not occur until infants begin to focus their attention on this class of syntactic elements.

An interesting connectionist model recently proposed by Behnke (1998) allows for both of the latter two possibilities to explain why language effects emerge later in development for some phonetic contrasts than for others. According to Behnke, delays may occur either because general limitations in auditory processing during infancy make it difficult for infants to differentiate certain phonetic contrasts (e.g., contrasts involving brief or low-amplitude phones) and/or because differentiation of some contrasts may remain difficult until the child has gained lexical knowledge that serves to fine-tune phonetic processing.

A fourth plausible account for the lack of language effect in our infants is that we may not have selected the right level of analysis to isolate language effects in perception of this contrast. In this study, we described language differences with respect to the presence versus absence of specific phones in French versus English. There are considerable data to support an approach in which language-specific processing is described in terms of the perceptual weighting of the acoustic cues underlying a given contrast (Harnsberger, 1999). Thus, to isolate effects of language experience, it may be more meaningful to consider the multiple cues underlying this stop–fricative contrast and the relative salience of each for French and English listeners. Morosan and Jamieson (1989) suggest that French and English adults differ in perceptual weighting of cues to the /d–ð/ contrast. Their perceptual data suggest that French adults rely on frication duration to identify /ð/. Thus, differences in cue weighting may contribute to the poorer performance and greater variability in older French infants and French adults and the varied assimilation patterns for the latter. The use of natural speech in this study does not allow us to isolate such differences. Further research using synthetic stimuli is needed to investigate the relationships between age, language experience, cue salience, and assimilation patterns.

The second finding in this study concerns the differences between infants and adults. Previous developmental cross-language studies generally support a maintenance view of language experience. In the maintenance view, specific language experience serves to prevent a developmental decline in perceptual differentiation for some contrasts. However, our data indicate that language experience serves to facilitate perception of /d–ð/ after 12 months of age; in the absence of language experience, adult performance is still comparable to the level observed in infancy. It should be noted that adults have been shown to perform better than infants in several previous cross-language studies (e.g., Polka and Werker, 1994; Polka and Bohn, 1996; Best *et al.*, 1990). However, these differences have been observed either only within non-native listeners or within both native and non-native listeners and thus indicate more general age-related changes. The differences in perceptual differentiation observed between English infants and adults in the present study most likely reflect some general age-related differences in task performance between these two age groups. However, given that the comparison of French adults and infants fails to reveal comparable improvement indicates that some of the age effects in the English group are attributable to a facilitative effect of language experience.

It is worth considering why a maintenance view has found clear support in the previous research while evidence in support of facilitation is lacking. There are at least two reasons. First, previous studies supporting a maintenance view have typically examined dichotomous measures of perceptual performance, such as a preset criterion or presence/absence of release from habituation, and thus have not always considered more subtle quantitative differences in performance. Second, they have typically not included a complete control group of native listeners at each age to compare with the non-native listeners. With the opportunity to compare both native and non-native subjects in each age group, the interpretation of continuous measures of perceptual accuracy becomes more meaningful. Thus, our data elaborate rather than challenge previous findings because the design used here permits us to bring more information to the interpretation of developmental patterns.

Although a facilitative effect of language experience on phonetic differentiation has not been clearly demonstrated in any previous developmental cross-language study, this finding is not surprising. None of the current models of phonetic development claim that phonetic perception is adult-like in a 1-year-old. Moreover, a facilitative role of language experience is entirely consistent with the literature on development of phonemic perception. Studies of phonemic perception in children learning English show that accuracy in phoneme perception improves during the preschool years such that perception of most native phonemic contrasts becomes adult-like between 2 and 10 years of age (Templin, 1957; Barton, 1980; Luksaneey-

anawin *et al.*, 1997). A great deal of research has focused on early infancy, yet there is much to be learned about the role of language experience in phonetic development during childhood. Werker and Tees (1983) reported that 4-, 8-, and 12-year-olds, as well as adults, had difficulty differentiating two Hindi contrasts that were difficult for 10- to 12-month-olds and adults. For the Hindi voicing contrast, none of the 4-year-olds they tested was able to reach criterion, whereas about half of the older children and adults did. Burnham (1986) has reported similar findings showing a dip in identification of a non-native voicing contrast at 6 years followed by some recovery in performance in older children and adults. It is not clear what factors contribute to these later changes in phonetic perception. Werker and Tees (1983) suggest that the rigid processing strategies often displayed by 4-year-olds may make it especially difficult for them to attend to nonphonemic differences. Burnham has proposed that a decline at 6 years occurs as a result of increased attention to phonemic structure typically encouraged at the onset of formal education (Burnham *et al.*, 1991; Burnham, 1986).

Our finding that perceptual differentiation did not change in the absence of language experience was more surprising. In all previous studies, clear language effects in adults have been associated with a decline in perceptual differentiation at an earlier point in development. It is possible that differentiation of this contrast does decline at a later point in development and then eventually recovers to the level observed in infancy.[15] Research underway in our lab will address this issue. Nevertheless, the present pattern of results suggests that when differentiation of a contrast is difficult in infancy there is no further development decline, whereas relevant language experience can act to boost this initial level of perception. For such contrasts advances in linguistic processing may be essential to fine-tune phonetic perception to levels observed in mature native listeners. Accordingly, for such contrasts robust language effects may not be evident until such linguistic maturity has been achieved. In this case, such contrasts may be especially useful to investigate the influence of language processing in the development of phonetic perception.

In summary, the present study has shown that perceptual differentiation of /d–ð/ is not influenced by language experience in the first year of life, but is clearly affected by language experience by adulthood. Further research is needed to determine whether this pattern of age and language effects is observed in other test paradigms or for other phonetic contrasts. Our future research will also aim to establish when facilitation effects emerge in perception of the /d–ð/ contrast, to determine whether a decline in Francophone perception is evident at this point in development, and to identify factors that contribute to the facilitative effect observed for this contrast.

ACKNOWLEDGMENTS

This work was supported in part by a scholarship from Fonds pour la Formation de Chercheurs et l'Aide à la Recherche to the second author as well as a grant from the Natural Science and Engineering Research Council of Canada awarded to the first author (No. OGP0105397). We would like to thank the adults, parents, and infants who participated in this study and also the following individuals who contributed to the data collection phase of the study: Lynne Herscovitch, Elisabeth Christe, Rachel Clarke, Miranda Entwistle, Domenic Haines, and Michelle Hayhoe.

[1]Speech perception tasks used with infants generally do not assess the limits of their sensory capacities. For this reason we use the term differentiation throughout this paper in an effort to be clear that the perceptual measures of interest are not to be equated with sensory capacity but also reflect task-specific demands on memory, attention, and motivation.

[2]A different dialect of German was used in this study than in Polka and Werker (1994) reported above; see Polka and Bohn (1996) for further discussion of the discrepant results in these two studies.

[3]Comparisons of several different spectral representations of the initial aperiodic noise segments failed to reveal differences for either contrast.

[4]Attrition due to poor performance on the /b–v/ control contrast was similar across the infant groups; English 6- to 8-month-olds=34%, French 6- to 8-month-olds=38%, English 10- to 12-month-olds=37%, and French 10- to 12-month-olds=23%.

[5]For four of these infants the fathers also spoke Armenian, French, Ukranian, or Italian; for one infant the mother also spoke Italian, and for one infant both parents also spoke Italian.

[6]Five additional trials were presented if the infant was within 2 trials of reaching the 8/10 criterion. A slightly higher proportion of change trials is often used with infants to ensure sufficient reinforcement to sustain task performance.

[7]Some debate surrounds the calculation of probability of attaining a preset criterion in the headturn procedure. However, regardless of calculation method used, p levels are always lower for 7/8 (87.5%) than for 8/10 (80%) for a 25-trial sample, probability estimate for both criteria is at least 0.05. Six of the nine infants meeting criterion 2 got 9/10 correct on day 1; the estimated probability for this criterion is at least 0.01.

[8]Every analysis of A-prime scores reported here was repeated using percent correct as the dependent variable and produced the same pattern of results.

[9]The formula used was

$$A' = 0.5 + (H - FA)(1 + H - FA)/[4H(1 - FA)],$$

where H=proportion of hits and FA=proportion of false alarms.

[10]In the analysis of percent correct scores this interaction was highly significant [$F(2,59)=6.84$, $p<0.005$].

[11]A language by age by contrast ANOVA including both contrasts revealed a main effect for contrast and a language by contrast interaction. Simple effect analysis of contrasts revealed that French subjects performed worse on /d–ð/ than on /b–v/ whereas performance on the two contrasts did not differ for English subjects. Simple effects of analysis of language also showed that language effects were evident for /d–ð/ but not for /b–v/.

[12]Following the headturn task, these subjects were asked to identify the initial consonant in each doze, those, boat, and vote syllable using a closed response set (b, p, d, t, th, v, or other) and then to immediately rate how well the item matched their selection on a scale of 1 (very poor match) to 5 (very good match). We included ''th'' as a response alternative because Francophone adults are generally aware that English contains this problematic ''th'' sound.

[13]This outcome most likely reflects an unsuccessful attempt to guess when they are hearing ''th.''

[14]With the closed-set identification task used we cannot be sure that French adults can accurately label English /ð/ or simply chose the ''th'' response instead of ''other'' (which was never chosen) when they were not sure what sound they heard.

[15]It may be that the Francophone adults' minimal exposure to English is relevant to a possible recovery. Of interest in this respect, we noted that differences in amount of exposure to English among the Francophone adults (measured via a questionnaire) appear to be unrelated to /d–ð/ differentiation accuracy.

Barton, D. **(1980)**. ''Phonemic perception in children,'' in *Child Phonology Volume 2: Perception*, edited by G. Yeni-Komshian, J. Kavanagh, and C. A. Ferguson (Academic, New York), pp. 97–116.

Behnke, K. **(1998)**. ''The acquisition of phonetic categories in young infants: A self organizing artificial neural network approach,'' *MPI Series in*

Psycholinguistics (Ponsen and Looijen bv, Wageningen).
Best, C. T. (**1990**). ''Adult perception of non-native contrasts differing in assimilation to native phonological categories,'' J. Acoust. Soc. Am. **88**, S177.
Best, C. T. **(1991)**. ''Phonetic influences on the perception of non-native speech contrasts by 6–8- and 10–12-month-olds,'' Paper presented at the meeting of Society for Research in Child Development, Seattle, WA, April.
Best, C. T. **(1993)**. ''Emergence of language-specific constraints in perception of non-native speech perception: A window on early phonological development,'' in *Developmental Neurocognition: Speech and Face Processing in the First Year of Life*, edited by B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. MacNeilage, and J. Morton (Kluwer, Dordrecht).
Best, C. T. **(1994a)**. ''Learning to perceive the sound pattern of English,'' in *Advances in Infancy Research*, edited by C. Rovee-Collier and L. Lipsitt (Albex, Hillsdale, NJ).
Best, C. T. **(1994b)**. ''The emergence of native-language phonological influences in infants: A perceptual assimilation model,'' in *The Development of Speech Perception: The Transition From Speech to Spoken Words*, edited by J. Goodman and H. C. Nusbaum (MIT Press, Cambridge, MA).
Best, C. T. **(1995)**. ''A direct realist view of cross-language speech perception,'' in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Baltimore), pp. 171–204.
Best, C. T., Faber, A., and Levitt, A. (**1996**). ''Perceptual assimilation of non-native vowel contrasts to the American English vowel system,'' J. Acoust. Soc. Am. **99**, 2602(A).
Best, C. T., McRoberts, G. W., Goodell, E., Wormer, J. S., Insabella, G., Kim, P., Klatt, L., Luke, S., and Silver, J. (**1990**). ''Infant and adult perception of non-native speech contrasts differing in relation to the listener's native phonology,'' Paper presented at Meeting of the International Conference on Infant Studies, Montreal, April, 1990.
Best, C. T., McRoberts, G. W., Lafleur, R., and Silver-Isenstadt, J. (**1995**). ''Divergent developmental patterns for infants' perception of two nonnative consonant contrasts,'' Infant Behav. Dev. **18**, 339–350.
Best, C. T., McRoberts, G. W., and Sithole, N. N. (**1988**). ''Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants,'' J. Exp. Psychol.: Human Percept. Perform. **14**, 345–360.
Best, C., Singh, L., Bouchard, J., Connelly, G., Cook, A., and Faber, A. **(1997)**. ''Developmental changes in infants' discrimination of non-native vowels that adults assimilate to two native categories,'' Poster presented at the Meeting of Society of Research in Child Development, 3–6 April 1997.
Burnham, D. K. (**1986**). ''Developmental loss of speech perception: Exposure to and experience with a first language,'' Appl. Psycholing. **7**, 201–240.
Burnham, D. K., Earnshaw, L. J., and Clark, J. E. **(1991)**. ''Development of categorical identification of native and non-native bilabial stops: Infants, children and adults,'' J. Child Language **18**, 231–260.
Eilers, R. E., and Minifie, F. D. (**1975**). ''Fricative discrimination in early infancy,'' J. Speech Hear. Res. **18**, 158–167.
Eilers, R. E., Wilson, W. R., and Moore, J. M. (**1977**). ''Developmental changes in speech discrimination in infants,'' J. Speech Hear. Res. **20**, 766–780.
Eimas, P. D., Siqueland, E. R., Jusczyk, P., and Vigorito, J. (**1971**). ''Speech perception in infants,'' Science **171**, 303–306.
Fant, G. **(1970)**. *Acoustic Theory of Speech Production* (The Hague, Mouton).
Fletcher, H. **(1953)**. *Speech and Hearing in Communication* (Van Nostrand, Princeton, NJ).
Grier, J. B. (**1971**). ''Nonparametric indexes for sensitivity and bias: Computing formulas,'' Psychol. Bull. **75**, 424–429.
Harnsberger, J. D. **(1999)**. ''The effect of linguistic experience on perceptual similarity among nasal consonants: A multidimensional scaling analysis,'' Research on Spoken Language Processing, Progress Report **23**, 227–239.
Holmberg, T. L., Morgan, K. A., and Kuhl, P. K. **(1977)**. ''Speech perception in early infancy: Discrimination of fricative consonants,'' Presented at the 94th Meeting of the Acoustical Society of America, Miami Beach, FL, 12–16 December 1997.
Jamieson, D. G., and Morosan, D. E. (**1986**). ''Training non-native speech contrasts in adults: Acquisition of English /ɵ/–/ð/ contrast by francophones,'' Percept. Psychophys. **40**, 205–215.
Jamieson, D., and Morosan, D. E. (**1989**). ''Training new, non-native speech contrasts: A comparison of the prototype and perceptual fading techniques,'' Can. J. Psychol. **43**, 88–96.
Kuhl, P. J., Williams, K. A., Lacerda, F., Stevens, K. N., and Lindblom, B. (**1992**). ''Linguistic experience alters phonetic perception in infants by 6 months of age,'' Science **255**, 606–608.
Lalonde, C. E., and Werker, J. F. (**1995**). ''Cognitive influences on cross-language speech perception in infancy,'' Infant Behav. Dev. **18**, 459–475.
Ling, D., and Ling A. **(1978)**. *Aural Rehabilitation* (Alexander Graham Bell Association for the Deaf, Washington, DC).
Logan, J. S., Lively, S. E., and Pisoni, D. B. (**1991**). ''Training Japanese listeners to identify English /r/ and /l/,'' J. Acoust. Soc. Am. **89**, 874–886.
Luksaneeyanawin, S., Burnham, D. K., Francis, E., and Pansottee, S. (**1997**). ''The role of L1 background and L2 instruction in the perception of fricative contrasts: Thai and English children and adults,'' Asia Pac. J. Speech, Lang. Hear. **2**, 25–42.
Maddieson, I. **(1984)**. *Patterns of Sounds* (Cambridge University Press, Cambridge, England).
Marascuilo, L. A. (**1966**). ''Large scale multiple comparison,'' Psychol. Bull. **69**, 280–290.
Mertus, J. **(1990)**. BLISS *User's Manual* (Brown University, Providence, RI).
Morgan, J. L., Shi, R., and Allopenna, P. **(1996)**. ''Perceptual basis of rudimentary grammatical categories: Toward a broader conceptualization of bootstrapping,'' in *Signal to Syntax*, edited by J. L. Morgan and K. Demuth (Erlbaum, Mahwah, NJ), pp. 263–283.
Morosan, D. E., and Jamieson, D. G. (**1989**). ''Evaluation of a technique for training new speech contrasts: Generalization across voices, but not word position or task,'' J. Speech Hear. Res. **32**, 501–511.
Pegg, J. E., and Werker, J. F. (**1997**). ''Adult and infant perception of two English phones,'' J. Acoust. Soc. Am. **102**, 3742–3753.
Pickett, J. M. **(1999)**. *The Acoustics of Speech Communication* (Allyn and Bacon, Needham Heights, MA).
Pisoni, D. B., Aslin, R. N., Perey, A. J., and Hennessy, B. L. (**1982**). ''Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants,'' J. Exp. Psychol. Human Percept. Perform **8**, 297–314.
Polka, L. (**1995**). ''Lingusitic influences in adult perception of non-native vowel contrasts,'' J. Acoust. Soc. Am. **97**, 1286–1296.
Polka, L., and Bohn, O-S. (**1996**). ''A cross-language comparison of vowel perception in English-learning and German-learning infants,'' J. Acoust. Soc. Am. **100**, 577–592.
Polka, L., and Werker, J. F. (**1994**). ''Developmental changes in perception of non-native vowel contrasts,'' J. Exp. Psychol. Human Percept. Perform. **20**, 421–435.
Shi, R., and Werker, J. F. (**2001**). ''Six-month-old infants preferences for lexical over grammatical words,'' Psychol. Sci. **12**, 71–76.
Shi, R., Werker, J. F., and Morgan, J. L. (**1999**). ''Newborn infants' sensitivity to perceptual cues to lexical and grammatical words,'' Cognition **72**, B11–B21.
Tees, R. C., and Werker, J. F. (**1984**). ''Perceptual flexibility: Maintenance or recovery of the ability to discriminate non-native speech sounds,'' Can. J. Psychol. **38**, 579–590.
Templin, M. C. **(1957)**. *Certain Language Skills in Children* (University of Minnesota Press, Minneapolis).
Tsushima, T., Takizawa, O., Sasaki, M., Shiraki, S., Nishi, K., Kohno, M., Menyuk, P., and Best, C. **(1994)**. ''Discrimination of English /r-l/ and /w-y/ by Japanese infants at 6–12 months: Language-specific developmental changes in speech perception abilities,'' in *Proceedings of the International Conference on Spoken Language Processing*, Acoustical Society of Japan, Vol. 4, pp. 1695–1698.
Werker, J. F., Frost, P. E., and McGurk, H. (**1992**). ''Cross-language influences on bimodal speech perception,'' Can. J. Psychol. **46**, 551–568.
Werker, J. F., Gilbert, J. H. V., Humphrey, K., and Tees, R. C. (**1981**). ''Developmental aspects of cross-language speech perception,'' Child Dev. **52**, 349–353.
Werker, J. F., Lloyd, V., Pegg, J. E., and Polka, L. **(1995)**. ''Putting the baby in the bootstraps: Toward a more complete understanding of the role of input in speech processing,'' in *Signal to Syntax*, edited by J. Morgan and K. Demuth (Erlbaum, Hillsdale, NJ), pp. 427–447.
Werker, J. F., and Pegg, J. E. **(1992)**. ''Infant speech perception and phonological acquisition,'' in *Phonological Development: Models, Research, and Implications*, edited by C. Ferguson, L. Menn, and C. Stoel-Gammon

(York, Parkton, MD), pp. 285–312.
Werker, J. F. and Tees, R. C. (**1983**). ''Developmental changes across childhood in the perception of non-native speech sounds,'' Can. J. Psychol. **37**, 278–286.
Werker, J. F., and Tees, R. C. (**1984a**). ''Cross-language speech perception: Evidence for perceptual reorganization during the first year of life,'' Infant Behav. Dev. **7**, 49–63.
Werker, J. F., and Tees, R. C. (**1984b**). ''Phonemic and phonetic factors in adult cross-language speech perception,'' J. Acoust. Soc. Am. **75**, 1866–1878.

The effect of intensity perturbations on speech intelligibility for normal-hearing and hearing-impaired listeners[a)]

Nicolle H. van Schijndel,[b)] Tammo Houtgast, and Joost M. Festen[c)]
Department of Otolaryngology, University Hospital VU, P.O. Box 7057, 1007 MB Amsterdam, The Netherlands

(Received 22 July 1999; accepted for publication 29 January 2001)

Hearing-impaired listeners are known to suffer from reduced speech intelligibility in noise, even if sounds are above their hearing thresholds. This study examined the possible contribution of reduced acuity of intensity coding to this problem. The ''distortion-sensitivity model'' was used: the effect of reduced acuity of auditory intensity coding on intelligibility was mimicked by an artificial distortion of the speech intensity coding, and the sensitivity to this distortion for hearing-impaired listeners was compared with that for normal-hearing listeners. Stimuli (speech plus noise) were wavelet coded using a Gaussian wavelet (1/4-octave bandwidth). The intensity coding was distorted by multiplying the modulus of each wavelet coefficient by a random factor. Speech-reception thresholds (SRTs) were measured for various degrees of intensity perturbation. Hearing-impaired listeners were classified as suffering from suprathreshold deficits if intelligibility of undistorted speech was worse than predicted from audibility by the speech intelligibility index model [ANSI, ANSI S3.5-1997 (1997)]. Hearing-impaired listeners without suprathreshold deficits were as sensitive to the intensity distortion as the normal-hearing listeners. Hearing-impaired listeners with suprathreshold deficits appeared to be less sensitive. Results indicate that reduced acuity of auditory intensity coding may be a factor underlying reduced speech intelligibility in noise for the hearing impaired. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1358301]

PACS numbers: 43.71.Ky, 43.66.Dc [RVS]

I. INTRODUCTION

Speech recognition (or intelligibility) in noisy environments is a problem for many hearing-impaired listeners. This problem may result from inaudibility of part of the speech spectrum. However, even if sounds are above hearing thresholds over the whole frequency range, some hearing-impaired listeners still have problems perceiving speech in noise. Their speech processing is not as good as that of normal-hearing listeners due to suprathreshold deficits (Moore, 1996; Noordhoek *et al.*, 2000). Examples of suprathreshold deficits are reduced spectral resolution (frequency selectivity), reduced temporal resolution, impaired frequency discrimination, or impaired loudness perception.

This study examines a deficit related to impaired loudness perception, i.e., reduced intensity coding. Reduced intensity coding may be thought of as a less accurate intensity representation in the auditory periphery. This may be due, for example, to a loss of auditory-nerve fibers, resulting in a more noisy intensity coding. Reduced intensity coding may lead to higher just-noticeable differences (jnd's) in intensity or fewer jnd's. A few studies suggest that intensity coding may be disrupted for some listeners with cochlear damage (Florentine *et al.*, 1993; Buus *et al.*, 1995; Moore, 1995). Reduced intensity coding acuity is likely to affect speech intelligibility in noise. However, the literature does not report any attempts to relate distorted auditory intensity coding to speech-perception problems in noise.

The main question in this study is whether poor auditory intensity coding is at least partly responsible for the observed poor speech intelligibility in noise by hearing-impaired listeners. This is examined by introducing an artificial distortion in the intensities of speech. The distortion simulates the effect of reduced acuity of auditory intensity coding on speech perception. Speech-reception thresholds (SRTs) for various degrees of the applied artificial distortion are compared for normal-hearing and hearing-impaired listeners in order to clarify the contribution of reduced auditory intensity coding acuity to impaired speech intelligibility. This type of experiment may be called a ''distortion-sensitivity approach'' (Houtgast, 1995).

Under the distortion-sensitivity model, a specific type of distortion is applied to combined speech and noise stimulus. Intelligibility tests are administered in order to determine whether the artificial distortion is related or not to the suprathreshold deficit of hearing-impaired listeners. Therefore, intelligibility is measured as a function of the degree of the distortion, and sensitivity to the distortion is compared for normal-hearing and hearing-impaired listeners. Figure 1 is a schematic illustration of the possible outcomes of such an experiment.

In the undistorted condition, using the original, unmodified speech, hearing-impaired listeners usually recognize speech more poorly than normal-hearing listeners. When comparing speech intelligibility by normal-hearing and

[a)]Part of the data of the normal-hearing listeners was presented at the 16th International Congress on Acoustics and the 135th Meeting of the Acoustical Society of America [Proceedings of ICA/ASA '98, Seattle, WA, 2599–2600 (1998)].

[b)]Current address: Philips Research Laboratories, Digital Signal Processing (WY 82), Prof. Holstlaan 4, 5656 AA Eindhoven, The Netherlands. Electronic mail: Nicolle.van.Schijndel@philips.com
[c)]Electronic mail: JM.Festen@azvu.nl

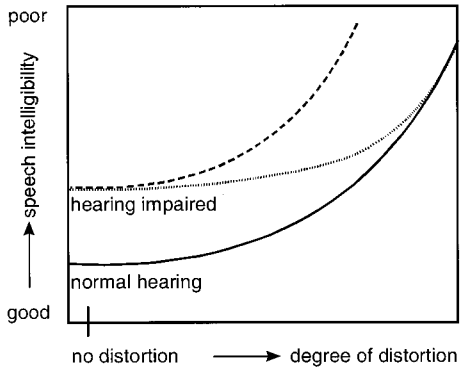


FIG. 1. Illustration of the distortion-sensitivity model. Performance measured as a function of the degree of distortion of hearing-impaired listeners is compared with that of normal-hearing listeners (solid line). Two possible outcomes of such an experiment: ''convergence'' (dotted and solid line), in which case the performance of the hearing-impaired listeners approaches that of the normal-hearing listeners for increasing degrees of distortion, or ''no convergence'' (dashed and solid line), in which case the performance of the hearing-impaired listeners suffers as much from increasing degrees of distortion as that of the normal-hearing listeners. See the text for further explanation.

hearing-impaired listeners as a function of the degree of distortion, essentially, two different trends may be hypothesized. First, performance of the normal-hearing and hearing-impaired listeners converges (solid and dotted line). Second, performance of the normal-hearing and hearing-impaired listeners does not converge (solid and dashed line).

In the convergence case, the performance difference between normal-hearing and hearing-impaired listeners becomes smaller as a function of the distortion level. For high levels of distortion, performance becomes essentially equal. Thus, hearing-impaired listeners are less sensitive to the distortion than normal-hearing listeners. In statistical terms, this is an interaction between listener groups and level of distortion or, stated differently, between hearing deficit and distortion. In terms of interpretation, the effect of the artificial distortion is smaller for hearing-impaired listeners because the hearing deficit already affects the speech processing in a similar way. Thus, the specific type of artificial distortion for which convergence is observed hints at the suprathreshold deficit causing the speech intelligibility problems of the hearing impaired.

In the no-convergence case, performance by normal-hearing and hearing-impaired listeners does not come close together. Hearing-impaired listeners are just as sensitive to distortion as normal-hearing listeners. This suggests that the effects of this type of artificial distortion are not related to the suprathreshold hearing deficits causing impaired speech intelligibility. It should be noted that, as the type of suprathreshold hearing deficit may be listener dependent, studying the results of individual listeners is important.

The distortion-sensitivity model can be illustrated by a simple example. Let us assume that a hearing-impaired listener suffers from a severe high-frequency hearing loss. The applied artificial distortion is low-pass filtering of the speech signal. Speech intelligibility is measured as a function of the cutoff frequency of the low-pass filter. Low-pass filtering reduces the speech intelligibility. Compared with normal-hearing listeners, the hearing-impaired listener is less sensitive to the low-pass filtering. This is because the high frequencies in the broadband signal are not perceived anyway. Convergence of the performance of normal-hearing and hearing-impaired listeners suggests that low-pass filtering relates to the problem experienced by the hearing-impaired listener, i.e., the listener misses some part of the high-frequency spectrum.

In this study, *artificial* distortion of the intensity coding tries to mimic poor *auditory* intensity coding. To simulate poor intensity coding, a model of ''normal'' auditory intensity analysis is required. Auditory analysis is a spectro-temporal analysis. This is modeled by wavelet decomposition. Wavelet analysis is used for mimicking auditory time–frequency analysis because of its logarithmic frequency scale (see, e.g., Rioul and Vetterli, 1991). In van Schijndel *et al.* (1999) it was shown that auditory spectral and temporal resolution are roughly matched by using a Gaussian-shaped mother wavelet (prototype analysis function) with a bandwidth of 1/4 octave. Using this perceptually relevant time–frequency analysis, specific manipulations of the wavelet coefficients may be used to simulate specific changes in auditory coding. Therefore, a reduced acuity in auditory intensity coding may be simulated by introducing random perturbation in the intensity of the wavelet coefficients.

In summary, the aim of this study is to investigate if reduced speech intelligibility by hearing-impaired listeners may be explained by reduced intensity coding. This question is addressed by a distortion-sensitivity model in which an artificial distortion of the intensities in a speech-plus-noise stimulus between wavelet decomposition and recomposition is applied. Intelligibility is measured as a function of the degree of distortion, and the sensitivity of hearing-impaired and normal-hearing listeners is compared. The rationale behind the distortion-sensitivity model is that when a hearing-impaired listener is less sensitive to the intensity distortion than normal-hearing listeners this may indicate that poor auditory intensity coding is causing part of the speech intelligibility problems.

II. METHOD

A. Distortion of wavelet coded intensities

In this study, intensity coding of sound is distorted to mimic the effects of poor auditory intensity coding. By means of the speech-reception threshold test (SRT; for an explanation, see Sec. II D), speech intelligibility of sentences is measured as a function of the degree of applied artificial intensity distortion. In order to simulate auditory intensity coding, a perceptually relevant spectro-temporal analysis method has been developed.

To model auditory spectro-temporal coding, sounds were described in the time–frequency domain by means of a wavelet transform. Compared with the short-time Fourier transform, the wavelet transform matches auditory system coding more closely because it uses a logarithmic frequency scale (e.g., Rioul and Vetterli, 1991). In this study, the criterion for the choice of the mother wavelet is its spectral (and temporal) resolution. Results of van Schijndel *et al.* (1999)

suggest that a Gaussian-windowed sinusoid with a shape factor between 0.15 and 0.3 roughly matches the auditory time–frequency window. Therefore, as the prototype analysis function, a complex sinusoidal carrier with a Gaussian envelope was chosen. This Gaussian wavelet is described by

$$s(t) = \sqrt{\alpha f_0}\exp(i2\pi f_0 t)\exp[-\pi(\alpha f_0 t)^2], \quad (1)$$

in which f_0 is the carrier frequency, α is the shape factor, and $\sqrt{\alpha f_0}$ normalizes the energy of the analysis function. This time-frequency window has an effective bandwidth of $\Delta_f = \alpha f_0$ and an effective duration of $\Delta_t = 1/(\alpha f_0)$ (see Gabor, 1947). The shape factor α was set to 0.1735. Thus, the effective bandwidth of the analysis function was $\frac{1}{4}$ octave [about the auditory critical band (see Scharf, 1970)]. As a result, the effective duration of the time–frequency window is 5.76 ms at 1 kHz (1.44 ms at 4 kHz). The effective number of sinusoidal periods contained within the Gaussian envelope equals 5.8 (i.e., $1/\alpha$).

This Gaussian wavelet is used to construct a wavelet decomposition that covers the time–frequency plane. Shifts of this prototype function cover the temporal domain; scales of the prototype function cover the spectral domain. The scaling is controlled by varying the carrier frequency f_0. For simultaneous sampling in time and frequency, the Nyquist sampling theorem was used (see Allen, 1977; Allen and Rabiner, 1977). This theorem is based on the bandwidth and duration of the analysis function. Because the Gaussian wavelet does not have compact support[1] in time nor in frequency, the 25-dB down points were taken as an estimate of the upper limit of bandwidth and duration of the analysis functions. This leads to a sampling of one wavelet per three periods of the wavelet carrier frequency along the time axis, and eight wavelets per octave along the frequency axis. The theoretical number of complex coefficients needed to describe the signal is about 2 per input sample (see Allen, 1977). In this study, the information of the signals was limited to the frequency range from 250 to 4000 Hz. As a result, the number of coefficients computed per input sample could be limited to about unity. Thus, 1 second of speech (sampling frequency: 15 625 Hz) was described by 16×10^3 complex wavelet coefficients in which no information below 250 Hz and above 4 kHz was preserved.

Using these coefficients, sounds can be reconstructed by an overlap-add procedure. Theoretically, the reconstruction is not perfect. However, using the above-described time–frequency tiling, differences between the original signal and the reconstructed signal are very small and not noticeable to a listener.

After the wavelet analysis, the modulus of each wavelet coefficient was perturbed to mimic the effect of a reduced accuracy in intensity coding. This was achieved by multiplying each individual complex wavelet coefficient by a random factor.[2] As a result, silence will still be silence after perturbation. The random perturbation factor ε (in dB) was chosen from a uniform distribution with zero mean and boundaries $-Pmax$ and $+Pmax$. Thus, the modulus of each individual coefficient was multiplied by a different random factor $10^{\varepsilon/20}$. After perturbation, the energy contained in each frequency band over the whole sentence was scaled to equal the original energy in this band.

The perturbation of the intensity coding was applied to the combined speech and noise signal. This probably simulates impaired auditory processing more realistically than a procedure in which speech and noise are processed separately and then combined.

B. Subjects

Twenty-five sensorineurally hearing-impaired listeners participated in the experiment. They were all native Dutch speakers, aged 24 to 70 years with a mean age of 41 years. Their intelligibility scores for monosyllabic words in quiet were at least 75% correct. In general, hearing loss was symmetrical. Thresholds in the better-hearing ears averaged over 0.5, 1, and 2 kHz (the pure-tone average, or PTA) ranged from 7 to 58 dB HL, with a mean PTA of 38 dB HL. The pure-tone, air-conduction thresholds in the better-hearing ears were at least 30 dB HL at one or more frequencies between 250 and 4000 Hz. Information about the hearing status of the hearing-impaired listeners' better-hearing ear, i.e., the ear used for data collection, is given in Table I (see also Sec. II D 1).

Twenty-two normal-hearing listeners (aged 19 to 29 years with a mean age of 22 years) served as a control group. All were native Dutch speakers. Pure-tone air-conduction thresholds of the normal-hearing listeners did not exceed 15 dB HL at any octave frequency from 250 to 4000 Hz.

C. Stimuli and apparatus

Speech material consisted of lists of 13 everyday Dutch sentences of eight to nine syllables read by a female or a male speaker (Plomp and Mimpen, 1979; Smoorenburg, 1992). The masking noise was spectrally shaped for each speaker individually according to the long-term average spectrum of all sentences.

Signals were generated by TDT (Tucker Davis Technologies) System II hardware. For frequency shaping in the intelligibility tests and in the determination of the broadband uncomfortable loudness level (see Sec. II D 1), the total signal was passed through a programable filter (TDT PF1). A finite-impulse-response (FIR) filter with 160 coefficients was implemented on the TDT-PF1 hardware. The stimuli were presented monaurally through Sony MDR-V900 headphones. To avoid the risk of cross hearing, the listener's better-hearing ear was tested. For calibration, noise levels were measured on a Brüel & Kjær type 4152 artificial ear with a flat-plate adapter. The entire experiment was controlled via a personal computer. Subjects were tested individually in a soundproof room.

D. Procedures

First, the dynamic range of each test ear was determined. Then, speech intelligibility was measured, in which stimuli were presented in the middle of the dynamic range, using frequency shaping. These tests are described below. To familiarize the subjects with the procedure, a training session preceded data collection. All conditions were measured

TABLE I. Individual hearing status of the better-hearing ear, i.e., the ear used for data collection, for the hearing-impaired listeners. The mean and standard deviation of the results for the normal-hearing listeners are presented at the bottom of the table. Given are the hearing threshold and uncomfortable loudness level for 1/3-octave bands of noise, together with age, and broadband UCL attenuation (i.e., the attenuation needed to arrive at the broadband UCL when a broadband noise burst is spectrally shaped according to the narrow-band UCL).

		Hearing threshold (dB SPL)					UCL (dB SPL)					Att.
Listener	Age (years)	0.25 kHz	0.5 kHz	1 kHz	2 kHz	4 kHz	0.25 kHz	0.5 kHz	1 kHz	2 kHz	4 kHz	(dB)
a	24	30.7	21.1	35.1	57.3	66.6	>120.0	117.7	105.7	113.3	112.7	20
b	33	36.2	33.1	41.8	43.2	48.4	>120.0	>120.0	>120.0	116.0	115.0	20
c	34	46.6	37.6	44.1	54.6	48.3	118.7	114.3	106.0	112.0	98.0	20
d	33	33.7	22.6	26.5	56.6	70.6	>120.0	>120.0	118.0	118.3	>120.0	21
e	37	41.1	25.8	22.2	54.0	58.8	>120.0	>120.0	>120.0	>120.0	>120.0	22
f	37	38.4	34.0	42.9	52.6	42.3	118.3	117.7	108.0	116.7	103.0	20
g	55	63.8	52.9	45.0	40.6	37.8	>120.0	113.3	118.0	>120.0	>120.0	15
h	25	52.4	42.8	52.2	47.4	52.3	>120.0	>120.0	117.0	116.3	109.7	20
i	37	46.9	29.5	35.6	52.2	53.2	119.7	>120.0	106.3	105.7	104.7	15
j	27	41.1	42.5	49.8	56.2	56.8	>120.0	119.7	116.3	116.0	110.3	20
k	43	45.4	44.7	39.9	62.9	61.7	>120.0	>120.0	>120.0	>120.0	>120.0	22
l	39	39.9	35.9	30.7	36.2	17.9	110.0	108.7	104.7	106.3	101.0	20
m	70	40.4	17.3	15.6	42.1	62.5	>120.0	119.7	116.0	111.7	112.7	25
n	66	37.0	24.0	25.8	50.8	64.1	>120.0	112.5	110.3	107.2	113.3	15
o	36	59.0	56.0	63.4	75.4	67.2	107.0	106.3	107.0	109.3	98.0	20
p	33	63.7	58.0	59.4	70.9	73.5	128.7	121.5	124.7	128.0	118.3	20
q	48	70.0	64.1	65.5	71.6	80.0	127.3	125.3	119.7	121.7	116.3	30
r	37	55.2	52.4	52.4	54.0	65.5	125.0	117.0	109.3	104.0	107.7	20
s	56	67.3	51.0	45.4	57.3	47.0	>120.0	119.0	108.0	113.0	102.0	20
t	32	48.8	57.7	65.8	76.4	80.9	130.0	122.2	120.2	122.8	>120.0	15
u	47	81.9	63.3	65.5	63.2	59.0	>120.0	113.3	>120.0	119.7	119.7	22
v	39	57.1	60.4	64.8	78.5	69.0	124.0	122.7	114.0	115.3	108.3	20
w	46	58.3	57.2	68.7	73.6	60.9	115.7	110.3	109.3	118.0	108.7	20
x	44	49.3	63.1	61.0	42.9	29.0	>120.0	111.5	108.3	105.7	98.0	25
y	48	29.2	20.8	7.7	22.5	53.8	>120.0	111.0	102.7	105.3	104.0	10
NH mean	22	29.8	13.6	6.7	9.5	0.3	119.6	113.7	110.3	110.2	102.4	19
NH s.d.	2.5	4.2	4.6	3.9	3.8	4.2	1.3	6.9	8.5	8.4	10.6	4

twice to determine test–retest reliability. An essential part of the distortion-sensitivity model is the comparison of the performance of an individual hearing-impaired listener to the performance of normal-hearing listeners. Therefore, for all listeners, the same order of conditions was used. In addition, for different conditions, different, fixed sentence lists were used. As a result, the order and list effects are the same for all listeners. This makes it possible to compare between the individual results of listeners without interference of order or list effects. The conditions were presented in the following order. First, a list without perturbation of the intensity coding was presented, continuing to lists processed with increasing degrees of intensity perturbation (see also Sec. II D 2). This procedure was repeated in the retest. As a result, two lists were used in each condition.

1. Dynamic range

The dynamic range of each listener was estimated by measuring the hearing threshold and the uncomfortable loudness level (UCL) for narrow bands of noise. The UCL was corrected for broadband stimulation, as described below.

Thresholds and UCLs were measured with 1/3-octave noise bands with center frequencies at 250, 500, 1000, 2000, and 4000 Hz. Hearing thresholds were measured using a Békésy tracking procedure (300-ms noise bursts; repetition rate 2.5 Hz; step size 1 dB). The measurement was ended after 11 reversals. The average of all but the first reversal level was taken as the hearing threshold. Narrow-band UCLs were measured with noise bursts presented increasing in level by 3 dB for each presentation (300-ms noise bursts; repetition rate 1.4 Hz). Listeners were asked to push a button when the noise bursts became uncomfortably loud. Then, the level of the noise burst was immediately diminished by a random amount between 21 and 30 dB, and the ascending procedure was repeated until six responses were obtained. The average of the levels at which the button was pushed was taken as the narrow-band UCL.

To correct the UCL for broadband stimulation, a 4-s broadband (0–7.8 kHz) noise burst was presented, spectrally shaped according to the narrow-band UCLs and starting 40 dB below the narrow-band UCLs. The level of the broadband noise burst was gradually increased in steps of 5 dB. After each presentation the listener was asked whether the signal was experienced as uncomfortably loud. If this was the case, the corresponding level was taken as the broadband UCL.

2. Speech intelligibility

a. Speech-reception threshold in noise for an adapted spectrum (SRTa). The speech-reception threshold (SRT, Plomp and Mimpen, 1979) was used to measure speech intelligibility. The SRT in noise is defined as the signal-to-noise ratio (SNR) at which 50% of entire sentences are reproduced correctly. In the SRT procedure a list of 13 sentences is presented. The first sentence is presented at a

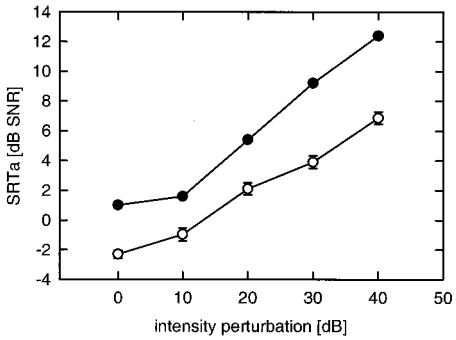


FIG. 2. Average SRTa as a function of intensity perturbation for ten normal-hearing listeners (open symbols). The error bars represent the standard error of the mean. Also, a typical example of the performance for hearing-impaired listeners is plotted (filled symbols: data for one hearing-impaired listener).

signal-to-noise ratio of −8 dB. It is repeated, each time at a 4-dB-higher level, until the listener reproduces the sentence correctly. The subsequent 12 sentences are presented only once, while the speech level is varied in an adaptive, up–down procedure with a step size of 2 dB. The SRT is calculated by taking the average signal-to-noise ratio of sentence 5 to 14. Sentence 14 is not actually presented, but its ''would-be'' level is known from the response to sentence 13. It is included in the calculation to use the information provided by the response to sentence 13.

In the speech intelligibility tests of this study, speech and noise are adapted to fit in the dynamic range of individual listeners. The adapted noise is presented in the middle of the dynamic range, while the adapted speech is presented relatively to this noise level. The adapted speech-reception threshold is called SRTa. In the SRTa tests in this study, all stimuli were bandpass filtered from 250 to 4000 Hz. The SRTa was measured as a function of intensity perturbation.

The aim of this study is to assess the effect of a reduced *auditory* intensity coding using *artificial* perturbations of the intensity coding of the speech-plus-noise stimulus. Because of the applied intensity perturbations, the auditory system is not provided with accurate intensity information. However, the applied intensity perturbations also introduce spectro-temporal fluctuations. To study the effects of distorted intensity coding, it is important to ensure that spectro-temporal effects do not dominate the speech intelligibility of hearing-impaired listeners. The larger the intensity perturbations, the bigger the spectro-temporal effects will be. Therefore, in the present study speech intelligibility was measured for intensity perturbations that only slightly affect performance. Preliminary data were collected to determine the appropriate range of intensity perturbations to apply.

The SRTa was measured as a function of the degree of intensity perturbation (0, 10, 20, 30, and 40 dB) for ten normal-hearing listeners. Figure 2 presents the results. Mean data for the normal-hearing listeners are indicated by open symbols. Error bars indicate the standard error of the mean. For the normal-hearing listeners at 10 dB of intensity perturbation, the SRTa is slightly affected (difference with no perturbation: 1.3 dB). For more severe perturbations, the SRTa increases almost linearly with perturbation, ranging from −0.9 dB when $Pmax$ is 10 dB to 6.6 dB when $Pmax$ is 40 dB.

The SRTa was also measured for 12 hearing-impaired listeners. Not all hearing-impaired listeners showed the same behavior for higher degrees of intensity distortion. Some hearing-impaired listeners seem to suffer from extreme distortions, while others do not. A typical example of the performance of a hearing-impaired listener of the former group is indicated by filled symbols. This hearing-impaired listener appears to be hardly affected by the 10 dB intensity perturbation; the decrease in performance compared with the reference condition (no perturbation) is only 0.3 dB. However, for larger degrees of intensity perturbations, speech intelligibility deteriorates more quickly than observed for the normal-hearing listeners.

Linear regression lines were fitted through the SRTa data at 10, 20, 30, and 40 dB of intensity perturbation for normal-hearing and hearing-impaired listeners. The 95-percent upper boundary of the slopes of the normal-hearing listeners is taken as the boundary between a normal sensitivity to higher degrees of distortion and higher-than-normal sensitivity. With this criterion, four of the 12 hearing-impaired listeners suffered markedly more than the normal-hearing listeners from extreme perturbations.

These data suggest that severe degrees of intensity perturbation affect some hearing-impaired listeners more than the normal-hearing listeners. This may be related to the effect of intensity perturbations on loudness perception. The dynamic range of the hearing-impaired listeners is markedly smaller than that of the normal-hearing listeners (see Table I). As a result, similar intensity perturbations do not result in similar loudness perturbation. The hearing-impaired listeners were probably subjected to higher degrees of loudness perturbation than the normal-hearing listeners. If this explanation is valid, one would expect a correlation between the dynamic range and the sensitivity to extreme distortions. A linear regression analysis on the combined data of normal-hearing and hearing-impaired listeners shows a significant correlation of −0.55 between sensitivity to extreme distortions and dynamic range at 4 kHz ($p<0.05$). Excluding the normal-hearing listeners, the correlation was not significant (correlation coefficient: −0.51, $p=0.09$).

Another cause for the higher sensitivity to extreme distortions may be the spectro-temporal fluctuations introduced by the artificial intensity distortion. These fluctuations in the combined speech and noise may result in additional masking in the temporal domain, i.e., forward and backward masking, and in the spectral domain, i.e., upward and downward spread of masking. Hearing-impaired listeners are known to suffer from excessive masking. [For review, see Moore (1995).] For some hearing-impaired listeners this excessive masking may have resulted in a higher sensitivity to extreme distortions. For large amounts of intensity perturbations these unwanted spectro-temporal by-products may even dominate the speech intelligibility of hearing-impaired listeners, causing performances for normal-hearing and hearing-impaired listeners to diverge.

To recapitulate, these preliminary data were collected to

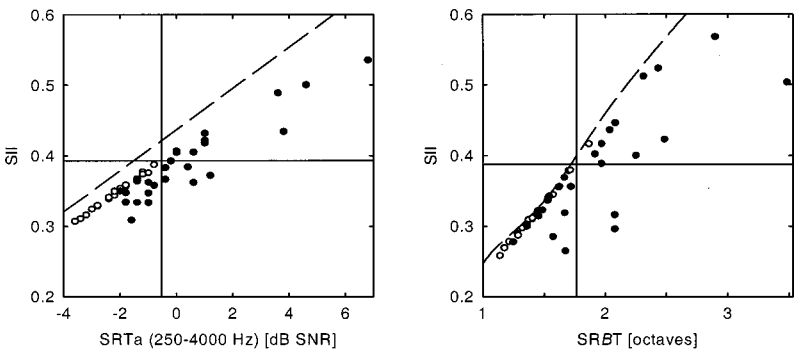


FIG. 3. Speech intelligibility index (SII) versus SRTa and SR*B*T for normal-hearing listeners (open circles) and hearing-impaired listeners (filled circles). Solid lines represent the upper boundaries of the one-tailed 95%-confidence intervals for normal-hearing listeners. Dashed lines represent the maximum SII when the audibility of the speech is not influenced by the hearing threshold, upward spread of masking, and level distortion. The vertical distance from the dashed line to the data points of the normal-hearing listeners is mainly the result of upward spread of masking.

determine the range where intensity perturbations only slightly affect performance to avoid the risk of spectro-temporal effects. A small but consistent effect on speech intelligibility was observed for intensity perturbations of 10 dB. Therefore, the range 0–10 dB was measured in the speech intelligibility experiment. As a measure for the sensitivity to the intensity distortion, the SRTa at 10 dB minus the SRTa at 0 dB is used.

b. Speech-reception bandwidth threshold (SRBT). To classify the hearing-impaired listeners into a group ''with'' and ''without'' suprathreshold deficits, the speech-reception *bandwidth* threshold (SR*B*T) was measured. The SR*B*T is a measure of speech intelligibility introduced by Noordhoek *et al.* (1999). The SR*B*T is highly sensitive for suprathreshold deficits, as is shown in a recent study of Noordhoek *et al.* (2000).

The SR*B*T procedure is similar to the SRT procedure, except that the bandwidth (center frequency 1 kHz) of speech sounds is varied instead of their levels when estimating the 50% intelligibility threshold. Complementary band-stop noise is added to the bandpass-filtered speech. Both speech and noise are presented in the middle of the listener's dynamic range.

E. Speech intelligibility index

As a measure for the quality of speech processing, the SRTa and SR*B*T data were converted to a speech intelligibility index. The speech intelligibility index (SII) (ANSI, 1997) is a physical measure of how much information of the speech is available to the listener. The SII model accounts for hearing threshold, self-masking in speech, upward spread of masking, and level distortion at high presentation levels. To calculate the SII, speech spectra, noise spectra, and hearing thresholds must be known. Therefore, sound-pressure levels of speech and noise (divided in 1/3-octave bands) were measured with the headphone positioned on a Brüel & Kjær type 4152 artificial ear with a flat-plate coupler. These levels were converted to equivalent free-field levels.

III. RESULTS AND DISCUSSION

A. Suprathreshold deficits

In the speech intelligibility tests (SRTa and SR*B*T) sounds were spectrally shaped, as described in Sec. II C, to fit in the dynamic range of individual listeners. A comparison of the results for normal-hearing and hearing-impaired listeners provides insight into the speech intelligibility performance of the hearing impaired when sounds are presented above hearing threshold.

For the normal-hearing listeners, the SRTa ranged from −3.6 to 0 dB, with an average of −2.1 dB (standard deviation 0.9 dB); for the hearing-impaired listeners, the SRTa ranged from −2.0 to 6.8 dB, with an average of 0.4 dB. The individual standard error (test–retest) averaged over all listeners was 1.1 dB.

For the normal-hearing listeners, the SR*B*T ranged from 1.14 to 1.87 octave, with an average of 1.44 octave (standard deviation 0.18 octave); for the hearing-impaired listeners, the SR*B*T ranged from 1.25 to 3.49 octave, with an average of 1.94 octave. The individual standard error (test–retest) averaged over all listeners was 0.16 octave.

The upper limit of the one-tailed 95%-confidence interval of the data for the normal-hearing listeners was used to distinguish the hearing-impaired listeners with difficulty of listening in noise. Relative to this boundary, the SRTa was elevated for 15 of the 25 hearing-impaired listeners; the SR*B*T was elevated for 13 of the 25 hearing-impaired listeners. This indicates that a substantial number of the hearing-impaired listeners has problems recognizing speech in noise, even if sounds are presented in the middle of the dynamic range of the listeners.

Speech intelligibility problems may be due to suprathreshold deficits. However, other possible explanations are inaudibility of part of the speech spectrum (if the dynamic range of a listener is very small) or high presentation levels causing extra upward spread of masking and level distortion. Therefore, to investigate the effect of suprathreshold deficits on speech intelligibility, individual SRTa and SR*B*T data were converted into SII units. An elevation of the SII values of a hearing-impaired listener compared with that of the normal-hearing listeners indicates the presence of suprathreshold deficits. The higher the SII, the more serious the speech processing deficits. Figure 3 shows the individual SII values of the SRTa and SR*B*T test for the normal-hearing listeners (open circles) and the hearing-impaired listeners (filled circles). The SII values are plotted as a function of the individual results on the two speech intelligibility tests. The upper limit of the one-tailed 95%-confidence interval of the SIIs of the normal-hearing listeners is chosen as the boundary between normal and elevated SII. This is indicated by a horizontal solid line. The boundary between normal and elevated SRTa or SR*B*T is indicated by a vertical solid line.

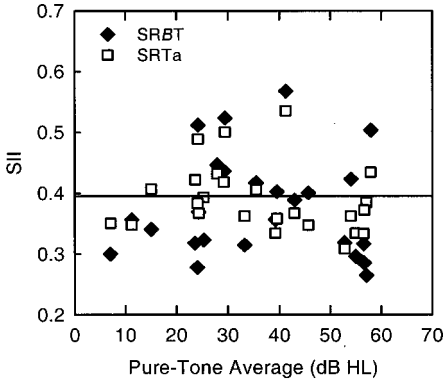


FIG. 4. Speech intelligibility index (SII) for hearing-impaired listeners on the two intelligibility tests (SRTa,SR*B*T) versus pure-tone average for better hearing ear. The horizontal line represents the upper limit of the one-tailed 95%-confidence interval for the SIIs, both the SII-SR*B*T and the SII-SRTa, of the normal-hearing listeners.

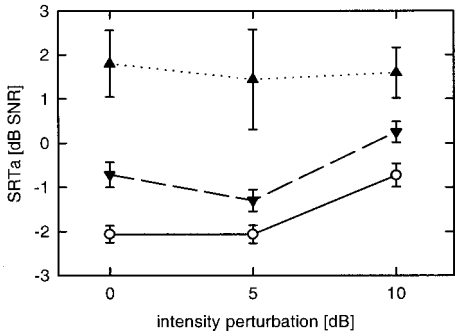


FIG. 5. SRTa as a function of intensity perturbation for normal-hearing listeners (open circles) and hearing-impaired listeners. Hearing-impaired listeners are divided into two groups: *with* (upward-pointing triangles) and *without* (downward-pointing triangles) suprathreshold deficits. Data points in conditions 0- and 10-dB intensity perturbation are for 22 normal-hearing listeners, and for 11 hearing-impaired listeners *with* suprathreshold deficits and 14 *without*. Data points in condition 5-dB intensity perturbation are for 12 normal-hearing listeners, and for 5 hearing-impaired listeners *with* suprathreshold deficits and 8 *without*. The error bars show the standard error of the mean.

The dashed lines in Fig. 3 represent the maximum SII value when the audibility of the speech is not influenced by the hearing threshold, upward spread of masking, and level distortion. Ten of the 25 hearing-impaired listeners have a higher than normal SII-SRTa and 11 have a higher than normal SII-SR*B*T; of the latter group, 7 also have a higher than normal SII-SRTa. These results show that a substantial number of hearing-impaired listeners have speech intelligibility problems because of suprathreshold deficits. In Sec. III B, hearing-impaired listeners are divided into groups *with* and *without* suprathreshold deficits.

The relation between the occurrence of suprathreshold deficits and hearing loss is illustrated in Fig. 4. SII-SR*B*T and SII-SRTa for the hearing-impaired listeners are plotted as a function of PTA. The horizontal line is the 95%-confidence limit of the SIIs (SII-SR*B*T and SII-SRTa combined) for the normal-hearing listeners. Figure 4 shows no correlation between hearing loss and SII. This indicates that some hearing-impaired listeners with only a mild hearing loss experienced hampered speech perception due to suprathreshold deficits. In contrast, some hearing-impaired listeners with severe hearing loss did not suffer from suprathreshold deficits. This finding agrees with the results of Noordhoek *et al.* (2000).

B. The distortion-sensitivity model

The distortion-sensitivity model compares speech intelligibility as a function of the degree of distortion for normal-hearing and hearing-impaired listeners. The aim is to determine whether artificial distortion relates to a suprathreshold deficit causing impaired speech perception. The hypotheses underlying this model were schematically illustrated in Fig. 1. In Fig. 5 the results of this study are depicted.

The average results of the normal-hearing listeners are represented by the open circles. The hearing-impaired listeners are divided into two groups: (1) *without* suprathreshold deficits and (2) *with* suprathreshold deficits. The division is based on the SII-SR*B*T because the SII-SR*B*T is independent of the values plotted in Fig. 5 and, as was mentioned in Sec. II D, the SR*B*T test is highly sensitive to suprathreshold deficits. This resulted in a group of 14 listeners *without* suprathreshold deficits, of which the average SRTa values are represented by downward-pointing triangles, and a group of 11 listeners *with* suprathreshold deficits, of which the average scores are represented by upward-pointing triangles. Not all listeners were tested at 5 dB of intensity perturbation. Data points in this condition are for 12 normal-hearing listeners, and for 5 hearing-impaired listeners *with* suprathreshold deficits and 8 *without*. The error bars represent the standard error of the mean.

Compared with normal-hearing listeners, hearing-impaired listeners *without* suprathreshold deficits show SRTa measures that are shifted upwards by 1 dB. No convergence of the data is observed. A Student t-test with unequal variances comparing the ''sensitivity to the distortion'' (SRTa at 10 dB minus SRTa at 0 dB) of normal-hearing and hearing-impaired listeners did not show significant convergence either. This group of hearing-impaired listeners is just as sensitive to the distortion as the normal-hearing listeners.

The performance of the hearing-impaired listeners *with* suprathreshold deficits does converge toward the performance of the normal-hearing listeners for increasing amounts of perturbation. This implies that the hearing-impaired listeners with suprathreshold deficits are less sensitive to the distortion than the normal-hearing listeners. A Student t-test with unequal variances confirmed this ($p<0.05$).

Thus, the hearing-impaired listeners *without* suprathreshold deficits are as sensitive to the intensity perturbations as the normal-hearing listeners. This is not surprising since the SII model shows that their speech intelligibility problems can be explained solely on the basis of audibility. Their suprathreshold speech processing is as good as that of normal-hearing listeners. However, the hearing-impaired listeners *with* suprathreshold deficits are less sensitive to the intensity distortion. As already noted in Sec. II D, the same degree of intensity perturbation will result in a larger degree of loudness perturbation for some hearing-impaired listeners, because they have a smaller dynamic range. For the hearing-

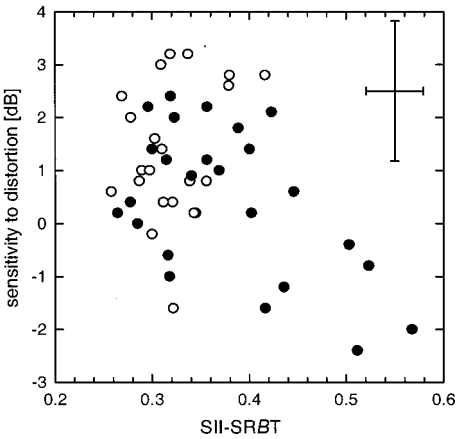


FIG. 6. Sensitivity to intensity distortion (SRTa at 10-dB intensity perturbation minus SRTa without perturbation) as a function of the SII-SR*B*T for normal-hearing listeners (open circles) and hearing-impaired listeners (filled circles). The error bars indicate plus and minus the individual standard error (test–retest) averaged over all listeners.

impaired listeners as a group, the dynamic range was roughly a factor 2 to 3 smaller than the dynamic range of the normal-hearing listeners. As a result, converting the 10-dB intensity perturbation to a loudness measure, the hearing-impaired listeners were presented with loudness perturbations a factor 2 to 3 larger than the normal-hearing listeners. Strikingly, the hearing-impaired listeners with suprathreshold deficits were less sensitive to these larger loudness perturbations. Thus, the possible conversion of the intensity factor to a loudness perturbation factor for each listener will result in a more pronounced convergence of performance for normal-hearing and hearing-impaired listeners.

In conclusion, along the lines of the distortion-sensitivity model, the observed convergence of speech intelligibility performance suggests that the artificial intensity distortion is related to the suprathreshold speech-processing problems of hearing-impaired listeners.

The division of the hearing-impaired listeners into two groups showed that hearing-impaired listeners *with* suprathreshold deficits are less sensitive to the intensity distortion than normal-hearing listeners, whereas the group *without* is just as sensitive as the normal-hearing listeners. To explore this relation between suprathreshold deficits and distortion sensitivity further, it is interesting to look at the individual results. As mentioned in Sec. II D, the difference in SRTa between 10- and 0-dB intensity perturbation was taken as a measure of the individual sensitivity to the distortion. In Fig. 6, this sensitivity is plotted as a function of the SII-SR*B*T. Open symbols represent the data for normal-hearing listeners, the filled symbols those for hearing-impaired listeners. The individual standard error of the SII-SR*B*T (test and retest) averaged over all listeners was 0.029; the individual standard error of the sensitivity to the distortion (test and retest) averaged over all listeners was 1.3 dB. The error bars indicate plus and minus 1 individual standard error. For some listeners, sensitivity to the distortion was negative, suggesting that performance improved when intensity perturbation was applied. However, the negative sensitivity may be explained by factors influencing speech intelligibility other than the distortion of intensity information. First, *measured* sensitivity may be negative as a result of measurement error. Second, order or list effects may push the sensitivity to negative values. As indicated before, to allow comparison between listeners, subjects listened to the same lists in each condition, in the same order. As a result, order and list effects may be present in the data across tests. Due to the order/list effect, the points of Fig. 6 may be shifted downwards systematically, pushing sensitivity to negative values for listeners with a very low sensitivity. However, since the order/list effects are the same for all listeners, they do not influence the correlation.

Even though the individual standard errors are large, a trend can be observed in Fig. 6: a decrease in sensitivity as the SII-SR*B*T increases. A linear regression analysis on the data of the hearing-impaired listeners showed a significant correlation of -0.54 ($p<0.05$). From this it may be concluded that the higher the SII (more severe speech-processing deficits) the less sensitive the hearing-impaired listeners are to the intensity distortion.

To summarize the above, the results provide evidence that speech intelligibility for the group of hearing-impaired listeners with suprathreshold deficits is affected less by intensity perturbation than for normal-hearing listeners. Moreover, looking at the individual results of all hearing-impaired listeners, the sensitivity to the intensity perturbation correlates negatively with the SII-SR*B*T. In other words, the larger the effect of suprathreshold deficits on speech processing, the less sensitive a hearing-impaired listener is to intensity perturbation. Under the distortion-sensitivity model, this implies that distortion of intensity coding relates to the effects of suprathreshold deficits underlying the poor speech intelligibility in noise. The underlying deficit might be poor auditory intensity coding.

How do these results relate to studies examining the effect of intensity distortions for normal-hearing listeners in quiet? In quiet, speech intelligibility is very robust for intensity distortions (Licklider, 1946). Even when the spectral resolution is severely limited, only a few quantization levels are enough to make speech intelligible (Loizou *et al.*, 1999; Shannon *et al.*, 2001). Thus, even severe *artificial* distortion of intensity information (a few quantization levels) does not affect speech-perception performance in quiet for normal-hearing listeners. Therefore, even if hearing-impaired listeners suffer from severely distorted *auditory* processing of intensity information, this deficit will probably not affect their speech-perception performance in quiet. Thus, if hearing-impaired listeners have problems perceiving speech in quiet, these problems must be due to factors other than distorted *auditory* intensity coding. Noordhoek *et al.* (2000) showed that these problems in quiet can be explained solely by reduced audibility.

The present study did not examine speech intelligibility problems in quiet, but examined problems in noise. In noise, the situation is different. For noisy speech at critical signal-to-noise ratios, artificial intensity distortion is much more detrimental to intelligibility (Licklider, 1946; the present study). To give an indication of the degree of distortion used in the present study, an intensity distortion of about 10 dB

was just noticeable to the listeners. This degree of distortion already reduced speech intelligibility by about 1 dB. This implicates that relatively minor distortions in auditory intensity coding may already result in speech intelligibility problems in noise. The present study provides some indications that, indeed, reduced auditory intensity coding may contribute to the speech intelligibility problems in noise experienced by hearing-impaired listeners.

IV. SUMMARY AND CONCLUSIONS

In this study, speech intelligibility was measured as a function of intensity perturbation of speech-plus-noise stimuli. The sensitivity to the distortion by hearing-impaired listeners was compared with that by normal-hearing listeners. The data on the speech intelligibility tests were converted to SII values. An elevation of the SII of a hearing-impaired listener, as compared with the SIIs of normal-hearing listeners, indicates a suprathreshold speech-processing deficit; the higher the SII, the more speech intelligibility is affected by suprathreshold deficits. The hearing-impaired listeners were divided into two groups on the basis of their SII-SR*B*T: a group *with* and a group *without* suprathreshold deficits. This classification did not relate to hearing loss: some listeners with a severe hearing loss did not show suprathreshold deficits, whereas some listeners with a mild hearing loss showed severe suprathreshold deficits. Data revealed that hearing-impaired listeners *without* suprathreshold deficits were just as sensitive to intensity perturbations as normal-hearing listeners; hearing-impaired listeners *with* suprathreshold deficits appeared to be less sensitive to intensity perturbations than normal-hearing listeners. The convergence for increasing degrees of intensity perturbation suggests that the applied artificial distortion relates to the suprathreshold deficit causing speech intelligibility problems. A small but significant correlation between the SII-SR*B*T of hearing-impaired listeners and the sensitivity to the intensity distortion was observed. It is concluded that intensity perturbation may partly characterize the effect of a suprathreshold deficit causing a reduced speech intelligibility in noise. The underlying hearing deficit may be a reduced acuity of auditory intensity coding.

ACKNOWLEDGMENTS

This work was supported by The Netherlands Organization for Scientific Research (NWO). We would like to thank Ingrid Noordhoek for the stimulating discussions with respect to the SII model. Thanks are due to the reviewers for their useful comments.

[1]A function $f(t)$ has compact support if it is zero outside the interval $T_0 < t < T_0 + \Delta T$.

[2]An alternative approach to simulate the effect of reduced auditory intensity coding is to manipulate the quantization levels of the wavelet coefficients. As already mentioned, in this study another approach was taken: the modulus of each wavelet coefficient was multiplied by a random factor. The different approaches reflect a possibly different nature of the auditory deficit. The approach of this study reflects the idea of noisiness of the intensity coding of the auditory system. The alternative approach reflects the idea of auditory coding with low intensity ''selectivity.'' Both approaches result in an artificial distortion of intensity information. Therefore, for both approaches, convergence is to be expected if distorted auditory coding of intensity information reduces speech perception.

Allen, J. B. (**1977**). ''Short term spectral analysis, synthesis, and modification by discrete Fourier transform,'' IEEE Trans. Acoust., Speech, Signal Process. **25**, 235–238.

Allen, J. B., and Rabiner, L. R. (**1977**). ''A unified approach to short-time Fourier analysis and synthesis,'' Proc. IEEE **65**, 1558–1564.

ANSI (**1997**). ANSI S3.5-1997, ''American National Standard Methods for Calculation of the Speech Intelligibility Index'' (American National Standards Institute, New York).

Buus, S., Florentine, M., and Zwicker, T. (**1995**). ''Psychometric functions for level discrimination in cochlearly impaired and normal listeners with equivalent-threshold masking,'' J. Acoust. Soc. Am. **98**, 853–861.

Florentine, M., Reed, C. M., Rabinowitz, W. M., Braida, L. D., and Durlach, N. I. (**1993**). ''Intensity perception. XIV. Intensity discrimination in listeners with sensorineural hearing loss,'' J. Acoust. Soc. Am. **94**, 2575–2586.

Gabor, D. (**1947**). ''Acoustical quanta and the theory of hearing,'' Nature (London) **159**, 591–594.

Houtgast, T. (**1995**). ''Psycho-acoustics and speech recognition of the hearing impaired,'' in *Proceedings of the European Conference on Audiology* (Noordwijkerhout, The Netherlands), pp. 165–169.

Licklider, J. C. R. (**1946**). ''Effects of amplitude distortion upon the intelligibility of speech,'' J. Acoust. Soc. Am. **18**, 429–434.

Loizou, P. C., Dorman, M., and Tu, Z. (**1999**). ''On the number of channels needed to understand speech,'' J. Acoust. Soc. Am. **106**, 2097–2103.

Moore, B. C. J. (**1995**). *Perceptual Consequences of Cochlear Damage* (University Press, Oxford).

Moore, B. C. J. (**1996**). ''Perceptual consequences of cochlear hearing loss and their implications for the design of hearing aids,'' Ear Hear. **17**, 133–161.

Noordhoek, I. M., Houtgast, T., and Festen, J. M. (**1999**). ''Measuring the threshold for speech reception by adaptive variation of the signal bandwidth. I. Normal-hearing listeners,'' J. Acoust. Soc. Am. **105**, 2895–2902.

Noordhoek, I. M., Houtgast, T., and Festen, J. M. (**2000**). ''Measuring the threshold for speech-reception by adaptive variation of the signal bandwidth. II. Hearing-impaired listeners,'' J. Acoust. Soc. Am. **107**, 1685–1696.

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

Rioul, O., and Vetterli, M. (**1991**). ''Wavelets and signal processing,'' IEEE Signal Process. Mag. **8**(4), October, 14–38.

Scharf, B. (**1970**). ''Critical bands,'' in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, New York), Vol. 1, pp. 157–202.

Shannon, R. V., Fu, Q., Wang, X., Galvin, III, J., and Wygonski, J. (**2001**). ''Critical cues for auditory pattern recognition in speech: Implications for cochlear implant speech processor design,'' in *Physiological and Psychophysical Bases of Auditory Function (Proceedings of the 12th International Symposium on Hearing)*, edited by D. J. Breebaart, A. J. M. Houtsma, A. Kohlrausch, V. F. Prijs, and R. Schoonhoven (Shaker BV, Maastricht, the Netherlands), pp. 500–508.

Smoorenburg, G. F. (**1992**). ''Speech reception in quiet and in noisy conditions by individuals with noise-induced hearing loss in relation to their tone audiogram,'' J. Acoust. Soc. Am. **91**, 421–437.

van Schijndel, N. H., Houtgast, T., and Festen, J. M. (**1999**). ''Intensity discrimination of Gaussian-windowed tones: Indications for the shape of the auditory frequency–time window,'' J. Acoust. Soc. Am. **105**, 3425–3435.

Age-related changes in detecting a mistuned harmonic

Claude Alain[a)]
Rotman Research Institute, Baycrest Centre for Geriatric Care, 3560 Bathurst Street, Toronto, Ontario M6A 2E1, Canada and Department of Psychology, University of Toronto, Mississauga, Ontario L5L 1C6, Canada

Kelly L. McDonald and Jodi M. Ostroff
Rotman Research Institute, Baycrest Centre for Geriatric Care, 3560 Bathurst Street, Toronto, Ontario M6A 2E1, Canada

Bruce Schneider
Department of Psychology, University of Toronto, Mississauga, Ontario L5L 1C6, Canada

(Received 20 September 2000; revised 16 January 2001; accepted 26 February 2001)

The effects of age on discriminating simultaneous sounds were investigated by comparing the hearing threshold in detecting a mistuned harmonic in young, middle-aged, and older adults. The stimuli were complex sounds containing multiple harmonics, one of which could be ''mistuned'' so that it was no longer an integer multiple of the fundamental. Older adults had higher thresholds than middle-aged or young adults. The effect of age was greater for short than for long duration sounds and remained even after controlling for hearing sensitivity. The results are consistent with an age-related decline in parsing simultaneous auditory events, which may contribute to the speech perception difficulties in the elderly. © *2001 Acoustical Society of America.*
[DOI: 10.1121/1.1367243]

PACS numbers: 43.71.Lz, 43.71.Pc, 43.66.Fe, 43.66.Lj [MRL]

I. INTRODUCTION

A common complaint of older adults is that they have difficulties understanding speech, especially when the competing signals are voices rather than homogeneous background noise (e.g., Duquesnoy, 1983; Prosser *et al.*, 1991) or in the presence of reverberation (Gordon-Salant and Fitzgibbons, 1993). Although sensory neural hearing loss is highly correlated with the speech reception threshold (Humes and Lisa, 1990; van Rooij and Plomp, 1990; van Rooij *et al.*, 1989), it is unlikely that peripheral factors alone can account for all of the observed changes in auditory perception that accompany aging. Speech perception problems experienced by older adults are somewhat greater than would be expected based solely upon reduction in auditory sensitivity (Marshall, 1981) and they remain significant even after controlling for age-related differences in hearing sensitivity (Divenyi and Haupt, 1997a). In addition, some individuals with near-normal sensitivity exhibit speech discrimination problems, while others with hearing loss receive little benefit from hearing aids despite restoration of near-normal sensitivity levels. It is also unclear how age-related changes in hearing sensitivity would account for difficulties in understanding speech, which primarily occur in the presence of noise or competing speech signals.

Difficulties in understanding speech in the presence of noise or speech signals may result from age-related changes in central auditory functions that are crucial for the perceptual segregation of co-occurring sounds (e.g., Divenyi and Haupt, 1997a). To follow a conversation during a cocktail party, a listener must be able to separate the acoustic elements that correspond to the conversation of interest from those that correspond to other ''competing'' conversations. This requires parsing the acoustic wave into two or more sound sources (i.e., auditory objects) and integrating the ongoing acoustic information from a particular source (''auditory scene analysis''). The perceptual segregation of sounds occurring simultaneously is enhanced by differences in frequency regions and spatial origin while fusion is favored by synchronous changes in amplitude or frequency that occur in different spectral regions [for reviews see Bregman (1990) and Hartmann (1988)]. For example, sounds that have similar stimulus onset, intensity, amplitude modulation, and/or frequency periodicity are more likely to be perceived as originating from the same source than sounds that differ in onset, intensity, and/or frequency periodicity. Thus deficits in processing either one or more of these acoustic cues could have dramatic consequences on the perception of complex auditory signals by hindering the perceiver's ability to adequately separate the spectral components of the speech event from the background noise (figure-ground segregation), thereby making the speech more difficult to understand.

The present study focused on age-related changes in one of the cues that lead to segregation of sounds. Frequencies that are harmonically related to one another are usually perceived as originating from a single source. However, if one of these frequencies departs from a harmonic relationship with the others, it is likely to be perceived as emanating from a second source. Thus age-related deficits in the ability to detect mistuned harmonics may deleteriously affect an older adult's ability to segregate concurrent auditory events. In the present study, we used a paradigm similar to that of Moore *et al.* (1985) who used loudness-matched complex sounds that either had all tuned harmonics or one mistuned harmonic. Participants were presented with signals; one that had

a)Electronic mail: calain@rotman-baycrest.on.ca

tuned harmonics (or ''partials'') and one with a mistuned harmonic (i.e., the harmonic was no longer an integer multiple of the fundamental). They indicated by pressing a button which signal contained the mistuned harmonic. Participants were better at detecting inharmonicity for higher than for lower harmonics and for long than for short duration signals (see also Hartmann *et al.*, 1990; Moore *et al.*, 1985).

In the present study, participants were presented with short and long duration signals to explore whether age effects on detecting a mistuned harmonic would interact with sound duration. Given that older adults have deficits in temporal processing (e.g., Schneider and Hamstra, 1999), one might expect to see greater age differences in thresholds for short than for long duration signals. We also included a group of middle-aged adults to characterize in more detail age-related changes in concurrent sound segregation. Lastly, we examined the relationship between the ability to distinguish a mistuned harmonic and the ability to recognize speech in a multi-talker noise background. Because the voiced components of speech emanating from a single talker will be harmonically related, age-related declines in the ability to detect mistuned harmonics may lead to difficulties in segregating voices when two or more people are speaking simultaneously. In particular, it may lead to difficulties in attending to a single voice against a multi-talker background. Previous results have shown that older adults have higher thresholds than younger adults for correctly reporting the last word of a sentence in the Speech Perception in Noise (SPIN) test, a test in which the noise background is a multi-talker babble (e.g., Gordon-Salant and Fitzgibbons, 1993; Hutchinson, 1989; Pichora-Fuller *et al.*, 1995). In the present study, we compared the ability to detect inharmonicity with performance on the SPIN test in a subset of participants.

II. METHOD

A. Participants

Ten young (mean age=22 years, s.d.=2; range=19–25; 4 men), 10 middle-aged (mean age=43 years, s.d.=6; range =35–49; 5 men), and 10 older (mean age=72 years, s.d.=6; range=65–82; 4 men) adults participated in the study. All participants had pure-tone thresholds less than or equal to 30 dB HL in the range 250 to 3000 Hz in the tested ear. Each participant signed a consent form according to University of Toronto guidelines. The young adults were recruited from local colleges whereas middle-aged and older adults were recruited from the community and local volunteer groups.

B. Stimuli and task

The complex sounds had a fundamental frequency of 200 Hz and were composed of the first ten harmonics at equal levels. Stimulus duration was either 100 or 400 ms including 10-ms rise/fall time. Stimuli were generated digitally with 16-bit resolution and a sampling rate of 20 kHz, passed through a digital-to-analogue converter, and then low passed filtered at 10 kHz using an anti-aliasing filter (Tucker Davis Technology, Gainesville, FL). Stimuli were presented at 80 dB SPL to the left ear through TDH-49 headphones.

Thresholds were defined as the percentage of mistuning required to detect the signal. They were determined using a two-interval, two alternative forced-choice procedure with a three-down and one-up rule that estimates the 79% correct point on the psychometric function (Levitt, 1971). Each trial consisted of two stimuli separated by an inter-stimulus interval of 500 ms. The interval containing the mistuned harmonic was either the first or the second at random (probability=0.5 for each interval) and the mistuning from trial to trial was either upward or downward at random. The listener's task was to select which of the two stimuli contained the mistuned harmonic. After each response, visual feedback was provided on the performance. Inter-trial interval was controlled by the participant whose response initiated the presentation of the next trial. After three consecutive correct responses, the mistuning of the partial was decreased by a factor of 1.4. After each incorrect response, the mistuning was increased by the same factor. Threshold was determined as the mean level for the final 8 of 12 reversals.

Three different harmonics, the second, fifth, and eighth, were mistuned in separate blocks. Findings from a pilot study showed that the mistuning of the second harmonic was particularly difficult to detect: To ensure that all participants could detect the second mistuned harmonic, the initial value of mistuning was set at 12%. For the fifth and eighth harmonics, the initial mistuning was set at 10%. For all participants, thresholds were first estimated for the long duration signal in separate blocks of trials, but the order of presentation of the second, fifth, or eighth harmonic was counterbalanced between participants. The thresholds reported are based on three estimates per condition. Participants were tested individually in a single-walled sound-attenuating chamber. They participated in three testing sessions, each lasting approximately 60 min.

C. Speech in noise SPIN test

In a second experiment, a subset of participants performed the SPIN test. They were presented with sentences embedded in babble and asked to identify the last word of the sentence. Participants were also required to indicate whether the last word was a high or low context word (e.g., for the sentence ''I ate dinner at the RESTAURANT'' participants would have to repeat the word ''restaurant'' and indicate that it is a high context word). The SPIN threshold was estimated by calculating the signal/noise ratio at which participants correctly identified 50% of the low context words embedded in babble. The SPIN test always followed the experiment on perception of the mistuned harmonic. Seven young (mean age=22 years, s.d.=2; range=19–24; 2 men), four middle-aged (mean age=42 years, s.d.=6; range =37–48; 2 men), and six older adults (mean age=72 years, s.d.=6; range=66–82; 3 men) participated in the second experiment. All participants were native English speakers.

III. RESULTS

Table I summarizes the audiometric thresholds for young, middle-aged, and older participants. ANOVA with age group as a between subjects factor and frequency (i.e.,

TABLE I. Group mean audiometric (HL) threshold (and standard deviation) in young, middle-aged, and older participants.

Groups	Frequency 250	500	1000	2000	3000
Young	2.0(5.9)	1.5(3.4)	1.0(4.6)	−0.5(5.5)	−0.5(6.4)
Middle-Aged	7.0(5.9)	5.5(4.4)	9.0(7.0)	10.0(9.4)	7.0(8.2)
Older	9.5(7.3)	7.5(4.9)	9.0(7.8)	14.0(9.1)	14.5(7.6)

threshold at 250, 500, 1000, 2000, and 3000 Hz) as a within subjects factor yielded a main effect of group, $F(2,27) = 12.91$, $p < 0.001$. Young adults had lower thresholds than middle-aged or older adults, $p < 0.01$ in both cases. There was no significant difference in audiometric threshold between middle-aged and older adults.

Figures 1(A) and (B) show the group mean threshold for the short and long duration stimuli, respectively. The threshold for detecting the mistuned harmonic is expressed as a percent of mistuning relative to the ''tuned'' harmonic. Overall, the threshold estimates were lower for long than for short duration stimuli, $F(1,27) = 113.68$, $p < 0.001$. Participants were also better at detecting inharmonicity when the fifth or eighth harmonic was mistuned than when the second harmonic was mistuned, $F(2,54) = 12.08$, $p < 0.001$. The interaction between stimulus duration and harmonic number tended toward significance, $F(2,54) = 2.88$, $p = 0.07$. The thresholds for the second harmonic were less affected by signal duration than the thresholds for the fifth or eighth harmonics.

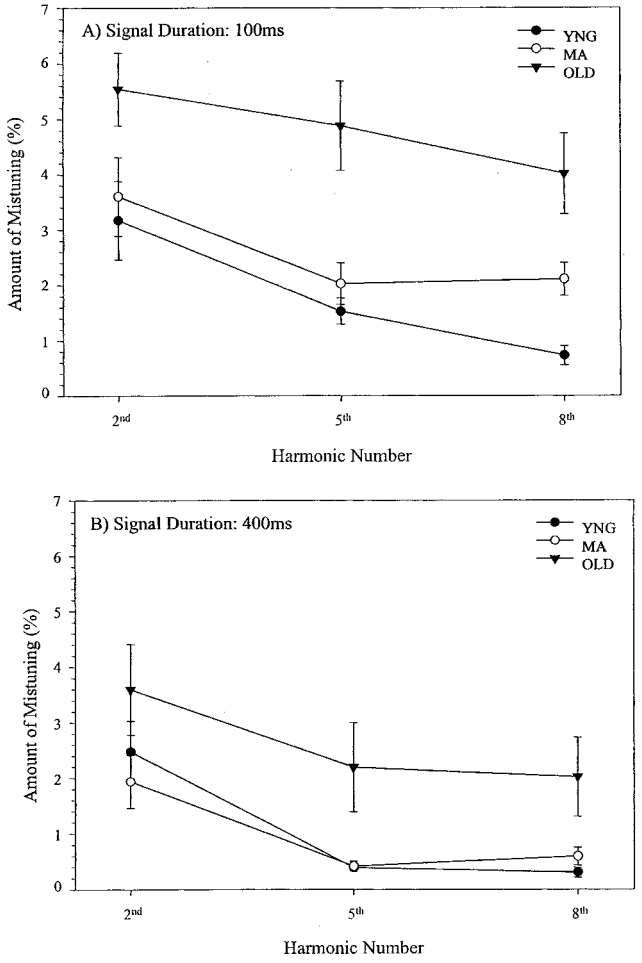


FIG. 1. (A) Thresholds for detecting inharmonicity for the 100-ms duration sounds in young, middle-aged, and older adults. (B) Thresholds for detecting inharmonicity for the 400-ms duration sounds in young, middle-aged, and older adults. The threshold for inharmonicity is expressed as percent mistuning of the harmonic concerned.

A mixed design ANOVA with group as a between-subject factor and stimulus duration and harmonic number as within-subject factors yielded a main effect of group, $F(2,27) = 7.16$, $p < 0.01$. Pairwise comparisons revealed that older adults had elevated thresholds compared with middle-aged and young adults, $p < 0.01$ in both cases. The group $\times$harmonic number interaction was not significant nor was the interaction between group, stimulus duration, and harmonic number. However, the group$\times$stimulus duration interaction was significant, $F(2,27) = 6.95$, $p < 0.01$, reflecting greater age effects for short than for long duration stimuli.

These observed age effects on detecting a mistuned harmonic could be partly attributed to the elevated audiometric thresholds in older adults. However, this cannot account for the performance differences between older and middle-aged listeners since there were no significant differences in audiometric thresholds (Table I) between these two groups. To show that there is an age effect after controlling for audiometric differences, we used the average audiometric threshold for 250, 500, 1000, 2000, and 3000 Hz, and the slopes of the audiogram for these five frequencies as covariates in an analysis of covariance. (To obtain the slope we fit straight lines to individual plots of dB HL versus log frequency.) The analysis of covariance yielded a main group effect, $F(2,26) = 4.31$ and 6.82, for pure-tone threshold average and slope, respectively, $p < 0.05$ in both cases. These analyses indicate that the age effects on detecting a mistuned harmonic remained even after controlling for age differences in hearing sensitivity.

Group differences in detecting a mistuned harmonic may be partly due to differences in learning the task. In order to examine the role of learning in detecting the mistuned harmonic, the thresholds measured at the first, second, and third sessions were compared between the three groups. The ANOVA yielded a main effect of session, $F(2,54) = 12.39$, $p < 0.001$, indicating that thresholds decreased significantly from the first to the third session. All pairwise comparisons were significant, $p < 0.05$. Although the group$\times$session interaction was not significant, the group$\times$session$\times$harmonic number interaction tended toward significance, $F(8,108) = 2.17$, $p = 0.056$. For the second harmonic, young and middle-aged adults improved their performance as a function of session whereas older adults did not. For the fifth and eighth harmonic, the effect of practice was similar in young, middle-aged, and older adults.

In each group a subset of participants also completed the SPIN test. The signal-to-noise ratios in dB to detect 50% of the low context words were −0.29, 1.77, and 1.97 in young, middle-aged, and older adults, respectively. Older adults had a higher SPIN threshold than younger adults, $t(11) = 2.57$, $p < 0.05$. There was no group difference between young and middle-aged or between middle-aged and older adults. In this subgroup of participants, there was a main effect of age on

audiometric thresholds, $F(2,14)=16.17$, $p<0.001$. Pairwise comparisons revealed that young adults had lower audiometric thresholds than middle-aged or older adults, $p<0.01$ in both cases. There was no significant difference in audiometric threshold between middle-aged and older adults. We also compared the threshold for detecting a mistuned harmonic for this sub-group of participants. As in the whole-group analysis, we found a main effect of signal duration, $F(1,14)=49.42$, $p<0.001$, with lower thresholds for long than for short stimuli. The main group effect was significant, $F(2,14)=8.49$, $p<0.01$, revealing higher thresholds in older adults compared to young and middle-aged adults. There were no differences between young and middle-aged adults. The interaction between group and signal duration was also significant, $F(2,14)=6.62$, $p<0.01$, reflecting greater age differences for short than for long signals. In this small sample of participants, the correlation between the threshold for detecting a mistuned harmonic and SPIN threshold was not significant.

IV. DISCUSSION

One of the most common complaints of older adults is that they have difficulty following a conversation, especially in complex listening situations where more than one sound source is present at a time. This speech perception problem in older adults has been attributed to several factors including changes in hearing sensitivity, in central auditory processing, and in higher cognitive functions such as memory and attention. Here, we show that older adults have difficulties in detecting a mistuned harmonic. Such difficulties may contribute to the speech perception problems experienced by older adults because perception of a mistuned harmonic, like speech signals embedded in babble or noise, depends on the ability to parse auditory events based on their spectral pattern. Deficits in processing spectral patterns may have dramatic consequences on the perception of concurrent auditory signals by making the perceiver unable to adequately separate the spectral components of the speech event from the background noise (figure-ground segregation), thereby making the perception of speech more difficult.

Previous research showed that the speech perception problems in older adults are prevalent in the presence of multiple sound sources (e.g., Divenyi and Haupt, 1997a). In the present study, older adults showed higher SPIN thresholds than young adults, consistent with numerous studies showing age-related declines in speech reception threshold (e.g., Bergman, 1980; Frisina and Frisina, 1997; Gelfand *et al.,* 1988). The increased threshold for detecting a mistuned harmonic in individuals who also showed deficits in speech reception threshold is consistent with the hypothesis that age-related changes in processes critical for sound segregation may contribute to the observed speech perception problems in older adults. Yet, the threshold in detecting a mistuned harmonic did not significantly correlate with the SPIN threshold. Although this may seem counterintuitive, the small sample size did not allow sufficient power to draw any conclusions or rule out a potential relationship between age-related decline in concurrent sound segregation and the SPIN threshold.

The age difference in detecting a mistuned harmonic was present even after controlling for audiometric threshold. This indicates that age-related increases in detecting inharmonicity do not depend solely on peripheral factors but must also involve an age-related decline in central auditory functions. The fact that middle-aged adults did not show elevated thresholds in detecting mistuning despite higher audiometric thresholds relative to young adults provides further evidence for an age-related decline in central auditory functions. If the threshold for detecting a mistuned harmonic was solely related to hearing sensitivity then performance should have been similar in middle-aged and older adults. The fact that middle-aged adults performed as well as the young adults suggests that in middle-aged adults top-down controlled processes may compensate for impoverished sensory input. Although our results point toward age-related changes in central auditory functions, one cannot exclude the possibility that age-related changes in detecting inharmonicity could also be mediated by peripheral factors that are distinct from those mediating absolute sensitivity.

Several factors could contribute to age-related changes in detecting a mistuned harmonic. One possibility is that older adults have broader auditory filters than young and middle-aged adults. Previous studies have shown that sensorineural hearing loss is often associated with broader auditory filters than normal (e.g., Glasberg and Moore, 1986; Lutman *et al.,* 1991). One consequence of having broader auditory filters would be a reduction in the ability to resolve the individual harmonics of the complex sound thereby resulting in higher thresholds for detecting a mistuned harmonic. However, it remains equivocal whether mild hearing impairment and/or age per se affect frequency selectivity. For example, the decrease in frequency resolution is usually minimal for pure-tone thresholds ranging from 0 to 30 dB (Glasberg and Moore, 1986; Lutman *et al.,* 1991). There is also evidence suggesting that age, independent of hearing loss, does not impair frequency resolution (Peters and Moore, 1992; Sommers and Gehr, 1998; Sommers and Humes, 1993), and audiometric thresholds in young and older listeners do not always correlate with performance on tasks assessing frequency selectivity (Divenyi and Haupf, 1997b). The relation between frequency selectivity and auditory discrimination is also ambiguous. Performance on discriminating complex sounds is only weakly correlated with frequency selectivity (Moore and Peters, 1992). Furthermore, young and old listeners may differ on frequency discrimination tasks and yet have similar auditory filter widths (Moore and Peters, 1992).

Given that the perception of complex sounds requires the analyses of rapid ongoing fluctuations in sound frequency and intensity, an age-related decline in temporal resolution could contribute to the elevated thresholds in detecting mistuned harmonics in older adults. For example, older adults typically showed elevated thresholds in detecting a silent gap within a signal (e.g., Moore *et al.,* 1992; Schneider and Hamstra, 1999; Snell, 1997). This age-related effect is greater when short rather than long duration markers are used (Schneider and Hamstra, 1999). Similarly, greater age effects in frequency difference limens have been re-

ported for short than for long duration signals (Cranford and Stream, 1991). In the present study, the age difference in detecting a mistuned harmonic was greater for short than for long duration signals, which is consistent with the proposal that older adults have impaired temporal processing. The age-related decline in temporal resolution may be related to inaccurate phase-locking of auditory neurons. There is evidence from the animal literature showing that older mice have a reduced number of neurons in the inferior colliculi that respond to brief gaps compared to younger mice (Walton *et al.*, 1998). The recovery of the phase-locking response was also of a lower magnitude in the older mice (Walton *et al.*, 1998). Inaccuracy in the phase-locking of auditory neurons with aging would impair the ability to accurately extract the frequency of each harmonic thereby making the detection of inharmonicity more difficult.

The age difference in detecting a mistuned harmonic could also be caused by age-related slowing in perceptual and cognitive processing (Salthouse, 1996). According to the processing-speed theory, age-related decline in perceptual and cognitive tasks can be accounted for by the reduced speed with which various perceptual and cognitive operations can be performed in the available time. In the present context, this suggests that the speed with which the harmonic template could be processed by the auditory system would be reduced for older adults. With short duration signals, older individuals had little time to extract frequency periodicity, leaving the perceptual analysis partly incomplete. When long duration signals were used, these individuals had time to complete the perceptual analysis, thereby improving their performance in detecting the mistuned harmonic.

In all participants, the threshold for detecting inharmonicity decreased when both signal duration and harmonic number increased. This is consistent with other studies involving well-trained listeners (e.g., Moore *et al.*, 1985). Differences in threshold between low and high harmonic number are thought to reflect differences in sensitivity for detecting roughness or beating. Previous work has shown that as the frequency of a mistuned harmonic increased, the rate of fluctuation increased, making the beats and roughness effective cues, especially at long signal durations.

As previously observed in the literature, the threshold for the lowest harmonic was less affected by signal duration than the middle or highest harmonic. This is consistent with the proposal that different processes may mediate processing of low and high harmonics (e.g., Moore *et al.*, 1985; Moore and Glasberg, 1986). In the present study, the effect of age did not interact with both duration and harmonic number, suggesting that these different processes were similarly affected by aging.

All participants improved their performance over the course of the experiment. Although the older adults seemed to benefit most from repeated sessions, the effect of learning/practice on detecting inharmonicity was similar in all three age groups. It remains possible that with additional training older adults may asymptote at a similar threshold value as middle-aged or younger adults. However, it is unclear what amount of training would be required to reach such asymptote values. In the present study, the short duration signals were always presented after the long duration signals. Therefore, the lower threshold for short duration signals cannot be accounted for by a practice effect. Further research is needed to evaluate the effects of training on detecting a mistuned harmonic and to assess whether training modulates the age-related difference in detecting inharmonicity observed in the present study.

ACKNOWLEDGMENTS

We gratefully thank Jane Carey and Melinda Ciffolillo for technical assistance, and Patricia van Roon and David Shore for helpful comments on earlier versions of the manuscript. This research was funded by a grant from the Canadian Institute for Health Research to C.A.

Bergman, M. (**1980**). *Aging and the Perception of Speech* (University Park Press, Baltimore).

Bregman, A. S. (**1990**). *Auditory Scene Analysis. The Perceptual Organization of Sounds* (The MIT Press, London).

Cranford, J. L., and Stream, R. W. (**1991**). ''Discrimination of short duration tones by elderly subjects,'' J. Gerontol. **46**, 37–41.

Divenyi, P. L., and Haupt, K. M. (**1997a**). ''Audiological correlates of speech understanding deficits in elderly listeners with mild-to-moderate hearing loss. I. Age and lateral asymmetry effects,'' Ear Hear. **18**, 42–61.

Divenyi, P. L., and Haupt, K. M. (**1997b**). ''Audiological correlates of speech understanding deficits in elderly listeners with mild-to-moderate hearing loss. II. Correlation analysis,'' Ear Hear. **18**, 100–113.

Duquesnoy, A. J. (**1983**). ''The intelligibility of sentences in quiet and in noise in aged listeners,'' J. Acoust. Soc. Am. **74**, 1136–1144.

Frisina, D. R., and Frisina, R. D. (**1997**). ''Speech recognition in noise and presbycusis: Relations to possible neural mechanisms,'' Hear. Res. **106**, 95–104.

Gelfand, S. A., Ross, L., and Miller, S. (**1988**). ''Sentence reception in noise from one versus two sources: Effects of aging and hearing loss,'' J. Acoust. Soc. Am. **83**, 248–256.

Glasberg, B. R., and Moore, B. C. (**1986**). ''Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments,'' J. Acoust. Soc. Am. **79**, 1020–1033.

Gordon-Salant, S., and Fitzgibbons, P. J. (**1993**). ''Temporal factors and speech recognition performance in young and elderly listeners,'' J. Speech Hear. Res. **36**, 1276–1285.

Hartmann, W. M. (**1988**). ''Pitch, perception and the segregation and integration of auditory entities,'' in *Auditory Function: Neurobiological Bases of Hearing*, edited by G. M. Edelman, W. E. Gall, and W. M. Cowan (Wiley, New York), pp. 623–645.

Hartmann, W. M., McAdams, S., and Smith, B. K. (**1990**). ''Hearing a mistuned harmonic in an otherwise periodic complex tone,'' J. Acoust. Soc. Am. **88**, 1712–1724.

Humes, L. E., and Lisa, R. (**1990**). ''Speech-recognition difficulties of the hearing-impaired elderly: the contribution of audibility,'' J. Speech Hear. Res. **33**, 726–735.

Hutchinson, K. M. (**1989**). ''Influence of sentence context on speech perception in young and older adults,'' J. Gerontol. **44**, 36–44.

Levitt, H. (**1971**). ''Transformed up-down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467.

Lutman, M. E., Gatehouse, S., and Worthington, A. G. (**1991**). ''Frequency resolution as a function of hearing threshold level and age,'' J. Acoust. Soc. Am. **89**, 320–328.

Marshall, L. (**1981**). ''Auditory processing in aging listeners,'' J. Speech Hear. Res. **46**, 226–240.

Moore, B. C., and Peters, R. W. (**1992**). ''Pitch discrimination and phase sensitivity in young and elderly subjects and its relationship to frequency selectivity,'' J. Acoust. Soc. Am. **91**, 2881–2893.

Moore, B. C., Peters, R. W., and Glasberg, B. R. (**1985**). ''Thresholds for the detection of inharmonicity in complex tones,'' J. Acoust. Soc. Am. **77**, 1861–1867.

Moore, B. C. J., and Glasberg, B. R. (**1986**). ''Thresholds for hearing mistuned partials as separate tones in harmonic complexes,'' J. Acoust. Soc. Am. **80**, 479–483.
Moore, B. C. J., Peters, R. W., and Glasberg, B. R. (**1992**). ''Detection of temporal gaps in sinusoids by elderly subjects with and without hearing loss,'' J. Acoust. Soc. Am. **92**, 1923–1932.
Peters, R. W., and Moore, B. C. J. (**1992**). ''Auditory filter shapes at low center frequencies in young and elderly hearing impaired subjects,'' J. Acoust. Soc. Am. **91**, 256–266.
Pichora-Fuller, M. K., Schneider, B. A., and Daneman, M. (**1995**). ''How young and old adults listen to and remember speech in noise,'' J. Acoust. Soc. Am. **97**, 593–608.
Prosser, S., Turrini, M., and Arslan, R. (**1991**). ''Effects of different noises on speech intelligibility in the elderly,'' Acta Oto-Laryngol., Suppl. **476**, 136–142.
Salthouse, T. A. (**1996**). ''The processing-speed theory of adult age differences in cognition,'' Psychol. Rev. **103**, 403–428.
Schneider, B. A., and Hamstra, S. J. (**1999**). ''Gap detection thresholds as a function of tonal duration for younger and older listeners,'' J. Acoust. Soc. Am. **106**, 371–380.
Snell, K. B. (**1997**). ''Age-related changes in temporal gap detection,'' J. Acoust. Soc. Am. **101**, 2214–2220.
Sommers, M. S., and Gehr, S. E. (**1998**). ''Auditory suppression and frequency selectivity in older and younger adults,'' J. Acoust. Soc. Am. **103**, 1067–1074.
Sommers, M. S., and Humes, L. E. (**1993**). ''Auditory filter shapes in normal-hearing, noise-masked normal, and elderly listeners [published erratum appears in J. Acoust. Soc. Am. **94**, 2449–2450 (1993)],'' J. Acoust. Soc. Am. **93**, 2903–2914.
van Rooij, J. C. G. M., and Plomp, R. (**1990**). ''Auditive and cognitive factors in speech perception by elderly listeners. II. Multivariate analysis,'' J. Acoust. Soc. Am. **88**, 2611–2624.
van Rooij, J. C. G. M., Plomp, R., and Orlebeke, J. F. (**1989**). ''Auditive and cognitive factors in speech perception by elderly listeners. I. Development of test battery,'' J. Acoust. Soc. Am. **86**, 1294–1309.
Walton, J. P., Frisina, R. D., and O'Neill, W. E. (**1998**). ''Age-related alteration in processing of temporal sound features in the auditory midbrain of the CBA mouse,'' J. Neurosci. **18**, 2764–2776.

Acoustic-phonetic features for the automatic classification of fricatives

Ahmed M. Abdelatty Ali[a] and Jan Van der Spiegel
Department of Electrical Engineering, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6390

Paul Mueller
Corticon, Inc., 155 Hughes Road, King of Prussia, Pennsylvania 19406

(Received 7 September 1999; revised 7 August 2000; accepted 26 January 2001)

In this article, the acoustic-phonetic characteristics of the American English fricative consonants are investigated from the automatic classification standpoint. The features studied in the literature are evaluated and new features are proposed. To test the value of the extracted features, a statistically guided, knowledge-based, acoustic-phonetic system for the automatic classification of fricatives in speaker-independent continuous speech is proposed. The system uses an auditory-based front-end processing system and incorporates new algorithms for the extraction and manipulation of the acoustic-phonetic features that proved to be rich in their information content. Classification experiments are performed using hard-decision algorithms on fricatives extracted from the TIMIT database continuous speech of 60 speakers (not used in the design/training process) from seven different dialects of American English. An accuracy of 93% is obtained for voicing detection, 91% for place of articulation detection, and 87% for the overall classification of fricatives. [DOI: 10.1121/1.1357814]

PACS numbers: 43.72.Ne [DOS]

I. INTRODUCTION

The fricatives constitute the largest set of consonants in the English language which has nine standard fricative consonants, namely: the voiceless /f/ as in ''feel'', /th/ as in ''thought'', /s/ as in ''seal'' and /sh/ as in ''she'' and their voiced cognates /v/ as in ''van'', /dh/ as in ''then'', /z/ as in ''zeal'' and /zh/ as in ''azure''. The ninth fricative is the /h/ which has unique semi-vowel-like characteristics. These consonants can be distinguished by English-speaking listeners in identical phonetic contexts, which indicates that the features needed for such discrimination should be in the acoustical signal.

Several past studies have investigated such features. Hughes and Halle (1956), Harris (1958), Strevens (1960), Fant (1960), and Heinz and Stevens (1961) are examples of some of the earliest studies which characterize the fricative consonants. Using perceptual experiments on synthetic speech, analysis of spoken syllables, and primitive recognition experiments, those studies provide us with much data on the acoustic characteristics of fricatives. However, they have been largely qualitative in nature and relied on a small set of stimuli produced in few vowel contexts usually by a single speaker.

Later studies (Fant, 1973; Cole and Cooper, 1975; McCasland, 1979a, b; Guerlekian, 1981; Manrique and Massone, 1981; Baum and Blumstein, 1987; Behrens and Blumstein, 1988a, b; Crystal and House, 1988; Jongman, 1989; Stevens *et al.*, 1992; Hedrick and Ohde, 1993; Borden *et al.*, 1994) have added to our knowledge about fricatives. However, except for a few studies, the acoustic characteristics that exist in the literature are qualitative, relational, and speaker dependent. They characterize the fricatives well from the articulation standpoint of separate syllables. But they are not enough to perform accurate phonetic distinctions in speaker-independent continuous speech.

In the following sections, we discuss the articulation, perception, and acoustic-phonetic features of fricatives. Then, the results of our research on the characteristics of fricatives, and their automatic classification, are discussed.

II. ARTICULATION OF FRICATIVES

Fricatives are nonresonant consonants (like stops and affricates). They do not have a clear formant structure as that of vowels, semivowels, and nasals. They are produced by forming a constriction in the vocal tract that causes turbulence (frication) which in turn generates noiselike sound. This noise makes it possible to hear the sounds whether there is phonation or not. Therefore, using a single articulation, it is possible to create two different phonemes (called cognates), one voiced and the other unvoiced (voiceless) (Borden *et al.*, 1994).

The presence of a source at the constriction creates a resonant cavity in front of it as well as a back cavity behind the constriction (Fant, 1960, 1973). The presence of this back cavity creates poles/zeros in the transfer function of the vocal tract filter in addition to the poles formed by the front cavity. These additional poles/zeros could be bound or free depending on whether the two cavities are decoupled or not. The positions of these created zeros decide the shape of the fricative spectrum and give it its distinctive nature, as we will see later.

[a]Currently with Texas Instruments, Inc. Electronic mail: ahm@ee.upenn.edu

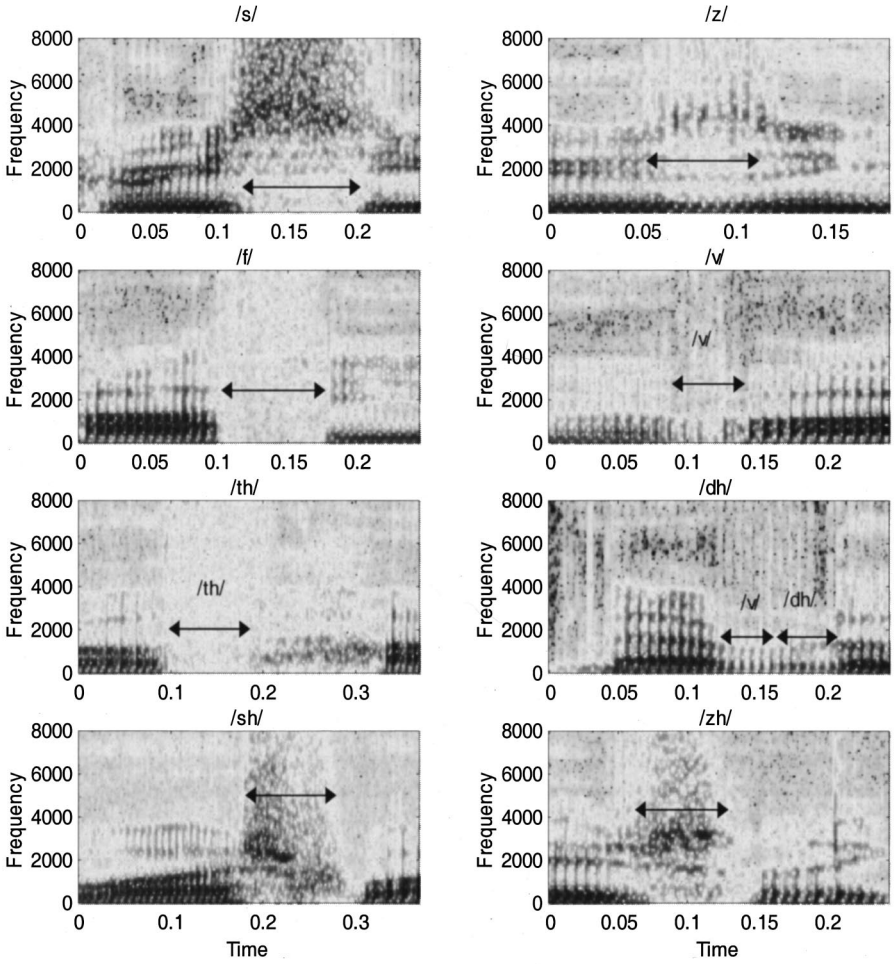


FIG. 1. Fourier-based uniform spectrograms of various fricatives. The similarities between the labiodentals (/f/ and /v/) and the linguadentals (/th/ and /dh/) are clear.

Compressing the continuous flow of air through a constriction is formed by closely approximating two articulators. In English, there are four main places of fricative articulation: labiodental, linguadental, alveolar, and palatal (Borden *et al.*, 1994). The labiodental (labial) fricatives, /f/ (voiceless) and /v/ (voiced), are formed by bringing the lower lip close to the inferior edges of the upper central incisors. The linguadental fricatives, /th/ (voiceless) and /dh/ (voiced), are formed by approximating the tip of the tongue with the upper incisors. The alveolars /s, z/ and palatals /sh, zh/ are called sibilants due to their hissing, shushing quality.

The alveolars are produced by forming a constriction between the alveolar ridge and the tongue. The constriction could be between the tip of the tongue and the ridge (a high-point /s/) or between the blade of the tongue and the ridge (a low-point /s/). Moreover, a groove is often formed along the tongue midline to channel the air stream. This is accomplished by raising the lateral edges of the tongue to the medial edges of the upper teeth while depressing the center of the tongue. The palatals are produced by forming a constriction further back, in the post alveolar area. The midline groove is slightly wider than for the alveolars and the lips may be somewhat rounded and protruded (Borden *et al.*, 1994).

The last fricative is the aspirate /h/, which is sometimes considered to be a semivowel or a voiceless vowel. The constriction is located in the larynx at the glottis. It is usually voiceless but can be voiced if embedded between voiced segments. The only required movement is the approximation of the vocal folds. The vocal tract takes the shape of the following vowel.

III. PERCEPTION OF FRICATIVES

Unlike their articulation, the perception and acoustic features of fricatives, as well as of other phonemes, are not adequately understood. There is some information about this subject in the literature based on numerous production and perceptual experiments performed over the past fifty years. Unfortunately, this information is still not enough for automatic speech recognition applications.

The most important acoustic cue that distinguishes the fricative manner of articulation is the extended period of noise (Borden *et al.*, 1994). This feature is usually straightforward to detect on a spectrogram especially for voiceless fricatives, which are usually longer in duration and stronger in intensity than their voiced cognates. Due to the presence of two resonant cavities, anterior and posterior to the constriction, zeros are introduced in the vocal tract transfer function whose locations depend on the place of articulation. Voiced fricatives display clearer formant structure than voiceless fricatives and are weaker in intensity. Fricatives in general, especially voiceless ones, are characterized by a high-frequency noise in the spectrum which is strongly em-

phasized compared to an unvoiced vowellike sound of similar articulation. Examples of spectrograms of different fricatives (excluding the /h/) are shown in Fig. 1. Due to its unique vowellike characteristics, the fricative /h/ is not included in any further discussion in this article.

The place of articulation of the fricatives is usually indicated, according to the literature (Borden *et al.*, 1994), by two primary cues: the spectrum and the intensity of frication. The sibilants (i.e., /s/, /z/, /sh/, and /zh/) are characterized by relatively steep, high-frequency spectral peaks, while the nonsibilants (i.e., /f/, /v/, /th/ and /dh/) are characterized by relatively flat and wider band spectra. Furthermore, the sibilants are distinguished as being alveolar (/s/ and /z/) or palatal (/sh/ and /zh/) on the basis of the location of their lowest spectral peak which is around 4 kHz for the alveolars and 2.5 kHz for the palatals for a typical male speaker. On the other hand, nonsibilant fricatives, which include the linguadentals (/th/ and /dh/) and the labiodentals (/f/ and /v/), are not easily distinguished because the dental constriction that characterizes them causes their spectra to be very similar.

The intensity differences also help to cue the place of articulation. Sibilants are marked by relatively high intensity levels compared to nonsibilants. This is due to the absence of an appreciable resonating cavity in front of the dental constrictions of the nonsibilants. There has been a debate in the literature about the significance of the intensity in detecting the place of articulation, with some researchers confirming its importance while others negating it. This will be discussed in detail in the following section.

Another cue that is thought to be contributing to the place of articulation detection is the formant transitions between the noise (friction) part and the vocalic part. There was also some debate about the importance of this cue, but the general agreement is that it is much less important for fricatives than it is for stop consonants. However, they are thought to play some role in discriminating between labiodentals and linguadentals (Borden *et al.*, 1994). Linguadentals are characterized by a slightly greater emphasis of the noise at high frequencies compared to labiodentals. However, the difference is so small that it becomes totally undetectable in multiple speaker experiments. Therefore, it is thought that formant transitions play a role in such discrimination. However, the nature of this role, as well as its mere existence, needs further evaluation.

To identify voicing, the most important cue is the presence of phonation (voicing). However, in the case of fricative consonants, this is a sufficient but not necessary condition. Listeners can make reliable judgments about the voicing class of a fricative based upon its duration (sometimes relative to the vowel preceding it), even in the absence of phonation. It is also possible that listeners use the intensity difference between the frication of voiced and voiceless fricatives to detect the voicing property (Borden *et al.*, 1994).

In spite of the usefulness of the above-mentioned features, which are summarized in Fig. 2, in understanding the acoustic properties of fricatives from the production standpoint, they are of limited help from the automatic recognition research standpoint. This is due to the following reasons:

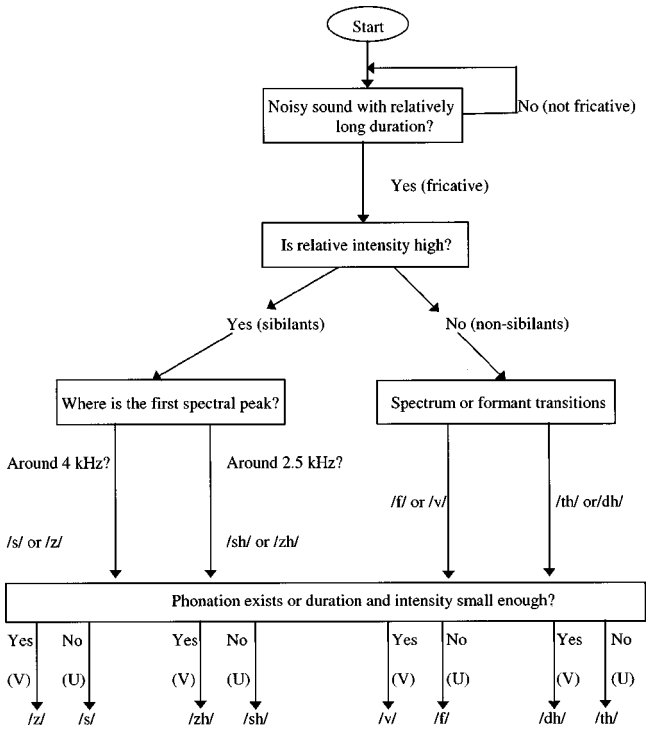


FIG. 2. Acoustic-phonetic characterization of fricatives (Borden *et al.*, 1994).

(i) Some of those features are qualitative and abstract in nature. Properties like flat/steep spectra are not easily usable in an automatic recognition experiment. They need to be translated to physical features (measurable parameters) in order to be fully utilized. Different translations yield different results.

(ii) Most of the features are relational and not absolute. In speech recognition, we can not deal with relational parameters and we need absolute descriptions. For example, when we say that nonsibilants are often weaker in intensity than sibilants, an obvious question arises: Is there a general, speaker- and context-independent threshold value above which the fricative is sibilant and below which it is nonsibilant, or is this valid only for fricatives in the same conditions? If only the latter is true, then this feature is not useful for automatic recognition because we do not have, in such application, a reference for comparison. However, if the former is true, then we will need an absolute threshold above which we categorize the fricative as strong (i.e., sibilant) and below which it is weak (i.e., nonsibilant). Unfortunately, such kind of recognition-oriented description is not yet available in the literature.

(iii) The mentioned features are usually deduced using experiments on a single speaker saying controlled utterances (syllables). Such conditions are acceptable for characterizing the acoustics of speech production and synthesis. But for continuous speech recognition, it is extremely vital to use multiple-speaker natural speech experiments to have credible conclusions.

(iv) As we will see later, a significant number of contradictions exist in the literature regarding the values and roles of various features.

IV. ACOUSTIC-PHONETIC FEATURES OF FRICATIVES

In this section, the acoustic-phonetic features of fricatives are investigated. Their strengths, weaknesses, variability, consistency, speaker dependence, and use are studied using acoustic and spectrogram analysis, statistical analysis, and automatic classification experiments. The results of these experiments are given in the next section for each feature separately. Emphasis will be on features that proved to be consistent, invariant, and useful from the recognition standpoint. It should be noted that the term ''feature'' in this article is used in its ''pattern recognition'' connotation to refer to any extractable property or measurable parameter that can be used as an input to a classifier. This is different from the ''linguistic'' connotation of the term (also known as ''distinctive features''), which usually refers to abstract properties that may, or may not, correspond to measurable parameters.

A. Duration and voicing detection

Duration has been investigated by several researchers as a cue for detecting the place of articulation as well as voicing (Harris, 1958; Cole and Cooper, 1975; Manrique and Massone, 1981; Baum and Blumstein, 1987; Crystal and House, 1988; Behrens and Blumstein, 1988b; Jongman, 1989; Stevens and Blumstein, 1992). Several studies have reported variations in the mean durations of fricatives with the place of articulation. Those variations, however, could not be used as an invariant cue (neither primary nor secondary) for the automatic recognition of the place of articulation. This is because of the significant overlaps in the distributions around the mean values as well as the inconsistent results from various experiments. It was also apparent from our own experiments that the duration was neither a consistent nor an invariant feature to rely on to play any role in the place of articulation detection.

What is more important is the use of duration as a cue in voicing detection. It is known that the duration of voiced fricatives is *often* less than that of voiceless ones. However, there was a debate in the literature regarding the consistency and reliability of this feature for voicing detection. Cole and Cooper (1975) reported strong dependence of the voicing perception on the duration. Decreasing the duration of voiceless fricatives in their perceptual experiments resulted in a change in their perception to their voiced cognates. Moreover, in the case of the presence of a preceding vowel, they found that the ratio between the fricative duration and the vowel duration was the decisive feature (and not just the fricative duration). They concluded that the fricative duration (relative to the preceding vowel if it existed in a final position) was a sufficient and primary cue for voiced–voiceless distinctions of fricatives.

Jongman (1989), however, reached an opposite conclusion in his perceptual experiments where he found the voicing characteristic of fricatives to be unaffected by duration variations. He found that listeners needed at least 30-ms duration to detect /z/ and /sh/, 50 ms to detect /f/, /s/ and /v/, and the whole fricative duration to detect /th/ and /dh/. Below those durations the place of articulation perceived was affected. The voicing detection, however, was not so strongly dependent on duration as the place of articulation was. This was a result that contradicted Cole and Cooper's.

In a different kind of experiment, Manrique and Massone (1981) examined the duration of the frication noise in Spanish voiced and voiceless fricatives and reached the conclusion that there was a significant difference between the mean durations of voiced and voiceless fricatives and that there was ''little or no overlap between duration ranges.'' This was later negated by Baum and Blumstein (1987) in their experiments who found significant overlaps in the duration ranges, which render such a feature useless. They attributed the discrepancy in the results to the fact that Manrique and Massone's experiments were performed on fricatives that were not homorganic and therefore the large difference in their reported durations could be due to other reasons besides the difference in the voicing characteristic.

The above-mentioned contradictions are an example of the conflicting information that exists in the literature of speech research and complicates the task of integrating speech knowledge in the current automatic speech recognition (ASR) systems. This underscores the need to perform acoustic-phonetic and statistical analyses to evaluate the features from the recognition standpoint.

The first observation regarding the use of duration as a primary cue for voicing detection is based more on logic than on experimentation. Voiced fricatives, like their voiceless cognates, are continuants. This means that, unlike stop consonants, they can be articulated as long as the speaker desires. Even if the voiced fricatives tend *naturally* to be shorter than voiceless ones, there is no inherent reason why a speaker can not produce them too long if he wants too. Therefore, it defies common sense to expect the absolute duration to be a primary cue in discriminating between voiced and voiceless fricatives. A speaker can say a /z/ as long as he wants (like in the word ''buzzzzzz'') and it will still be recognized as a /z/ and not as an /s/ (its voiceless cognate) as long as the speaker has invested enough phonation in its articulation.

This speculative conclusion has been verified by our experiments in which we were unable to find a threshold duration that consistently divides the fricatives into voiced and voiceless. Such conclusion, however, could lead to another possibility if pursued further. It is known that the presence of phonation (voicing), which is manifested by the low-frequency energy in the spectrograms, is a sufficient, but not necessary, condition for voicing. In most fricatives, this feature is not present at least for part of the phoneme (Stevens *et al.*, 1992). This feature, however, is necessary for the correct recognition of a peculiarly long voiced fricative, as mentioned above in our speculation. This leads us to the conclusion that what really matters may not be the duration of the fricative as a whole, but rather of its voiceless (unphonated) part. This means that a voiced fricative can be as long as we want as long as it is fully voiced, but when the voicing information is not present, there is an upper limit on the dura-

tion of the unvoiced portion above which it is perceived as voiceless.

This feature [i.e., the duration of the unvoiced portion (DUP)] was tested in our experiments and gave a 93% correct response as will be explained in more detail in the next section. This result is in agreement with the result obtained by Stevens *et al.* (1992), who achieved similar accuracy using the same feature for nonfinal fricatives.

To summarize, we found out that the duration of the fricative *per se* does not play a role in either the place of articulation or voicing detection. What really matters is the duration of the unvoiced portion (DUP) of the fricative (normalized with respect to the duration of the preceding vowel if fricative is in utterance-absolute-final position). A threshold DUP was statistically found which, interestingly, was independent of the place of articulation. Above that threshold the fricative can be classified as voiceless and below it as voiced. This result agrees with previous research and is in line with the acoustic-phonetics literature.

B. Relative amplitude (intensity) of frication

This is a feature that is thought to help discriminate between sibilants and nonsibilants (Hedrick and Ohde, 1993). Here ''relative'' means compared to the nearest vowel or to the maximum energy of the utterance. However, there is a disagreement about its role in such task.

Several researchers (Strevens, 1960; Heinz and Stevens, 1961; Guerlekian, 1981; Behrens and Blumstein, 1988a; Hedrick and Ohde, 1993; Borden *et al.*, 1994) observed the difference in intensity between sibilants and nonsibilants and commented on the possibility of using the amplitude as a cue for such discrimination. Heinz and Stevens (1961) also found that nonsibilants were never correctly perceived by the listeners in their perceptual experiments, when the intensity was too high, no matter how the spectrum looked like.

Guerlekian (1981) confirmed this finding by his results which showed that low amplitude of the noise relative to the vowel was perceived as /f/ while high amplitude was perceived as /s/ when the fricative noise portion and the transition portion were in conflict. However, Behrens and Blumstein (1988a) argued that although the amplitude did play a role, it did not seem to be a major or primary one. They performed their experiments using stimuli with noise and transition portions that were compatible instead of being conflicting as in Guerlekian's experiment. They found that decreasing the amplitude of sibilants caused them to be recognized as nonsibilants, but the opposite was not true. Had the amplitude been a primary feature, the opposite should have been true and the effect in both ways should have been significant, which was not the case. This led them to conclude that what could play a role might not be the relative total amplitude but rather the relative amplitude in a certain frequency band.

Hedrick and Ohde (1993) tested this hypothesis and confirmed it. They found that the relative amplitude in a certain band provided information for the perception of the place of articulation. However, their choice of the band was rather arbitrary, chosen as the third formant (F3) region for the /s/-/sh/ contrast and the fifth formant (F5) region for the /s/-/th/ contrast. Such a variable and phoneme-dependent choice, however, is not practical for recognition applications.

The above-mentioned results indicate clearly that the relative amplitude may be playing a role in sibilant/nonsibilant discrimination. However, it may not be the only feature used in performing such a task. Apparently the spectral shape is needed as well, either in the form of using the relative amplitude in a certain band or by detecting the flatness (diffuse) property of the spectrum of nonsibilants that is opposed to the steep spectrum of the sibilants.

Due to the difficulty and variability in choosing the band of relative amplitude as described by Hedrick and Ohde, we preferred to use the flatness property. To do that we used a quantity that we called the maximum normalized spectral slope (MNSS), which is described later in detail. This quantity incorporates the flatness property together with the relative amplitude in one value. It proved to be effective in detecting nonsibilants and gave better results than the relative amplitude alone.

Therefore, it was found that the relative amplitude does play a role in discriminating between sibilants and nonsibilants. However, such discrimination can not be performed accurately based on this feature alone and the spectral flatness is needed in addition to achieve satisfactory performance.

The relative intensity was also proposed by some researchers as a feature to discriminate between voiced and voiceless fricatives. Despite the fact that voiced fricatives tend to be weaker than their voiceless cognates, our statistical analysis has proved that this is not a reliable or invariant feature to perform such a task.

C. The spectral shape

Researchers agree on the importance of the spectral shape in the place of articulation detection of fricatives. Several studies had been carried out to characterize the spectral shape of different fricatives. However, the results of those studies were usually qualitative, relational, and speaker dependent.

For example, researchers agree that the flatness of the spectrum is a property that characterizes nonsibilants as opposed to the steep spectra of sibilants. However, flatness is an abstract, speaker-dependent feature that was not meant in this description literally. The nonsibilants do not really have a *flat* spectrum; their spectra are just flatter than those of the sibilants. Moreover, a ''flat'' spectrum obtained using a linearly distributed filter bank (as in traditional spectrograms) will not be flat when we use an auditory-based system like the Bark-scaled filter bank (used in our experiments). This is due to the nonlinear filter distribution that tends to emphasize the high-frequency regions.

A measurable parameter was needed to describe the flatness property and give good speaker-independent recognition results. The previously mentioned MNSS was found to satisfy this requirement as will be discussed in detail in the following sections.

The other problem was the discrimination among the sibilants themselves (i.e., /s/ and /z/ on one hand and /sh/ and /zh/ on the other). The spectral description is usually given in

the literature in terms of peak positions or the lowest peak position. However, there are different processing techniques to extract a peak position, which will yield different results. Also, the peak position could be measured absolutely or relative to the neighboring vowel, again yielding different results. Such considerations, together with the corresponding results, will be discussed later.

D. Formant transitions

The role of the second and third formant transitions (F2 and F3) in detecting the place of articulation were investigated by several researchers (Harris, 1958; Heinz and Stevens, 1961). The common conclusion was that the spectral shape was a much more important cue for the place of articulation than formant transitions. And that if they played any role, it would be in discriminating between labiodentals and linguadentals. Apart from this, perceptual experiments have shown that formant transitions are not a primary cue in the place of articulation detection.

The first formant (F1) transition, on the other hand, is used sometimes as a cue for detecting voicing. Stevens *et al.* (1992) have confirmed this fact. In our experiments, such transition was also investigated and proved to be a sufficient cue (but not necessary) for detecting voicing in fricatives. Its role, however, is only secondary to that of phonation and to the DUP. This will be discussed in more detail in the next section.

V. CLASSIFICATION EXPERIMENTS

In this section, some experiments on the acoustic characteristics of fricatives are described. The features used and the results obtained will be examined in detail and compared to previous knowledge that exists in the literature and was discussed previously. Features are proposed based on their presence and consistency during the training (design) phase in a small portion of the TIMIT database (five male and five female speakers). The classification experiments are performed on 1100 fricatives extracted from the continuous speech of 60 different speakers with seven different dialects of American English from the TIMIT database.

For a fricative to be included in the training or testing phase, it has to be correctly transcribed. Fricatives that were not perceived correctly (i.e., as transcribed in the TIMIT database) by a human listener (namely the first author), before performing the experiments, were excluded. The resulting database included fricatives in various contexts and positions with strong assimilation and coarticulation effects. The frequency of occurrence of each fricative in the training and testing data is similar to its frequency in the TIMIT database and in American English speech (Edwards, 1992). This way, the evaluation of the system will be skewed by the unequal frequencies of occurrence of various fricatives and hence will reflect its performance on real American English speech. Each fricative is manually extracted from the database using the TIMIT database transcription boundaries. It is then divided into three equal segments, and the extracted spectral parameters are averaged over the duration of the middle segment.

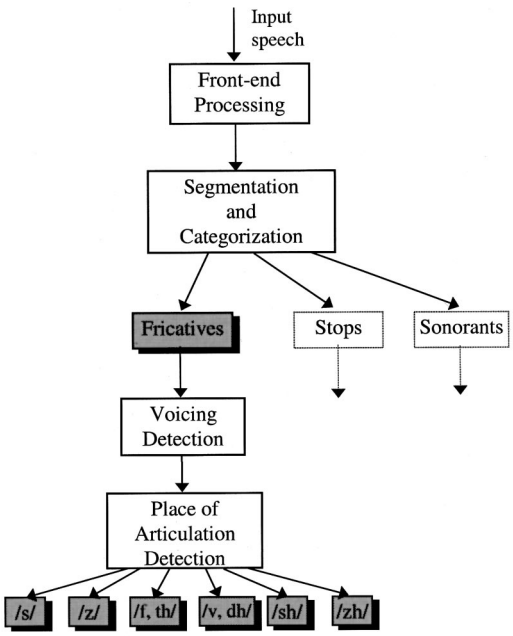


FIG. 3. Block diagram of the fricative recognition system used.

A block diagram of the system is shown in Fig. 3. The front-end signal processing system that is used in our experiments is a biologically oriented filter-bank system (Ali, 1999). It is based on the system developed by Seneff and described in detail in Seneff (1985, 1988). The system is modified to enhance its formant extraction ability and decrease its response to the pitch harmonics (Ali, 1999). The block diagram is given in Fig. 4. The system incorporates some of the auditory effects like critical-band filtering, nonlinear effects, saturation, short-term adaptation, forward-masking, and synchrony detection.

The filter bank used is a bank of 36 critical-band filters (Bark scale) with the distribution given by Zwicker (1961). It is preceded by a 20-dB/decade high-frequency preemphasis. This and other biologically oriented distributions have proved to yield better results in ASR applications especially in the presence of noise (Davis and Morenalstein, 1980; Hunt and Lefebrre, 1989; Jankowski *et al.*, 1995; Seneff, 1985).

This system gives two outputs, the mean-rate and the ALSD (average localized synchrony detector) output. The generalized synchrony detector (GSD) developed by Seneff (1988) computes an auto-correlation-like output by finding the soft-limited ratio of the expected (averaged) value of the sum and difference of the output of each filter and a delayed version of it. The ALSD uses the spacial averaging of neighboring filters' GSDs in order to suppress spurious peaks (pitch harmonics) and enhance the formants (Ali *et al.*, 2000;

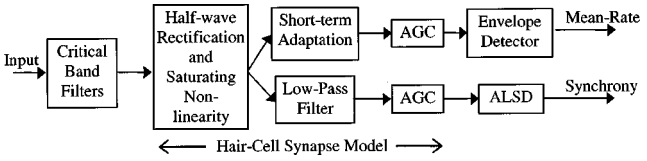


FIG. 4. Block diagram of an auditory-based front-end system.

Ali, 1999). In our experiments, both the mean-rate and ALSD outputs are utilized. The ALSD is used for its superb formant extraction ability, higher response to periodic signals, and higher immunity to noise, while the mean-rate is used for its higher sensitivity and better ability to describe the overall spectral shape. This is in agreement with our knowledge about auditory neurobiology. It is known that the average response and the temporal pattern of the neural firings play complementary roles which are very much similar to the roles used in our experiments and mentioned above (Delgutte and Kiang, 1984a, b, c, d; Delgutte, 1986, 1980, 1984).

A. Acoustic parameters used

Several features are described in the literature, which are useful in characterizing and recognizing fricatives. In this section, we will discuss the measurable parameters that proved to be successful and invariant enough to capture the previously mentioned abstract features (properties).

1. Duration and voicing detection

As mentioned before, the duration of the unvoiced portion (DUP) of the fricative is used as a voicing detection feature. To measure this duration, we have to develop a method to detect the absence or presence of voicing in the signal and hence detect the start and end points. Notice that for a totally voiced fricative, phonation never stops. Therefore, our method should detect its presence and hence indicate a DUP of zero.

Voicing is manifested in the output by low-frequency energy which is characteristic of voiced sounds, especially vowels and semi-vowels. Two methods are developed to detect such energy:

(1) The total energy of the lowest nine filters (less than 1 kHz) in the ALSD output. Such output is readily normalized in the ALSD processing and is sensitive to periodicity. Call it LOWG.
(2) The ratio between the low-frequency (below 1.5 kHz) and high-frequency (above 3 kHz) energies in the mean-rate output, normalized with respect to the maximum energy in the utterance. Call it LOWE.

The advantage of using two quantities instead of one is that they tend to complement each other. LOWG is very effective in detecting periodicity, as opposed to noise, and therefore it gives accurate results. However, due to the presence of spontaneous response, it is not sensitive enough and might miss weak energy phonation. LOWE, on the other hand, makes up for this by its relatively higher sensitivity. Each of the two quantities has its own threshold, which is found statistically using histogram analysis. If *either* of them exceeds its threshold, then phonation is assumed present. Therefore the DUP is the period where both quantities (LOWG and LOWE) are below their respective thresholds (LOWGTH and LOWETH). It is clear that, as mentioned before, for a fully phonated fricative, DUP is equal to zero. If DUP is below a certain, statistically determined, threshold (DUPTH), then the fricative is assumed voiced, otherwise it is voiceless. The algorithm is described in Fig. 5.

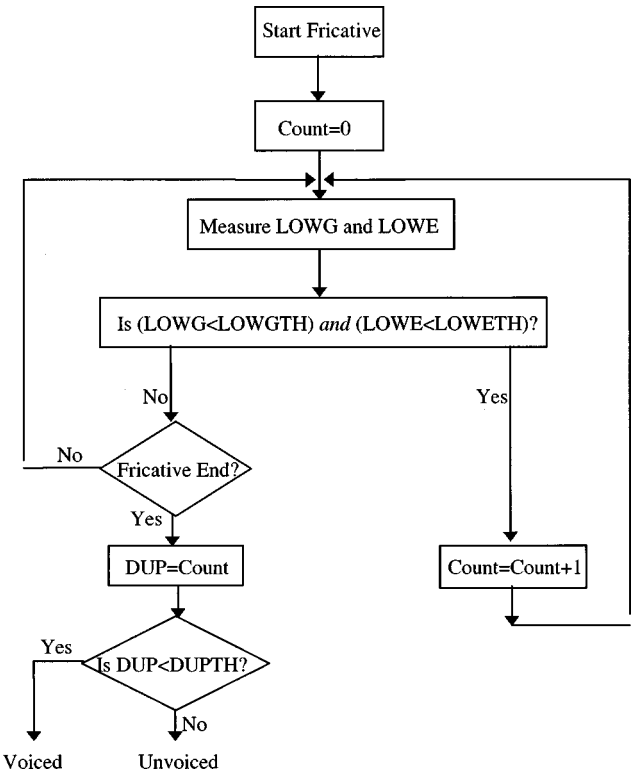


FIG. 5. Voicing detection in prevocalic fricatives.

The first formant transition was also investigated for voicing detection. This feature was extracted and tested using the formant tracker developed to run on the ALSD output. This feature proved to be a sufficient but not necessary cue. The presence of a transition in F1 that persisted for some time (exceeding 10 ms) indicated voicing. However, such a feature did not seem to affect the recognition performance significantly and its practical usefulness for automatic recognition is highly doubted.

The results for the voicing detection on the eight fricatives, /s/, /sh/, /f/, /th/, /z/, /zh/, /v/ and /dh/, are given in the confusion matrix shown in Table I. This represents 93% accuracy.

The following remarks need to be mentioned:

(i) The results in Table I are obtained using the DUP for prevocalic and absolute final fricatives. The number of absolute final fricatives, however, is too low to cause a significant effect. This is because, in continuous speech, most word-final fricatives tend to be medial due to continuity with the following word. It is expected that, had the number of the final fricatives been larger, normalization with respect to the previous vowel's duration would have been needed to preserve such good performance. It was found that most

TABLE I. Confusion matrix for voicing detection. Accuracy is 93%.

	Detected as voiced	Detected as unvoiced
Voiced	90%	10%
Unvoiced	6%	94%

TABLE II. Confusion matrix for voicing detection using LOWE alone. Accuracy is 83%.

	Detected as voiced	Detected as unvoiced
Voiced	75%	25%
Unvoiced	9%	91%

TABLE III. Confusion matrix for voicing detection using LOWG alone. Accuracy is 86%.

	Detected as voiced	Detected as unvoiced
Voiced	75%	25%
Unvoiced	4%	96%

of the errors in that case are due to final voiceless fricatives that are erroneously detected as voiced due to their short durations relative to the DUP threshold.

(ii) The threshold used for the DUP is statistically found to be 60 ms. However, it was found that if the DUP is above 100 ms, then, almost surely, the fricative is voiceless. This is in agreement with Stevens *et al.*'s result (1992), who found 60 ms to be a threshold for voiceless detection.

(iii) The advantage of using the two features LOWG and LOWE instead of just one of them can be confirmed by comparing the confusion matrices of Table I to those of Tables II and III. The superiority of the ALSD is clear. The advantage of using the two features and their complementary action is also obvious. The improvement apparently results from the use of the ALSD output with its powerful periodicity detection ability. An example of a case where the ALSD detected voicing that was missed by the mean-rate is shown in Fig. 6. In this case, the synchrony shows periodicity over the whole phoneme (evident by the low-frequency energy) while the mean-rate shows an interval in which the phoneme is voiceless.

(iv) The choice of the previously mentioned thresholds is determined statistically (using histogram analysis with the maximum posterior probability criterion) to minimize the probability of error and maximize the overall accuracy. It is important to note that those thresholds are independent of the energy of the speech (loudness), the rate of speaking, or the gender of the speaker.

(v) These experiments show an improvement in results compared to those obtained by Stevens *et al.* (1992). This is in spite of the fact that the database used was larger and with continuous speech as opposed to controlled utterances. The reason behind this improvement could be attributed to the different approach used here for detecting phonation (periodicity). This is a clear example of how the translation process from abstract features to measurable parameters could play a significant role in the recognition performance.

2. Relative amplitude and spectral flatness

Relative amplitude (intensity) is a useful feature in discriminating between sibilants and nonsibilants. However, its efficacy in such a task is doubted by some researchers, who noted that some spectral information is needed, in addition to the amplitude, to make such a discrimination. In this part of the experiment, we evaluate the relative amplitude as a feature for discriminating between the place of articulation of

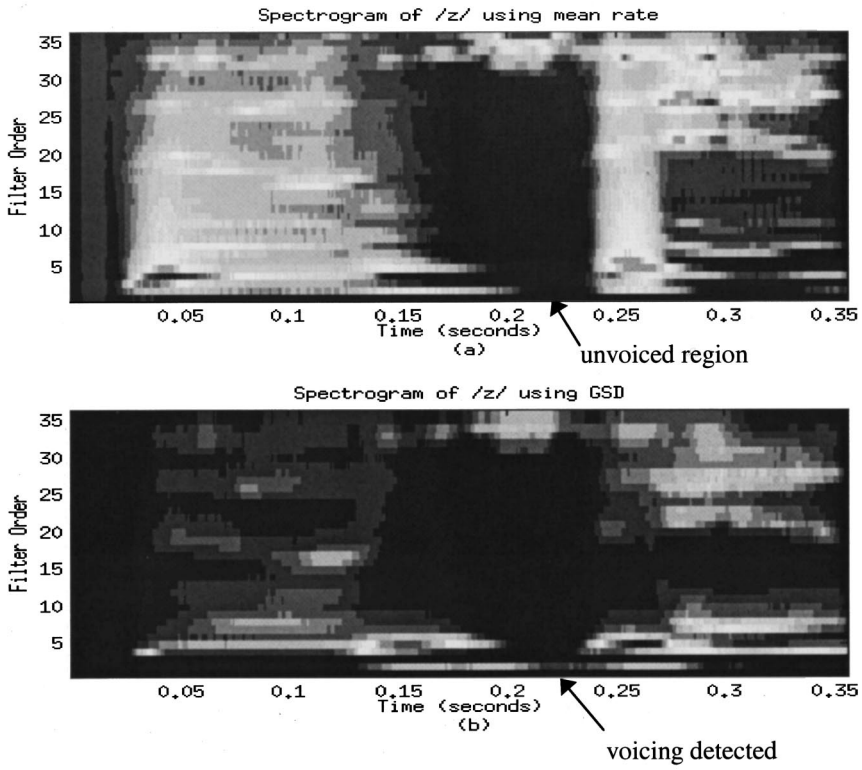


FIG. 6. The output Bark-scaled pseudo-spectrogram of the phoneme /z/ using (a) mean-rate detector and (b) synchrony GSD/ALSD. The /z/ lies in the region from 0.17 to 0.23 s between two vowels.

TABLE IV. Confusion matrix for sibilant/nonsibilant discrimination on the training set using the relative amplitude (RA). Total accuracy is 90%.

	Detected as sibilant	Detected as nonsibilant
/s/ and /z/	95%	5%
/f/, /v/, /th/ and /dh/	22%	78%
/sh/ and /zh/	98%	2%

sibilants and nonsibilants. The relative amplitude (RA) is defined in our work as the ratio between the total energy of the fricative (defined to be the sum of the outputs of the envelope detectors of all the filters) and that of the maximum neighboring vowel, that is,

$$\mathrm{RA}_j = \sum_{i:\text{all filters}} yenv_{ij}\big|_{j\in\text{fricative}} \Big/ \max_{j:\text{all utterance}}\left(\sum_{i:\text{all filters}} yenv_{ij}\right), \tag{1}$$

where $yenv_{ij}$ is the mean-rate output of the ith filter at the jth instant. The RA value for the whole fricative is obtained by averaging the above-mentioned RA_j over the duration of the middle third segment of the fricative.

Using the relative amplitude to discriminate between sibilants and nonsibilants is performed by choosing a threshold above which the sound is considered a sibilant, and below which it is detected as nonsibilant. An optimum threshold was statistically found using histogram analysis to be 0.2 and the resulting confusion matrix is given in Table IV. This is the static threshold value that minimizes the overall error. A better performance could be obtained if a dynamic threshold is used which increases to 0.25 if the fricative is preceded by silence.

The results in Table IV yield an interesting observation: For a minimum overall error, the relative amplitude is apparently better at detecting sibilants than it is at detecting nonsibilants. Although the overall performance is not bad (90%), most of the errors are in misrecognizing nonsibilants as sibilants. This clearly indicates that a considerable percentage (about 22%) of the nonsibilants have large relative amplitudes that are comparable to those of the sibilants. Yet they are still perceived correctly by humans as nonsibilants. This means that the high relative amplitude is a necessary, but not sufficient, condition to perceive sibilants because nearly all sibilants have high relative amplitudes but not all nonsibilants have low relative amplitudes.

Therefore, these results would lead us to the conclusion that decreasing the relative amplitude of sibilants should lead to their perception as nonsibilants because the necessary condition has been violated. However, increasing the relative amplitude of nonsibilants should not lead to their perception as sibilants because it is not a sufficient condition and because there are many nonsibilants which do have high relative amplitudes.

This logical conclusion is in complete agreement with the results obtained by Behrens and Blumstein (1988a) who found in their perceptual experiments that decreasing the relative amplitude of the sibilants /s/ and /sh/ resulted in a predominance of /f/ and /th/ responses, while raising the amplitude of /f/ and /th/ failed to result in a predominance of /s/ and /sh/ responses.

In spite of the relatively high percentage of correct identification shown above (90%), it is apparent that the relative amplitude can not be used alone for the task of differentiating between sibilants and nonsibilants. The 22% miss rate of the nonsibilants is too high to be acceptable. Moreover, the conclusion that high relative amplitude is a necessary but not sufficient condition for sibilant detection clearly indicates that another feature is needed for this task. This other feature, as mentioned before, should detect the relative flatness of the nonsibilant spectra compared to those of the sibilants.

One way to solve this problem is to combine both properties (i.e., the relative amplitude and the spectral flatness) in one feature that could be used solely for the discrimination between sibilants and nonsibilants. We developed a parameter which we called the maximum normalized spectral slope (MNSS). It is defined as the maximum value (over the whole spectrum at a certain instant) of the smoothed differences between neighboring filters, normalized with respect to the maximum energy of the utterance, that is,

$$\mathrm{MNSS} = \left(\max_{i:\text{all filters}} \mathrm{Diff}(yenv_{ij})\right)_{j\in\text{fricative}} \Big/ \max_{f:\text{all utterance}}\left(\sum_{i:\text{all filters}} yenv_{ij}\right), \tag{2}$$

where $yenv_{ij}$ is the ith filter mean-rate (envelope) output at the jth instant, while Diff() is a difference function which approximates the derivative with respect to frequency. It can be as simple as the difference between two neighboring filters, i.e.,

$$\mathrm{Diff}(yenv_{ij}) = yenv_{ij} - yenv_{(i-1)j}. \tag{3}$$

Another method to approximate the derivative is to use the derivative of a second-order fitting polynomial ($h_1 + h_2 f + h_3 f^2$). The derivative of such a polynomial (at $f=0$) is equal to h_2 which is found to be (Rabiner and Juang, 1993)

$$\mathrm{Diff}(yenv_{ij}) = h_{2j} = \frac{\sum_{i=-M}^{M} i \times yenv_{ij}}{T_M}, \tag{4}$$

where $2M+1$ is the number of points used in the fitting process, and T_M is defined as

$$T_M = \sum_{i=-M}^{M} i^2. \tag{5}$$

Using $M=1$, we obtain

$$\mathrm{Diff}(yenv_{ij}) = \frac{yenv_{(i+1)j} - yenv_{(i-1)j}}{2}. \tag{6}$$

Neglecting the constant we obtain

$$\mathrm{Diff}(yenv_{ij}) = yenv_{(i+1)j} - yenv_{(i-1)j}. \tag{7}$$

Both Eqs. (3) and (7) could be used for the approximation of the derivatives. Higher orders (i.e., large values of M) did not result in any practical improvement. Smoothing is not essen-

tially needed in this case but is rather needed in locating the most dominant peak as explained later. Finally, the MNSS value for the whole fricative is found by averaging over the duration of the middle third segment of the fricative.

It is important to note that the MNSS feature is extracted using the mean-rate output (and not the ALSD output). This is essential given the need for the MNSS to capture some relative amplitude information. This information is preserved in the mean-rate output, but not in the synchrony (ALSD) output due to the normalization performed in its calculation.

The MNSS parameter includes and describes the two properties of flatness and low relative amplitude simultaneously. Mathematically, a small RA will definitely lead to a small MNSS (because the difference between two small positive quantities will surely be a small quantity), and hence lead to a nonsibilant decision, which is the correct decision perceptually. On the other hand, a large RA will not necessarily lead to a large MNSS (because the difference between two large quantities may be large or small depending on how similar they are). The value will depend on how close the two quantities are (i.e., how flat the spectrum is). Therefore, a large RA will not necessarily lead to a large MNSS. The value of the MNSS will depend, in this case, on the difference. If the spectrum is flat (as is the case with the nonsibilants), the difference will be small and hence the MNSS will be small. In the case of a nonflat, steep spectrum (as is the case with the sibilants), the difference will be large and hence the MNSS will be large. Therefore a small RA will force the MNSS to be small (overriding the spectral shape) and hence lead to a nonsibilant decision. On the other hand, if the RA is large, the deciding property will be the shape (i.e., flatness) of the spectrum. Therefore, both the RA and flatness property are captured neatly in the MNSS feature so as to represent the acoustic properties that discriminate between sibilants and nonsibilants.

We should note that the normalization in the MNSS expression is done with respect to the maximum energy in the utterance (or with respect to the neighboring vowel), as opposed to the energy of the fricative itself, which was frequently done in the literature. The normalization with respect to the fricative energy is sometimes performed implicitly when using the log-spectrum, since the difference between two logarithmically compressed quantities is equal to the logarithm of their ratio. This is avoided in the MNSS feature. In our case, we prefer the normalization with respect to the utterance maximum energy in order to account for intensity (loudness) variation while preserving the ''desired'' RA information of the fricative in the MNSS quantity.

Large values of the MNSS are expected to correspond to sibilants, small values to nonsibilants. An optimum threshold was statistically found using histogram analysis to be 0.02 for unvoiced fricatives and 0.01 for voiced fricatives. Thus, if the MNSS is greater than the threshold, the fricative is detected as a sibilant, while if it is below it, it is a nonsibilant. The results of this recognition experiment are shown in Table V. A dynamic threshold, however, was found to be more accurate. A higher threshold (by adding 0.005) is needed when the fricative is preceded by a silence and a lower threshold (by subtracting 0.005) if followed by a silence. This adjustment was used in the final recognition experiments.

TABLE V. Confusion matrix for sibilant/nonsibilant discrimination on the training set using the maximum normalized spectral slope (MNSS). Accuracy is 94%.

	Detected as sibilant	Detected as nonsibilant
/s/ and /z	99%	1%
/f/, /v/, /th/ and /dh/	9%	91%
/sh/ and /zh/	85%	15%

The recognition accuracy using the MNSS is 94%, which is better than the 90% obtained by the relative amplitude. However, the improvement is much more significant than a 4% change. This improvement is clear from the significant decrease in the misses of the nonsibilants. Instead of the 22% miss rate obtained when using the relative amplitude, here we have only a 9% miss rate, which is much lower, and even comparable with the miss percentage of the sibilants. This is clearly due to the fact that we do not rely on the relative amplitude alone anymore, but on the spectrum as well. This result is in total agreement with our previous discussion and confirms the conclusion that a high relative amplitude is a necessary but not sufficient condition for detecting sibilants, and that the spectral shape is needed to make up for this deficiency. Apparently, the MNSS has indeed succeeded in integrating the two properties together into one parameter.

To have a better quantitative description of the improvement introduced by the use of the MNSS, an information transmission analysis is needed (Jongman, 1989). Unlike the overall accuracy, which takes into account only the *total* number of errors, the information transmission analysis takes the whole confusion matrix into account. Therefore, it is a better evaluation tool that enables us to detect problems like the high nonsibilant miss rate which may not be obvious in the total error rate.

To perform the information transmission analysis on the two confusion matrices in Tables IV and V we calculate the percentage of the place-of-articulation information that is carried by each of the two features using the following equations:

$$T(x,y) = -\sum_{i,j} p_{ij} \log_2 \frac{p_i p_j}{p_{ij}}, \tag{8}$$

$$H(x) = -\sum_{i} p_i \log_2 p_i \tag{9}$$

$$T_{\text{rel}} = T(x,y)/H(x), \tag{10}$$

where $T(x,y)$ is the transmission in bits per stimulus from input to output, p_i is the probability of an input stimulus i, p_j is the probability of an output stimulus j, p_{ij} is the probability of the joint occurrence of an input i and an output j, $H(x)$ is the entropy of the input, and T_{rel} is the relative information transmitted.

Performing this analysis on our results, we found that the relative amplitude was able to carry about 36% of the

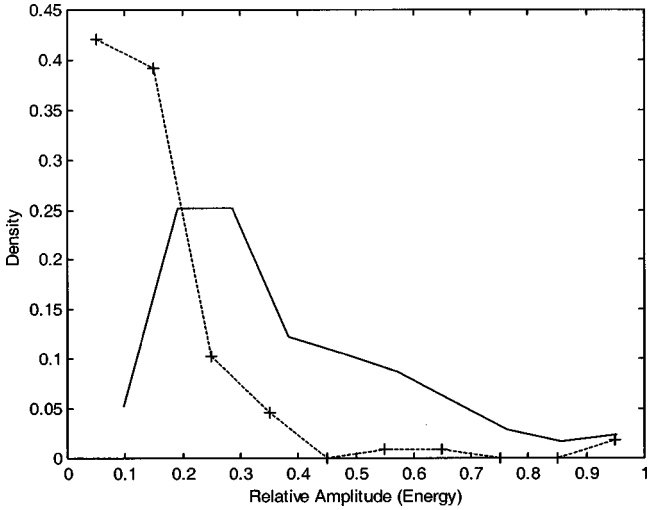


FIG. 7. Histogram of the relative amplitude (RA) for sibilants (solid) and nonsibilants (dashed).

place of articulation information, while the MNSS was able to carry about 46% of that information. This represents a much more significant improvement than that suggested by the total error rate, which clearly demonstrates the significance of the MNSS as a primary feature for place of articulation detection.

More insight can also be gained by studying the histograms of the two features. Figures 7 and 8 show the histograms for the relative amplitude and the MNSS, respectively, for sibilants and nonsibilants. The superiority of the MNSS in clustering these two groups is clear.

3. Spectral shape and peak location

The spectral shape is known to play a major role in the place of articulation detection of fricatives. In this section, we investigate its role through several features (parameters). Only the successful features, or those with meaningful results, are discussed. Features that failed to give satisfactory performance are not discussed at all.

As mentioned in a previous section, alveolar fricatives are characterized by a lowest-frequency peak that is located

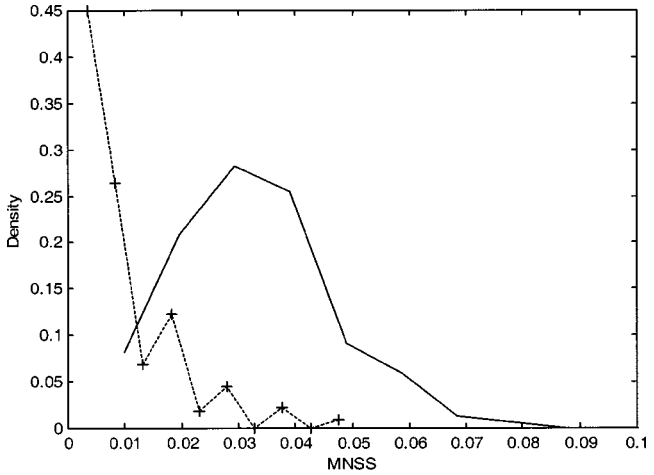


FIG. 8. Histogram of the MNSS for sibilants (solid) and nonsibilants (dashed).

TABLE VI. Confusion matrix for alveolar detection on the training set using the spectral center of gravity (SCG). Accuracy is 84%.

	Detected as alveolar	Detected as non-alveolar
/s/ and /z/	95%	5%
/f/, /v/, /th/ and /dh/	22%	78%
/sh/ and /zh/	28%	72%

at a higher frequency location compared to palatal fricatives. Since nonsibilants (labio- and lingua-dentals) have been successfully detected in the previous section using their relatively flat spectrum and low amplitude, the concern in this section will be on how to discriminate between alveolars and palatals.

One feature that can detect the alveolar fricatives is the spectral center of gravity (SCG) defined as

$$\mathrm{SCG}_j = \sum_{i:\text{all filters}>1.2\text{ kHz}} iyenv_{ij} \Big/ \sum_{i:\text{all filters}>1.2\text{ kHz}} yenv_{ij}. \quad (11)$$

The calculation is performed for filters above 1.2 kHz in order to eliminate the effect of voicing. Averaging over the middle third segment duration is then performed to obtain the SCG. The values of the SCG range between 15 and 36 (which are the minimum and maximum filter orders in the equation). It is expected that alveolars should have the highest SCG compared to other fricatives. Testing this feature in detecting alveolars using a statistically determined threshold (which minimizes the overall error) of filter number 30 (corresponding to 4 kHz) yields the confusion matrix shown in Table VI.

The results in Table VI indicate an accuracy of 84%. This is not adequate given the known and expected importance of the spectral shape in detecting alveolars. Moreover, the obtained errors in detecting alveolars (i.e., false positives) are too high to be acceptable. About 22% of the nonsibilants and 28% of the palatals were erroneously detected as alveolars. This result clearly indicates that the high SCG is a necessary, but not sufficient, cue for detecting alveolars. Its absence strongly suggests a nonalveolar fricative, while its presence could be due to an alveolar or a nonalveolar fricative. This clearly indicates that the SCG plays only an auxiliary role in the alveolar detection and that another feature is needed.

Another important feature is the peak location. The location of the most dominant peak proved to be useful in discriminating between palatals and alveolars. Palatals are characterized by a compact spectrum that has a dominant peak at a relatively low frequency. On the other hand, the alveolars' dominant peak is at a higher frequency, while the nonsibilants usually do not have a significant peak and their most dominant peak is also at a higher frequency. Therefore, this feature could be useful in extracting palatals.

To extract the MDP, however, we need to notice that it is not necessarily the peak with the largest amplitude, since in many cases, especially with a Bark-scaled or Mel-scaled spectrum, there is a considerable increase in energy with

TABLE VII. Confusion matrix for palatal detection on the training set using the mean-rate MDS. Accuracy is 91%.

	Detected as palatal	Detected as non-palatal
/s/ and /z/	7%	93%
/f/, /v/, /th/ and /dh/	11%	89%
/sh/ and /zh/	90%	10%

frequency. Therefore, some peaks may have large amplitudes, due to their high-frequency location, without being dominant. They may even be spurious peaks that do not have any relevance acoustically.

The first method to locate the most dominant peak (MDP) is to locate, in the mean-rate output, the peak that is most prominent compared to its neighborhood. To do that we can measure the smoothed slopes of the spectrum and locate their maximum value. This is the location of the MNSS described before, however, here smoothing is essential to avoid the effect of sharp variations. The NSS quantity is expected to be maximum at the location (or more accurately: the edge) of the peak.

The second method to locate the MDP is to use the ALSD. This system is known for its superb ability in detecting formants and peaks, based on the temporal pattern of the response. Moreover, the ALSD output is known to be relatively independent of the absolute value of the response, but rather looks for the relative dominance. Therefore, the second method to locate the MDP is to locate the maximum slope (MDS) or peak (MDP) in the ALSD response. Both the MDS and the MDP were found to yield comparable performance that was superior to that obtained using the mean-rate. The results of both methods are shown in Tables VII and VIII.

The difference in performance between the two methods is striking. The information transmission analysis of the two methods also yielded significant difference. The mean-rate MDS conveyed 25% of the place of articulation information, while the synchrony MDS conveyed about 41%. The average localized synchrony detector (ALSD) is apparently superior in its ability to extract the MDP as it is in extracting the formants. This discrepancy in performance is another example of how the front-end processing, as well as the mechanism used in translating the abstract feature into a measurable parameter, can play a major role in the recognition system performance.

The results in Tables VII and VIII clearly indicate that the spectral most dominant peak (MDP or MDS), when properly detected using the ALSD system, plays a major and primary role in the place of articulation detection and specifically in detecting palatals. The center of gravity (SCG), however, plays only a secondary role. Another spectral parameter, which was also found to play a minor helpful role in the place detection, is the dominance relative to the highest filter output (call it DRHF). This feature describes the amplitude roll-off in the high-frequency filters relative to the dominant peak. It is expected to be small for alveolars, and is defined as the difference between the MDP and the output of the highest filters:

$$\mathrm{DRHF}_j = \Big(\max_{i:\text{all filters}} (\mathrm{ALSD\_output}_{ij}) - \max_{k:\text{highest 3 filters}} (\mathrm{ALSD\_output}_{kj})\Big)_{j \in \text{fricative.}} \tag{12}$$

TABLE VIII. Confusion matrix for palatal detection on the training set using the synchrony MDS. Accuracy is 98.5%.

	Detected as palatal	Detected as non-palatal
/s/ and /z/	1%	99%
/f/, /v/, /th/ and /dh/	1%	99%
/sh/ and /zh/	92%	8%

B. Place of articulation detection

In the previous section, the acoustic features that are needed in the place of articulation detection are extracted and evaluated separately. In this section, we combine the different features in order to make a place of articulation decision, in a knowledge-based manner, that takes into account the information and understanding that we achieved from the literature as well as from our own experiments. Information transmission analysis and decision trees were used to help design the final system.

We perform the place of articulation detection using five main features:

(1) the maximum normalized spectral slope (MNSS),
(2) the location of the most dominant spectral slope (MDS),
(3) the location of the most dominant spectral peak (MDP),
(4) the spectral center of gravity (SCG), and
(5) the dominance relative to the highest filter (DRHF).

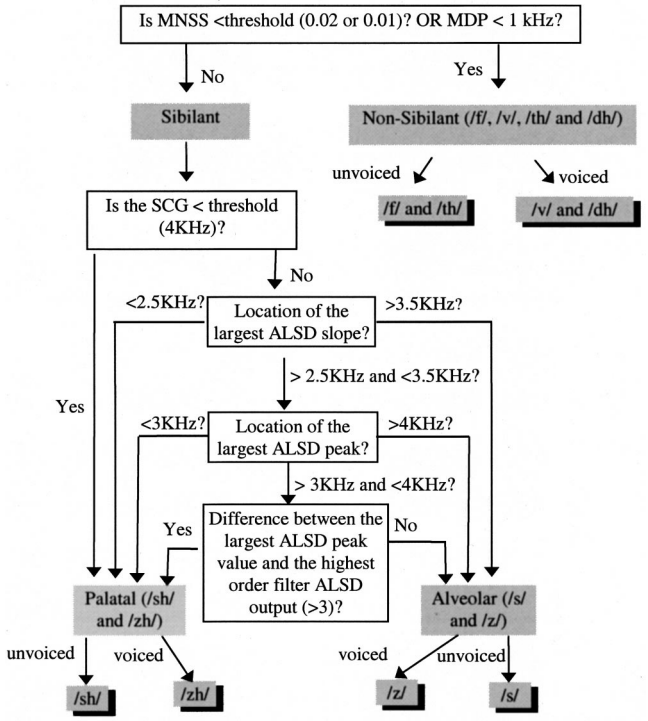


FIG. 9. Algorithm for place of articulation detection.

TABLE IX. Confusion matrix for place of articulation detection for ten speakers used in the system design. Accuracy is 95%.

	Detected as alveolar	Detected as dental	Detected as palatal
Alveolars /s/ and /z/	93%	5%	2%
Dentals /f/, /v/, /th/ and /dh/	1%	98%	1%
Palatals /sh/ and /zh/	5%	1%	94%

As previously mentioned, each fricative is manually extracted and divided into three equal segments. The above-mentioned spectral parameters are then averaged over the duration of the middle segment and used in the algorithm shown in Fig. 9.

The results of the place of articulation detection are represented in the confusion matrices of Tables IX and X.

The accuracy obtained using this algorithm is about 91%. The improvement due to using a dynamic threshold for the MNSS and combining multiple features is evident. An example of what the spectra of different places of articulation look like is shown in Fig. 10. The different characteristic features are clearly evident. Figure 10(a) shows the synchrony response of the three places of articulation. The weak response of the labiodental /f/, the high-frequency peak of the alveolar /s/, and the lower frequency dominant peak of the palatal /sh/ are obvious.

It is interesting to notice that the mean-rate spectrum shown in Fig. 10(b) failed to correctly localize the dominant peak position of the /sh/. This limitation is mentioned before and is due to the dependence of the mean-rate response on the amplitude, which tends to increase with frequency due to the use of the Bark scale and the front-end preemphasis. This makes it difficult for the mean-rate to correctly detect dominant peaks in the mid-frequency region. The ALSD response, on the other hand, is less sensitive to the amplitude and detects the peak position correctly. In order to detect the dominant peak using the mean-rate, we use the NSS, which is shown in Fig. 11. The peaks (or more accurately, their edges) are correctly detected. We can also see clearly the small value of the maximum NSS (i.e., MNSS) of the /f/ compared to the /s/ and the /sh/.

It is obvious from the results in Tables IX and X that the labiodentals (/f/ and /v/) and the linguadentals (/th/ and /dh/) are grouped together. In our study, we were unable to find

TABLE X. Confusion matrix for place of articulation detection for 60 speakers from seven different dialects of American English. Accuracy is 91%.

	Detected as alveolar	Detected as dental	Detected as palatal
Alveolars /s/ and /z/	90%	6%	4%
Dentals /f/, /v/, /th/ and /dh/	3%	92%	5%
Palatals /sh/ and /zh/	6%	3%	91%

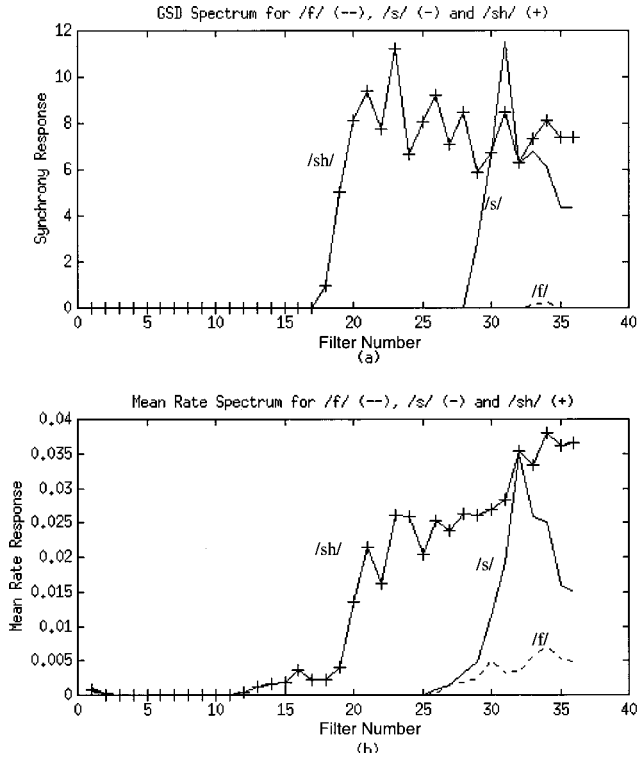


FIG. 10. The spectra of typical dental (/f/ - -), alveolar (/s/ −) and palatal (/sh/ +) using (a) synchrony and (b) mean-rate.

any static feature that could lead to the successful discrimination between those two groups. Formant transitions, which were speculated in the literature to be a possible cue for that purpose, also failed to give us any consistent context-independent results. A possible cue for the discrimination is the high-frequency fall-off, which usually is above 8 kHz and therefore could not be detected in the TIMIT database whose sampling frequency is 16 kHz. This point needs further investigation.

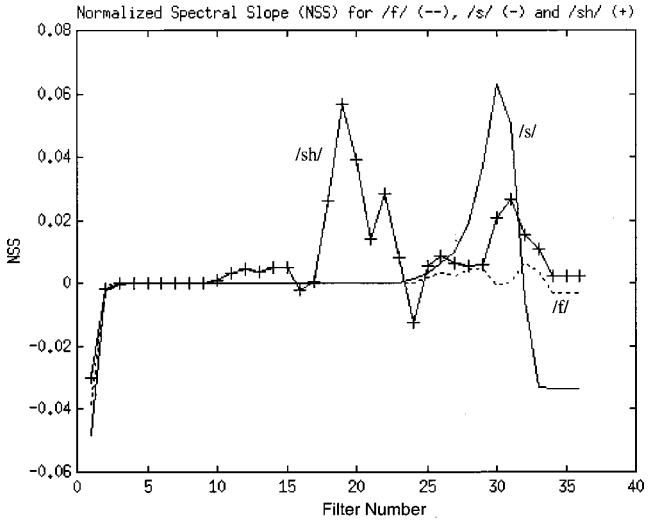


FIG. 11. The normalized spectral slopes (NSS) for a typical dental (- -), alveolar (−), and palatal (+).

TABLE XI. Confusion matrix for fricatives detection (in percentages). Overall accuracy is 87%.

	Detected as /s/	Detected as /f/, /th/	Detected as /sh/	Detected as /z/	Detected as /v/, /dh/	Detected as /zh/
/s/	88%	7%	2%	2%	0.5%	0.5%
/f/, /th/	5%	83%	5%	1%	5%	1%
/sh/	5%	2%	88%	1%	1%	3%
/z/	7%	2.5%	2%	83%	2%	3.5%
/v/, /dh/	0.5%	3%	0.5%	0.5%	93%	2.5%
/zh/	1%	1%	11%	6%	1%	80%

TABLE XII. Confusion matrix obtained using statistical discriminant analysis. Overall accuracy is 96%.

	Detected as alveolar /s, z/	Detected as dental /f, v, th, dh/	Detected as palatal /sh, zh/
Alveolar	96%	4%	0%
Dental	3%	97%	0%
Palatal	4%	2%	94%

TABLE XIII. Confusion matrix obtained using statistical discriminant analysis using all features except the SCG. Overall accuracy is 95%.

	Detected as alveolar /s, z/	Detected as dental /f, v, th, dh/	Detected as palatal /sh, zh/
Alveolar	93%	6.8%	0.2%
Dental	3%	97%	0%
Palatal	4%	2%	94%

TABLE XIV. Confusion matrix obtained using statistical discriminant analysis using all features except the MNSS. Overall accuracy is 87%.

	Detected as alveolar /s, z/	Detected as dental /f, v, th, dh/	Detected as palatal /sh, zh/
Alveolar	90%	7%	3%
Dental	17%	77%	6%
Palatal	4%	3%	93%

TABLE XV. Confusion matrix obtained using statistical discriminant analysis using all features except the MDS. Overall accuracy is 91%.

	Detected as alveolar /s, z/	Detected as dental /f, v, th, dh/	Detected as palatal /sh, zh/
Alveolar	95%	4.5%	0.5%
Dental	9%	88%	3%
Palatal	6%	3%	91%

TABLE XVI. Confusion matrix obtained using statistical discriminant analysis using all features except MDP. Overall accuracy is 91.4%.

	Detected as alveolar /s, z/	Detected as dental /f, v, th, dh/	Detected as palatal /sh, zh/
Alveolar	87%	10%	3%
Dental	3%	93%	4%
Palatal	4%	2%	94%

TABLE XVII. Confusion matrix obtained using statistical discriminant analysis using all features except DRHF. Overall accuracy is 96%.

	Detected as alveolar /s, z/	Detected as dental /f, v, th, dh/	Detected as palatal /sh, zh/
Alveolar	96%	4%	0%
Dental	3%	97%	0%
Palatal	3%	2%	95%

C. Overall classification

To perform the overall recognition of fricatives, voicing detection is performed using the algorithm in Fig. 5 to discriminate between voiced and unvoiced fricatives. Then the place of articulation detection is performed using the algorithm of Fig. 9 to classify the fricatives into dentals, alveolars, and palatals. The voicing detection is done first because some of the thresholds used in the place detection depend on the voicing decision. The overall recognition accuracy was 87%, keeping in consideration that /f/ and /th/ were put in one group, as well as /v/ and /dh/. This is shown in Table XI.

D. Statistical analysis

The features described in our experiments were used in knowledge-based decision-tree-like algorithms. The algorithms were designed based on the acoustic-phonetic characteristics of various sounds. The resulting interaction and relevance of each feature in the corresponding classification task are hence clear and well defined, as previously discussed. It is desirable, however, to evaluate the statistical relevance and role of various features when used in a data-driven algorithm.

We used statistical discriminant analysis, and Bayesian classification using the maximum posterior probability criterion (MAP), to cluster (classify) the given fricatives into three places of articulation using the extracted parameters, i.e., MDP, MDS, DRHF, SCG, MNSS, and the voicing decision. The MAP criterion is known to yield an optimum detection performance and hence represents a best-case scenario. The result of this analysis on the studied 1100 fricatives as training data is shown in Table XII. The classification accuracy on the training data is extremely high, which demonstrates the power of the extracted parameters in detecting place of articulation for fricatives in speaker-independent, continuous speech.

Tables XIII–XVIII show the effect of missing a parameter on the place of articulation detection. Figure 12 shows the effect on the overall classification accuracy when remov-

TABLE XVIII. Confusion matrix obtained using statistical discriminant analysis using all features except the voicing decision. Overall accuracy is 95%.

	Detected as alveolar /s, z/	Detected as dental /f, v, th, dh/	Detected as palatal /sh, zh/
Alveolar	95%	4%	1%
Dental	4%	96%	0%
Palatal	3%	4%	93%

ing one, or some, of the features. Some interesting conclusions could be made:

(i) The spectral center of gravity (SCG) effect is clearly minor as explained previously. Its expected role in distinguishing the alveolars is clear from the increase of the alveolars detected erroneously as dentals in Table XIII.

(ii) The maximum normalized spectral slope (MNSS) absence caused a significant deterioration as shown in Table XIV. Since the role of the MNSS was mainly to extract the dentals, its absence led to a large percentage of misses of the dental fricatives.

(iii) Both the MDS and the MDP have comparable importance on the overall detection as shown in Tables XV and XVI. The MDS, however, is slightly more influential. This is in agreement with our conclusions and our system design considerations.

(iv) The effect of the DRHF alone is negligible as seen from Table XVII. However, when both the DRHF and the MDP are excluded, a significant deterioration happens as shown in Fig. 12 (bar 3) where the recognition dropped to 85%. This is in agreement with our previous conclusions that demonstrated that the role of the DRHF is only auxiliary.

(v) Neglecting the spectral peak information (MDS, MDP, and DRHF) causes a significant deterioration (74%) as shown in Fig. 12 (bar 2). This illustrates the importance of the peak location in place of articulation detection. Also, using the peak slope alone (MDS) without the auxiliary peak information (MDP and DRHF) or the MNSS caused a terrible deterioration as shown in Fig. 12 (bar 1). This illustrates the importance of combining the different features in the decision process.

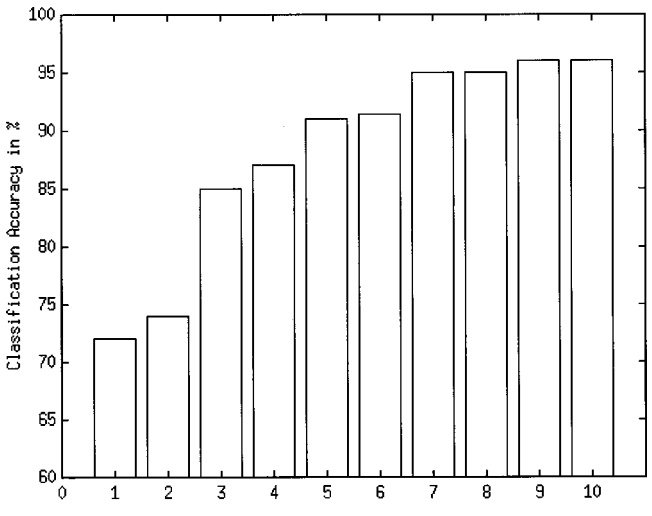


FIG. 12. Classification of place of articulation accuracy using statistical discriminant analysis with different feature sets. (10) Using all features. (9) Without the DRHF. (8) Without the voicing decision. (7) Without SCG. (6) Without MDP. (5) Without MDS. (4) Without MNSS. (3) Without the auxiliary spectral information (MDP and DRHF). (2) Without any spectral peak information (i.e., no MDS, MDP, or DRHF). (1) MDS, SCG, and voicing decision only (i.e., no MDP, DRHF, or MNSS).

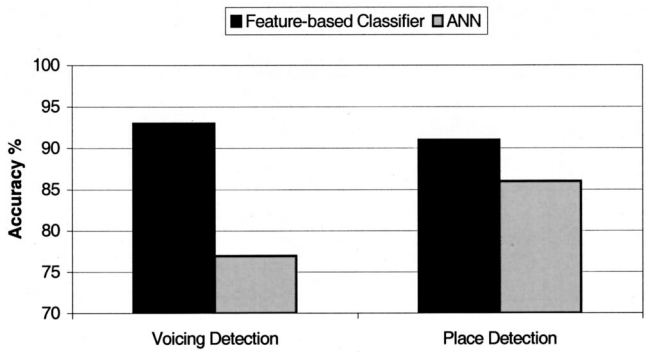


FIG. 13. Classification accuracies for the voicing and place of articulation detection using the feature-based systems described in this article and artificial neural networks using MFCCs as inputs.

(vi) The statistical results demonstrate that each parameter plays a well-defined, relatively independent, role. Sometimes, a certain parameter is responsible for detecting a certain place of articulation. The absence of any parameter causes a localized deterioration that depends on its role. For example, removing the SCG increases the errors of the alveolars, while removing the MNSS causes a large dental miss rate. This is different from the stop consonants where the different parameters interact in a complex manner that makes their roles highly interdependent, as discussed in Ali (1999).

E. Comparison with data-driven systems

To compare the performance of this system with a traditional data-driven system, we performed the same classification experiments using artificial neural networks (ANN). The networks used are multi-layer perceptrons (MLP) trained using the back-propagation algorithm. They have one input layer, one hidden layer, and one output layer. The inputs are 12 Mel-frequency cepstral coefficients (MFCC) (Yu and Cheah, 2000). The best performance obtained was 77% for voicing detection and 86% for the place of articulation detection. This was achieved using 12 input nodes, 15 hidden nodes, and 2 output nodes for the voicing network and 12 input nodes, 10 hidden nodes and 3 output nodes for the place network. Using the MFCC derivatives (in addition to the MFCC) did not result in any improvement.

As previously discussed, the acoustic-phonetic feature-based approach described in this article yielded 93% and 91% for the voicing and place detection, respectively. As shown in Fig. 13, this indicates a comparable performance for the place detection and a significant improvement when using the acoustic-phonetic approach in the voicing detection. This improvement in the voicing detection is due to the use of dynamic features (mainly durations) in the feature-based system. It is difficult for the ANN to capture this kind of feature from the MFCCs and their derivatives. It is also interesting to note that the number of inputs for the feature-based system was only 5, as opposed to 12 for the ANN.

VI. ENVIRONMENTAL ROBUSTNESS

Environmental robustness was one of the main concerns throughout this work that motivated many of the design and implementation decisions. There are several aspects of this system that inherently improve its robustness. Those are as follows.

(1) The system uses auditory-based front-end processing. This approach has proved to yield better robustness to noise, and other mismatches, relative to other traditional systems in various ASR problems.
(2) The system uses a feature-based approach for classification. This is believed to be more immune to noise relative to traditional template (pattern) matching approaches. The robustness is even more evident if the features are carefully designed to be relatively insensitive to noise.

A. Auditory-based processing

Auditory-based processing systems in general and the Seneff (1988, 1985) system (which is similar to the system used in this work) in particular were shown to outperform traditional Fourier-based approaches in the presence of noise and other adverse environments (Ghitza, 1994, 1992, 1988; Jankowski *et al.*, 1995; Junqua and Haton, 1996; Kim *et al.*, 1999; Ohshima, 1993; Sandhu and Ghitza, 1995; Stern *et al.*, 1992; Wang and Shamma, 1994). Although theoretical verification and explanation for such phenomenon do not exist, the practical evidence is overwhelming. It is argued that such robustness stems from their ability to closely emulate the auditory-periphery processing. In fact, it was found that every component in the Seneff system has a role in achieving such robustness and that removing any of the major parts could significantly alter the performance (Ohshima, 1993).

It was shown in Junqua and Haton (1996) that critical-band filtering improves the performance of the system in the presence of noise. The compressive nonlinearity was also shown in Ohshima (1993) to enhance the system performance in the presence of noise. This was theoretically validated by Wang and Shamma (1994) who demonstrated the role played by this nonlinearity in improving the signal-to-noise ratio of the system's output.

An important system component, which was repeatedly shown to significantly enhance the system robustness, is the short-term adaptation and forward-masking block. This module enhances the system response to dynamic events relative to static inputs. Dynamic features (like delta cepstra and delta-delta cepstra) were shown to improve ASR in the presence of noise (Lee *et al.*, 1996). By emphasizing the dynamic events, and attenuating the response to static inputs, the output becomes more immune to stationary noise. This is similar to the high-pass filtering that is used in some systems [like the RASTA processing (Hermansky and Morgan, 1994)], to attenuate the effect of additive and convolutional noise.

Another vital factor for the system robustness is the synchrony detection module. Extracting dominant periodic information was found by Ghitza (1994, 1992, 1988) to be key to robustness against noise. This was confirmed by others like Kajita and Itakura (1995), Ohshima (1993), and Sandhu and Ghitza (1995). This is due to the inherent relative insensitivity of periodicity to additive noise compared to the amplitude. Moreover, the amplitude normalization performed in the GSD and ALSD reduces the effect of spectral tilts on the output. The ALSD further improves the noise performance by reducing the response to spurious peaks compared to the dominant formants.

Wang and Shamma (1994) demonstrated that a compressive nonlinearity followed by a second filter narrower in band, relative to the filter bank, accounts for the system robustness to noise. This narrow-band filter is achieved by both the synchrony detectors, which behave like narrow-band comb filters, and the lateral inhibitory network (LIN). Moreover, the LIN demonstrates good immunity to white noise due to its detection of dominant spectral slopes (edges), as opposed to peaks. This cancels the effect of any flat components (like white noise) in the spectrum. Such a desirable property can be employed in the GSD and ALSD as well by using their spectral slopes instead of peaks as will be shown in the next section.

B. Feature extraction

The other aspect of the system's robustness is using the feature-based approach and the careful choice of features. In general, some features are more robust to noise and mismatches than others. For example, dynamic features were consistently demonstrated to be less affected by noise and mismatches compared to static spectral features. Some other features are inherently less affected by noise. For example, formants in general, and those extracted by synchrony detectors in particular, are known to be relatively robust to noise. Moreover, features like the MDS and the MNSS enjoy relative immunity to white noise due to the lateral inhibition effect employed in their extraction. This leads to extracting spectral edges (as opposed to peaks) and hence reduces the effect of any additive flat components like white noise. Features obtained by detecting periodicity are also shown to be relatively less sensitive to additive noise. And finally, the normalization employed in the MNSS and synchrony detection leads to features that are less affected by spectral tilts and additive noise. In addition, relational features are also known to improve the immunity of the system to speaker and channel variability.

C. Performance

Knowing the various robustness issues taken into consideration in the system design, it would be informative to have a quantitative assessment of the system performance in the presence of noise. To evaluate the robustness of the system, some of the classification experiments mentioned previously are performed in various adverse conditions. We need to emphasize that the system with all its components was designed (trained) on clean speech and was not modified at all to accommodate the adverse environments. In other words, this testing condition corresponds to mismatched training and testing environments.

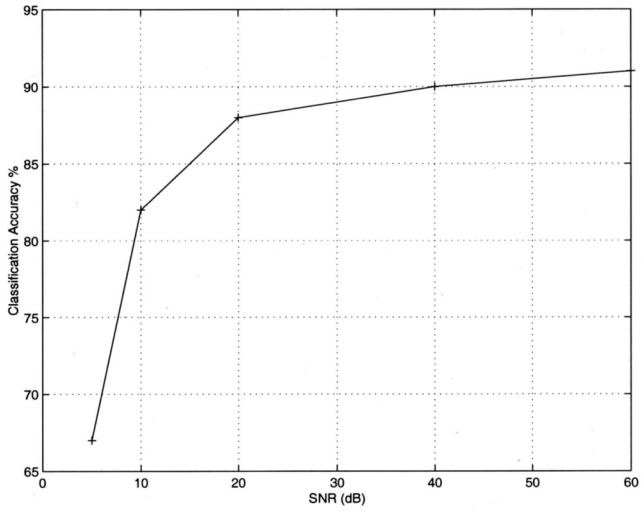


FIG. 14. Classification accuracy in the presence of additive white Gaussian noise for fricatives.

The experiments performed include the place of articulation detection of fricatives in the presence of additive white Gaussian noise at various signal-to-noise ratios (SNRs). The choice of white noise is motivated by the fact that this kind of noise was found to be the most detrimental to phoneme recognition and hence represents a worst case condition (Junqua *et al.*, 1994). The system is also tested for its sensitivity to spectral tilts (convolutional noise). This is done by low-pass, or high-pass, filtering the speech at some cutoff frequency. Obviously, the lower the low-pass cutoff frequency, and the higher the high-pass cutoff frequency, the worse the performance. In our tests, we chose low-pass filtered speech at 5 kHz for testing the fricative classification.

Figure 14 shows the classification results for fricatives in the presence of additive white Gaussian noise at different SNRs. The emphasis was on the place of articulation detection since it is expected to be more sensitive to noise relative to the voicing detection.

The patterns of the errors obtained were interesting. It is found that the majority of the fricative errors were confusions between alveolars and labials. Labials tend to be easily detected as alveolars (in the presence of noise), while the opposite happens much less frequently. This is mainly due to the high-frequency energy introduced by the noise, which alters the labial spectra into alveolarlike spectra. Velars, on the other hand, were less affected by noise, and if affected they tend to be detected as labials. This is due to the fact that velars tend to be relatively compact (peaky) and hence less affected by white noise.

It is clear that the system demonstrates a relatively good immunity to noise as evidenced by the negligible deterioration in performance even at SNRs as low as 20 dB. The accuracy at SNR = 10 dB is still considered acceptable. This robustness is mainly due to the auditory-based processing and acoustic-phonetic features used.

Figure 15 shows the results of low-pass filtering speech at 5 kHz for fricatives. Overall recognition accuracy drops from 87% to 81%. Virtually all the extra errors are in the place detection while the voicing detection remains unaffected by the low-pass filtering. Most of those errors are alveolars erroneously detected as labials, and to lesser extent as velars. Unlike training-based systems (Lippman and Carlson, 1997), the deterioration of this system due to the spectral tilt is graceful and most of the errors resemble those experienced by humans in similar conditions. To put the results in perspective, human performance for nonsense CVC syllables is also shown.

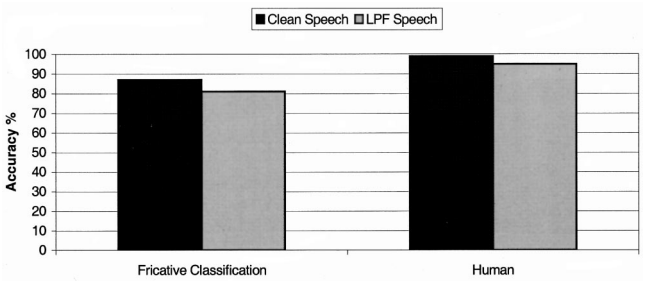


FIG. 15. Classification accuracy for clean speech and for speech low-pass filtered at 5 kHz. Fricative classification is performed using the algorithms and systems described in this work. ''Human'' refers to the human performance for CVC nonsense syllables.

VII. CONCLUSION

In this work, the acoustic-phonetic features characterizing the fricative consonants were analyzed in detail. Features that are discussed in the literature, as well as some new features developed by the author, were investigated and evaluated from the classification standpoint. Five features proved to be very useful in detecting the place of articulation, namely, the maximum normalized spectral slope (MNSS), the location of the most dominant slope (MDS), the location of the most dominant peak (MDP), the dominance relative to high filters (DRHF), and the spectral center of gravity (SCG). These features were able to achieve 91% accuracy.

As for voicing detection, we used the duration of the unvoiced portion (DUP) of the fricative as the main cue in detecting voicing. Using two parameters to formulate this feature, an accuracy of 93% was obtained.

The obtained results show an improvement compared to previous work. Similar experiments, which study the automatic recognition of fricatives using multispeaker continuous speech with comparable database size, are quite rare in the literature. Results obtained by Hughes and Halle (1956) for the place of articulation detection on a smaller database (190 fricatives from five speakers) gave between 77% and 80% recognition accuracy. They used several features describing the spectral shape to perform their recognition. Bengio *et al.* (1992) used artificial neural networks (ANNs) and obtained 74.8% accuracy, but their system was capable of discriminating between labiodentals and linguadentals as well. The 91% accuracy obtained in our experiments indicates an improvement that is mainly due to the different choice of features and parameters and to the knowledge invested in their design.

In voicing detection, the obtained results show an improvement over the 83% and 85% detection rate obtained by Stevens *et al.* (1992) for voiced fricatives. This improvement

is attributed mainly to the use of the ALSD, which, when used with the mean-rate, demonstrates better voicing detection ability.

Halberstadt (1998) used heterogeneous features in a multiple classifier segment-based architecture to perform phonetic classification. His best fricative classification accuracy ranged from 84.9% for strong fricatives to 89.7% for weak fricatives. Those results are clearly comparable with the 87% obtained in our work.

Some important conclusions have been discussed in this work and are summarized below:

(i) The average localized synchrony detector (ALSD) has proved to be very useful in voicing detection as well as place of articulation detection. Its ability to detect periodicity and extract dominant peaks accurately is superior to that of the mean-rate envelope detector.

(ii) It was clear from our results that the method used in translating the acoustic abstract feature into a measurable parameter makes a clear difference in the overall performance. Large discrepancies in performance were observed between approaches whose only difference between them is in how they extract a certain feature. This observation stresses and highlights the need for more research to enhance our knowledge about speech from the ASR standpoint. ASR system designers need the measurable features (parameters), and not the abstract ones, in order to integrate more knowledge in the available systems or to build knowledge-based systems.

(iii) The role of the relative amplitude in discriminating between sibilants and nonsibilants deserves a comment. We found that high relative amplitude is necessary, but not sufficient, for the detection of sibilants. In other words, low relative amplitude is a sufficient, but not necessary, condition for detecting nonsibilants. This clearly indicates that decreasing the amplitude of sibilants will cause them to be perceived as nonsibilants, but increasing the amplitude of nonsibilants will not cause them to be perceived as sibilants. This result explains, and is in total agreement with, the result obtained by Behrens and Blumstein (1988a). It also indicates that the relative amplitude is a useful, but not sufficient, cue for detecting sibilants.

(iv) A new feature is used, which captures the relative amplitude and the flatness property of the spectrum into one quantity, called the maximum normalized spectral slope (MNSS). This quantity proved to be useful in discriminating between sibilants and nonsibilants. In fact, this quantity was able, solely, to achieve this discrimination with an accuracy that was not achievable otherwise. According to our knowledge, no similar quantity had been proposed before to achieve such a task. The reason behind this lies in the fact that, in most front-end processing systems, logarithmic compression is done. This means that the difference (or slope) between neighboring filter outputs actually corresponds to the ratio of their amplitudes. This leads to an automatic scaling which removes any dependence on the absolute amplitude of the signal. This scaling is exactly what we do not want and is the reason why such quantity was not used before. In our system, no logarithmic compression is performed and therefore no scaling is done with respect to the fricative amplitude. This ensures that the ''desirable'' amplitude information of the fricative is still preserved because it is definitely needed. The normalization, in our case, is done with respect to the vowel and not with respect to the fricative amplitude.

(v) Auditory-based processing and feature-based recognition contribute significantly to the robustness of our system in the presence of noise. Recognition experiments were performed to illustrate the robustness of various classification tasks to additive and convolutional noise. The results indicate a relatively robust performance. Acceptable classification accuracies were achieved with speech contaminated with white noise at signal-to-noise ratios as low as 10 dB and with speech low-pass filtered at 5 kHz.

ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for their valuable comments and insightful suggestions. This work is supported by a grant from Catalyst Foundation.

Ali, A. M. A. (**1999**). ''Auditory-based Acoustic-phonetic Signal Processing for Robust Continuous Speech Recognition,'' Ph.D. dissertation, University of Pennsylvania.

Ali, A. M. A. *et al.* (**2000**). ''Auditory-based speech processing based on the average localized synchrony detection,'' in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing* (ICASSP-200).

Baum, S. R., and Blumstein, S. E. (**1987**). ''Preliminary observations on the use of duration as a cue to syllable-initial fricative consonant voicing in English,'' J. Acoust. Soc. Am. **82**, 1073–1077.

Behrens, S., and Blumstein, S. E. (**1988a**). ''On the role of the amplitude of the fricative noise in the perception of place of articulation in voiceless fricative consonants,'' J. Acoust. Soc. Am. **84**, 861–867.

Behrens, S. J., and Blumstein, S. E. (**1988b**). ''Acoustic characteristics of English voiceless fricatives: a descriptive analysis,'' J. Phonetics **16**, 295–298.

Bengio, Y., De Mori, R., Glammia, G., and Kompe, R. (**1992**). ''Phonetically motivated acoustic parameters for continuous speech recognition using artificial neural networks,'' Speech Commun. **11**(2–3), 261–271.

Borden, G. J. *et al.* (**1994**). *Speech Science Primer: Physiology, Acoustics and Perception of Speech* (Williams & Wilkins, Baltimore).

Cole, R. A., and Cooper, W. E. (**1975**). ''Perception of voicing in English affricates and fricatives,'' J. Acoust. Soc. Am. **58**, 1280–1287.

Crystal, T. H., and House, A. S. (**1988**). ''Segmental durations in connected-speech signals: Current results,'' J. Acoust. Soc. Am. **83**, 1553–1573.

Davis, S. B., and Mermelstein, P. (**1980**). ''Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences,'' IEEE Trans. Acoust. Speech, Signal Process. **ASSP-28**, 357–366.

Delgutte, B. (**1986**). ''Analysis of French stop consonants using a model of the peripheral auditory stop consonants using a model of the peripheral auditory systems,'' in *Invariance and Variability in Speech Processes*, edited by J. S. Perkell and D. H. Klatt (Erlbaum, Hillsdale, NJ).

Delgutte, B. (**1980**). ''Representation of speech-like sounds in the discharge patterns of auditory nerve fibers,'' J. Acoust. Soc. Am. **68**, 843–857.

Delgutte, B. (**1984**). ''Speech coding in the auditory nerve: II. Processing schemes for vowel-like sounds,'' J. Acoust. Soc. Am. **75**, 879–886.

Delgutte, B., and Kiang, N. Y. S. (**1984a**). ''Speech coding in the auditory nerve: I. Vowel-like sounds,'' J. Acoust. Soc. Am. **75**, 866–878.

Delgutte, B., and Kiang, N. Y. S. (**1984b**). ''Speech coding in the auditory nerve: III. Voiceless fricatives consonants,'' J. Acoust. Soc. Am. **75**, 887–896.
Delgutte, B., and Kiang, N. Y. S. (**1984c**). ''Speech coding in the auditory nerve: IV. Sounds with consonants-like dynamic characteristics,'' J. Acoust. Soc. Am. **75**, 897–907.
Delgutte, B., and Kiang, N. Y. S. (**1984d**). ''Speech coding in the auditory nerve: V. Vowels in background noise,'' J. Acoust. Soc. Am. **75**, 908–918.
Edwards, H. T. (**1992**). *Applied Phonetics: The Sounds of American English* (Singular).
Fant, G. (**1960**). *Acoustic Theory of Speech Production* (Mouton, 's-Gravenhage).
Fant, G. (**1973**). *Speech Sounds and Features* (MIT, Cambridge).
Ghitza, O. (**1988**). ''Temporal non-place information in the auditory nerve firing patterns as a front-end for speech recognition in a noisy environment,'' J. Phonetics **16**, 106–123.
Ghitza, O. (**1992**). ''Auditory nerve representation as a basis for speech processing,'' in *Advances in Speech Signal Processing*, edited by S. Furui and M. M. Sondhi (Marcel Dekker, New York).
Ghitza, O. (**1994**). ''Auditory models and human performance in tasks related to speech coding and speech recognition,'' IEEE Trans. Speech Audio Process. **2**, 115–132.
Guerlekian, J. A. (**1981**). ''Recognition of the Spanish fricatives /s/ and /f/,'' J. Acoust. Soc. Am. **70**, 1624–1627.
Halberstadt, A. K. (**1998**). ''Heterogeneous acoustic measurements and multiple classifiers for speech recognition,'' Ph.D. thesis, MIT.
Harris, K. S. (**1958**). ''Cues for the discrimination of American English fricatives in spoken syllables,'' Lang. Speech **1**, 1–17.
Hedrick, M. S., and Ohde, R. N. (**1993**). ''Effect of relative amplitude of frication on perception of place of articulation,'' J. Acoust. Soc. Am. **94**, 2005–2026.
Heinz, J. M., and Stevens, K. N. (**1961**). ''On the properties of voiceless fricative consonants,'' J. Acoust. Soc. Am. **33**, 589–596.
Hermansky, H., and Morgan, N. (**1994**). ''RASTA processing of speech,'' IEEE Trans. Speech Audio Process. **2**, 587–589.
Hughes, G. W., and Halle, M. (**1956**). ''Spectral properties of fricative consonants,'' J. Acoust. Soc. Am. **28**, 303–310.
Hunt, M. J., and Lefebvre, C. (**1989**). ''A comparison of several acoustic representations for speech recognition with degraded and undegraded speech,'' in Proc. ICASSP, pp. 262–265.
Jankowski, C. R., Vo, H. H., and Lippmann, R. P. (**1995**). ''A comparison of signal processing front ends for automatic word recognition,'' IEEE Trans. Speech Audio Process. **3**, 286–293.
Jongman, A. (**1989**). ''Duration of frication noise required for identification of English fricatives,'' J. Acoust. Soc. Am. **85**, 1718–1725.
Junqua, J.-C., and Haton, J.-P. (**1996**). *Robustness in Automatic Speech Recognition, Fundamentals and Applications* (Kluwer Academic, Boston).
Junqua, J. *et al.* (**1994**). ''A robust algorithm for word boundary detection in the presence of noise,'' IEEE Trans. Speech Audio Process. **2**, 406–412.
Kajita, S., and Itakura, F. (**1995**). ''Robust speech feature extraction using SBCOR analysis,'' in *Proc. IEEE ICASSP*, pp. 421–424.
Kim, D.-S. *et al.* (**1999**). ''Auditory processing of speech signals for robust speech recognition in real-world noisy environments,'' IEEE Trans. Speech Audio Process. **7**, 55–69.
Lee, C.-H. *et al.* (ed.) (**1996**). *Automatic Speech and Speaker Recognition, Advanced Topics* (Kluwer Academic, Boston).
Lippman, R. P., and Carlson, B. A. (**1997**). ''Robust speech recognition with time-varying filtering, interruptions, and noise,'' in *Proc. IEEE Workshop on Automatic Speech Recognition and Understanding*, pp. 365–372.
Manrique, A. M. B., and Massone, M. I. (**1981**). ''Acoustic analysis and perception of Spanish fricative consonants,'' J. Acoust. Soc. Am. **69**, 1145–1153.
McCasland, G. P. (**1979a**). ''Noise intensity and spectrum cues for spoken fricatives,'' J. Acoust. Soc. Am. Suppl. 1 **65**, S78–S79.
McCasland, G. P. (**1979b**). ''Noise intensity cues for spoken fricatives,'' J. Acoust. Soc. Am. Suppl. 1 **66**, S88.
Ohshima, Y. (**1993**). ''Environmental robustness in speech recognition using physiologically-motivated signal processing,'' Ph.D. thesis, Carnegie Mellon University.
Rabiner, L., and Juang, B.-H. (**1993**). *Fundamentals of Speech Recognition* (Prentice–Hall, Englewood Cliffs, NJ).
Sandhu, S., and Ghitza, O. (**1995**). ''A comparative study of Mel cepstra and EIH for phone classification under adverse conditions,'' *Proc. IEEE ICASSP*, pp. 409–412.
Seneff, S. (**1988**). ''A Joint Synchrony/Mean Rate Model of Auditory Speech Processing,'' J. Phonetics **16**, 55–76.
Seneff, S. (**1985**). ''Pitch and Spectral Analysis of Speech Based on an Auditory Synchrony Model,'' Ph.D. thesis, MIT.
Stern, R. M. *et al.* (**1992**). ''Multiple Approaches to Robust Speech Recognition,'' in *Proc. ICSLP*.
Stevens, K. N. *et al.* (**1992**). ''Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters,'' J. Acoust. Soc. Am. **91**, 2979–3000.
Strevens, P. (**1960**). ''Spectra of fricatives noise in human speech,'' Lang. Speech **3**, 32–49.
Wang, K., and Shamma, S. (**1994**). ''Self normalization and noise robustness in early auditory representations,'' IEEE Trans. Speech Audio Process. **2**, 421–435.
Yu, J., and Cheah, U. W. (**2000**). ''Automatic speech recognition neural networks,'' Technical Report, University of Pennsylvania.
Zwicker, E. (**1961**). ''Subdivision of the audible frequency range into critical bands (Fequenzgrupper),'' J. Acoust. Soc. Am. **33**, 248.

Experimental validation of the use of Kramers–Kronig relations to eliminate the phase sheet ambiguity in broadband phase spectroscopy

Rebecca L. Trousil, Kendall R. Waters,[a] and James G. Miller
Laboratory for Ultrasonics, Department of Physics, Washington University, Saint Louis, Missouri 63130

(Received 18 August 2000; accepted for publication 16 February 2001)

The technique of broadband phase spectroscopy proposed in 1978 by Sachse and Pao [J. Appl. Phys. **49**, 4320–4327 (1978)] determines the phase velocity as a function of frequency from the Fourier transforms of a received reference and through-sample signal. Although quite successful, this approach can be influenced by an ambiguity in the phase velocity calculation which stems from the boundedness of the inverse tangent operation used to calculate phase. Several empirical approaches to resolve the phase ambiguity have been reported. An alternative approach that has not previously been considered appeals to the causal nature of the measurements. This article experimentally validates a method which uses the causally consistent Kramers–Kronig relations to eliminate the ambiguity in phase spectroscopy-derived phase velocity calculations. Broadband pulse and narrow-band tone burst measurements were performed on three gelatin-based phantoms containing different concentrations of graphite particles (0%, 10%, and 20% by volume). The phantoms were constructed to have attenuation coefficients which vary approximately linear-with-frequency, a dependence exhibited by many soft tissues. The narrow-band phase velocity measurements do not suffer from a phase ambiguity, and thus they serve as a ''gold standard'' against which the broadband phase velocity measurements are compared. The experimental results illustrate that using the Kramers–Kronig dispersion relations in conjunction with phase spectroscopy-derived phase velocity measurements is an effective means by which to resolve the phase sheet ambiguity in broadband phase spectroscopy.
[DOI: 10.1121/1.1365114]

PACS numbers: 43.80.Ev [FD]

I. INTRODUCTION

Since Sachse and Pao[1] introduced the technique of broadband phase spectroscopy in 1978, measurements of ultrasonic phase velocity obtained in this manner have been used to characterize many materials of biological and biomedical interest including soft tissues,[2–6] fluids,[7,8] ultrasonic contrast agents,[9] and bone.[10–13] This approach has also been applied to manmade materials to characterize ceramics[14] and polymers,[15–19] to quantify the effects of porosity in solids,[20–24] and to monitor the curing process in resins.[25,26] In this technique, the phase velocity of a sample as a function of frequency is determined from the Fourier transforms of reference and through-sample signals. Specifically, the phase of each signal is computed from the inverse tangent of the ratio of the imaginary to the real part of the frequency response, and the phase difference between the reference and through-sample signals is used to determine the phase velocity. This technique is particularly efficient when the frequency dependence of phase velocity (i.e., the dispersion) is of interest, because the phase velocity over the entire bandwidth can be deduced from a modest set of measurements.

The broadband phase spectroscopy technique, although quite successful, is at times limited by the manner in which the phase is calculated. Because this approach relies on the inverse tangent to calculate phase, the reference and through-sample phases can each only be determined to within an integer multiple of 2π. The phase velocity is derived from the phase difference of these two signals, and thus a factor of $2\pi\cdot(m-n)$ arises in the expression for the phase velocity where $2\pi n$ accounts for the phase ambiguity in the reference signal and $2\pi m$ accounts for the analogous ambiguity in the through-sample phase. The quantity $(m-n)$ will be referred to as the phase sheet offset and its indeterminacy the phase sheet ambiguity.

A number of empirical approaches have been implemented to address the phase sheet ambiguity in phase spectroscopy. One method, based on the work of Tribolet,[27] uniquely determines the phase spectra of the rf signals by extrapolating the bandlimited unwrapped phases to zero frequency.[12,28,29] An *ad hoc* resolution of the phase sheet ambiguity has been achieved by choosing the phase sheet offset that yields the phase velocity exhibiting minimum dispersion.[9,30–32] It has also been suggested that in some instances the phase sheet ambiguity can be circumvented if the signals are properly symmetrized prior to calculating the Fourier transforms.[33,34]

The method proposed and experimentally validated in this article introduces an alternative approach to resolve the phase sheet ambiguity in broadband phase spectroscopy based on the causal nature of the measurements. The causally consistent Kramers–Kronig relations, combined with the

[a]Current address: Laboratoire d'Imagerie Paramétrique, UMR 7623 CNRS–Paris VI, 15, rue de l'Ecole de Médecine, 75270 Paris CEDEX 06, France.

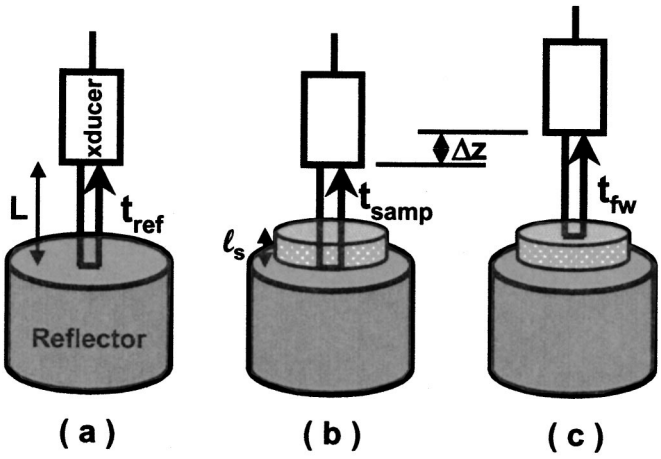


FIG. 1. Illustration of the time domain signals acquired for ultrasonic determination of sample thickness, phase velocity, and attenuation coefficient using a pulse-echo, immersion technique: (a) reference trace, and (b) through-sample trace. The additional front wall time trace (c) permitted the sample thickness to be measured ultrasonically.

measured amplitude attenuation coefficient (computed from the same signals acquired for phase spectroscopy), are used to predict the dispersion of the material of interest. A comparison of the Kramers–Kronig predicted dispersion with the frequency dependence of each of the possible phase velocity curves generated in broadband phase spectroscopy permits the correct phase sheet offset to be selected for phase spectroscopy-based measurements of phase velocity.

The article is organized in the following manner. In Sec. II, we begin by identifying the source of the phase sheet ambiguity in the phase spectroscopy technique. In Sec. III, we illustrate how the phase sheet ambiguity can be resolved from measurements of broadband attenuation and the Kramers–Kronig relations. In Sec. IV, the samples and the experimental techniques used to measure the phase velocity and attenuation coefficient are described. A comparison of narrow-band phase velocity measurements, phase spectroscopy-derived broadband phase velocity measurements, and Kramers–Kronig predicted dispersions in three gelatin-based phantoms are described in Sec. V. A discussion and summary follow in Secs. VI and VII, respectively.

II. THE PHASE SHEET AMBIGUITY IN BROADBAND PHASE SPECTROSCOPY

Broadband ultrasound spectroscopy techniques are used to estimate material properties from the Fourier transforms of time-localized signals. The original formalism derived expressions for the phase velocity and attenuation coefficient applicable for through transmission, contact transducer measurements.[1] The current study employs a pulse-echo, immersion setup (see Fig. 1). A summary of the broadband spectroscopy approach in the context of the current investigation is provided in the Appendix.

We begin by considering the difference in accumulated phase between a reference signal which has propagated a distance of $2L$ through a water path and a through-sample signal which has traveled the same distance except that a portion of the path has been replaced by a sample. From the arguments of the frequency responses given in Eqs. (A1) and (A2), the phase difference between the through-sample and reference signals, $\Delta\varphi(\omega)$, is

$$\Delta\varphi(\omega)\equiv\varphi_{\text{samp}}(\omega)-\varphi_{\text{ref}}(\omega)=2\omega l_s\cdot\left(\frac{1}{v_s(\omega)}-\frac{1}{v_h}\right), \tag{1}$$

where ω is angular frequency, l_s is the sample thickness, $v_s(\omega)$ is the frequency-dependent phase velocity of the sample, and v_h is the phase velocity in the host medium. The velocity in the host medium is treated as dispersionless.

Focusing on the experimental measurements required to obtain the phase difference in Eq. (1), we will illustrate how the phase sheet ambiguity arises. The reference and through-sample rf signals are digitized after oscilloscope trigger delays of $\tau_{\text{delay}}^{\text{ref}}$ and $\tau_{\text{delay}}^{\text{samp}}$, respectively. The digitized records are then symmetrized to reduce the rapidly varying linear-with-frequency phase ramp. Symmetrizing the reference and through-sample rf traces results in a second time shift for each signal denoted by $\tau_{\text{sym}}^{\text{ref}}$ and $\tau_{\text{sym}}^{\text{samp}}$, respectively. The details of the symmetrization process are described in Sec. IV B. The Fourier transforms of the symmetrized reference and through-sample rf signals are then computed, and the respective phases are determined from a four-quadrant inverse tangent of the ratio of the imaginary to real components of the Fourier transform. Because of the boundedness of the inverse tangent, the phases of the symmetrized rf signals, $\varphi_{\text{sym}}^{i}(\omega)$ where $i=$ ''ref'' or ''samp,'' can only be known to within integer multiples of 2π. Thus, the symmetrized phase contributions to the total phase accumulated for the reference and through-sample signals are $\varphi_{\text{sym}}^{\text{ref}}(\omega)+2\pi n$ and $\varphi_{\text{sym}}^{\text{samp}}(\omega)+2\pi m$, respectively. We remark that if the phase spectra were calculated from a two-quadrant inverse tangent, the phases would be ambiguous to integer multiples of π.

Making use of the Fourier shift theorem (using the $+i$ sign convention for the forward Fourier transform),[35] we write the total phase difference between the reference and through-sample signals as

$$\Delta\varphi(\omega)=[(\varphi_{\text{sym}}^{\text{samp}}(\omega)+2\pi m)-(\varphi_{\text{sym}}^{\text{ref}}(\omega)+2\pi n)]+\omega(\tau_{\text{delay}}^{\text{samp}}-\tau_{\text{delay}}^{\text{ref}})+\omega(\tau_{\text{sym}}^{\text{samp}}-\tau_{\text{sym}}^{\text{ref}}). \tag{2}$$

Equations (1) and (2) are combined to yield an expression for the phase velocity in the sample,

$$v_s(\omega)=\frac{2\omega l_s v_h}{2\omega l_s+v_h\cdot\{[\varphi_{\text{sym}}^{\text{samp}}(\omega)-\varphi_{\text{sym}}^{\text{ref}}(\omega)]+2\pi(m-n)+\omega(\tau_{\text{delay}}^{\text{samp}}-\tau_{\text{delay}}^{\text{ref}})+\omega(\tau_{\text{sym}}^{\text{samp}}-\tau_{\text{sym}}^{\text{ref}})\}}. \tag{3}$$

Because of the inherent ambiguity present when phase is calculated from an inverse tangent operation, the phase velocity is susceptible to an ambiguity when the phase spectroscopy technique is employed.

III. USE OF THE KRAMERS–KRONIG RELATIONS TO ELIMINATE THE PHASE SHEET AMBIGUITY

Based on the premise of causality and linearity, the ultrasonic Kramers–Kronig relationships connect the dispersion and attenuation coefficient in a material. If Kramers–Kronig relations can be formulated for the attenuation model of interest, the dispersion of the sample can be predicted from knowledge of the attenuation coefficient. We will show that the dispersion predicted from the Kramers–Kronig relations provides an additional piece of information which permits the phase sheet ambiguity in phase spectroscopy to be resolved.

It has been shown that dispersion relationships for media exhibiting linear-with-frequency attenuation coefficients can be obtained from differential dispersion relations[36] (e.g., the so-called nearly local Kramers–Kronig relations with one subtraction[37,38]). We remark, however, that in media for which the attenuation coefficient obeys a frequency power law but increases faster than the first power of frequency, two or more subtractions are necessary to formulate valid Kramers–Kronig relations.[39] The observation that the nearly local approximation breaks down when the attenuation increases faster than linear-with-frequency was reported by Szabo[40] and He.[41] Because the samples investigated in this study exhibit linear-with-frequency attenuation coefficients, the nearly local Kramers–Kronig relations with one subtraction will be employed to derive dispersion relations for these samples.

For media with attenuation coefficients of the form $\alpha(\omega)=\alpha(\omega_c)+(\beta/2\pi)(\omega-\omega_c)$ where β is the slope of attenuation and ω_c is the angular frequency at band center, the dispersion is approximated by

$$\frac{1}{v(\omega)}-\frac{1}{v(\omega_0)}=-\frac{2}{\pi}\cdot\frac{\beta}{2\pi}\ln\left(\frac{\omega}{\omega_0}\right), \tag{4}$$

where $v(\omega)$ is the frequency-dependent phase velocity, and ω_0 is, in general, any angular frequency within the useful bandwidth. When the dispersion is small, Eq. (4) can be expanded to first order in the dispersion so that

$$v(\omega)-v(\omega_0)\simeq\frac{[v(\omega_0)]^2\beta}{\pi^2}\ln\left(\frac{\omega}{\omega_0}\right). \tag{5}$$

Previous studies by this laboratory[37–39,42,43] and others[19,28,44] have predicted the dispersion of a sample from knowledge of the attenuation coefficient using ultrasonic Kramers–Kronig relations. The Kramers–Kronig dispersion relations have not, however, explicitly been used to address the issue of the phase sheet ambiguity in phase spectroscopy. We will show that comparing the predicted dispersion in Eq. (5) with the frequency dependence of the phase spectroscopy-derived phase velocity permits the correct phase sheet offset to be determined definitively.

IV. METHODS

A. Samples

Broadband phase velocity, broadband attenuation, and narrow-band phase velocity measurements were performed on three gelatin-based phantoms which were constructed to have attenuation coefficients that vary linearly with frequency. The phantoms are composed of 300 bloom granulated gelatin (Vyse Gelatin; Schiller Park, IL), 325 mesh graphite powder (Ventron; Danvers, MA), distilled water (Schnucks; Saint Louis, MO), Triton® X-100 surfactant (Aldrich Chemical; Milwaukee, WI), and 37% formaldehyde (Fisher Scientific; Fair Lawn, NJ). The phantoms were constructed to have 0%, 10%, and 20% graphite (by volume) uniformly distributed in the gelatin matrix. Twelve sites within a 12×8-mm^2 region of interest were insonified for each phantom. The nominal sample thickness ranged from 17 to 19 mm.

B. Broadband phase velocity and attenuation measurements

A 3.5-MHz center frequency, 2-in. focus, 0.5-in.-diam transducer (Panametrics V382; Waltham, MA) was excited by a broadband pulse from a pulser-receiver (Panametrics 5800; Waltham, MA) in pulse-echo mode. The reflected signals were amplified by 40 dB prior to being fed into an 8-bit digital oscilloscope (Tektronix 2440; Beaverton, OR). The time domain signals were temporally averaged 128 times to reduce random noise prior to being digitized at a 250-MHz sampling rate. Analysis was performed off-line. The three translational degrees of freedom associated with the transducer were adjusted with a motion controller (Aerotech Unidex 511; Pittsburgh, PA) under computer control.

In order to measure the sample thickness, phase velocity, and attenuation coefficient ultrasonically, three time domain signals were acquired at each site in a 12-site region of interest. Prior to data acquisition, the separation between the transducer and the steel reflector was varied until the maximum peak-to-peak signal was detected so as to place the reflector at the transducer focal distance. The reference signals were then acquired for a water-only propagation path of $2L$ [Fig. 1(a)]. A sample was then placed on the steel reflector, substituting a portion of the water path with sample, and the through-sample rf traces were acquired [Fig. 1(b)]. Finally, in order to measure the sample thickness ultrasonically at each site, the transducer was focused on the front wall of the sample, and the reflections from the front wall of the sample were digitized [Fig. 1(c)]. The temperature in the water tank and in the samples was maintained with a precision of ±0.1 °C throughout the course of the experiment in order to minimize temperature-dependent variations in the phase velocity.

An average host velocity (i.e., sound speed in water) was measured ultrasonically from correlations of broadband, time domain rf pulses which had propagated three known distances through the host medium. The broadband pulses were digitized at a sampling rate of 500 MHz which resulted in a 2 ns per point temporal resolution. A mean value for the

host velocity of 1487 ± 1 m/s was used in the calculations of sample thickness, phase velocity, and attenuation coefficient.

The sample thickness at each site in the region of interest was measured ultrasonically using

$$l_s = \frac{v_h}{2} \cdot (t_{\text{ref}} - t_{\text{fw}}) + \Delta z, \tag{6}$$

where t_{ref} is the water path only roundtrip time-of-flight between the transducer and the reflector, t_{fw} is the water path only roundtrip time-of-flight between the transducer and the front wall of the sample, and Δz is the distance the transducer was translated to change the focus from the reflector to the front wall of the sample. The timing difference was obtained from knowledge of the trigger delays for the reference and front wall rf signals and a correlation of the two digitized traces. Translation of the transducer over the distance Δz, which corresponded to approximately the sample thickness, was accomplished with the computer-controlled motion controller.

The broadband phase velocity was calculated at each site in the region of interest using Eq. (3). Calculation of the phases associated with the symmetrized rf waveforms was performed in the following manner. Prior to performing a fast Fourier transform of a time domain trace, the mean was subtracted from the rf signal to eliminate the dc component. The signal was then symmetrized by shifting the largest peak in the intensity of the time domain trace to the beginning of the record, wrapping earlier points to the end of the waveform. The frequency response of the signal was obtained by taking a fast Fourier transform of the symmetrized rf trace, and the phase was calculated from a four-quadrant inverse tangent of the ratio of the imaginary to the real part of the frequency response. This symmetrization choice was made for illustrative purposes which are described in detail in Sec. VI.

The phase that is extracted from the Fourier transform of a time domain signal is a wrapped phase (i.e., bounded from $-\pi$ to π) as discussed in Sec. II. In order to remove any 2π discontinuities from the phase spectrum, a phase unwrapping algorithm is applied to the data. The algorithm first unwraps the phase from a user-identified seed frequency within the bandwidth up to the upper end of the bandwidth and then returns to the seed frequency and unwraps the data down to the lower end of the bandwidth.

Once the unwrapped phases of the symmetrized reference and through-sample signals were obtained, the apparent phase velocity was calculated [Eq. (3)] at each site in the region of interest using three possible values for the phase sheet offset. The phase sheet offset which yielded the minimum dispersion was determined empirically, and then the two neighboring phase sheet offsets were also used to calculate the apparent phase velocity. For a given phase sheet offset, the mean phase velocity and standard deviation were computed for each sample based on measurements in a 12-site region of interest.

In order to predict the dispersion in the sample from the Kramers–Kronig relations, knowledge of the amplitude attenuation coefficient was needed. The attenuation coefficient in each sample was measured experimentally using a shadowed reflector technique which uses the same rf signals acquired for phase spectroscopy. The mean was subtracted from the reference and through-sample time domain signals, and then a 2.0-μs Hanning window was applied to the time domain traces to isolate the signal of interest. After calculating the fast Fourier transforms of the windowed rf signals, the amplitude attenuation coefficient was obtained from Eq. (A5). The average amplitude attenuation coefficient and standard deviation for each sample were calculated from the measured attenuation at each of the 12 sites in the region of interest. Because the sound speeds in the samples were similar to the velocity in the host medium, diffraction effects were considered to be negligible in this study. In general, however, the measured attenuation coefficient should be properly compensated for diffraction effects[45,46] prior to predicting the dispersion from the Kramers–Kronig relations.

A linear fit of the form $\alpha(\omega) = \alpha(\omega_c) + (\beta/2\pi)(\omega - \omega_c)$ was applied to the average attenuation coefficient data for each phantom to determine the slope of attenuation, β, which was subsequently used in Eq. (5) to predict the dispersion of the sample using the Kramers–Kronig relations. The attenuation coefficient is expressed in units of Np/cm and the slope of attenuation in units of Np/cm-MHz. The value of ω_c was taken to be the center of the useful bandwidth for each sample.

Because the Kramers–Kronig relations only predict the relative changes in phase velocity with frequency and not the absolute values, the phase spectroscopy-derived phase velocity at band center was used to set the absolute level of the Kramers–Kronig predicted dispersion [i.e., $\omega_0 = 2\pi f_c$ in Eq. (5)] so that the Kramers–Kronig prediction could be plotted on the same graph as the narrow-band and broadband phase spectroscopy measurements of phase velocity. Kramers–Kronig dispersion predictions and phase spectroscopy-derived phase velocity curves were generated for several phase sheet offsets, as described earlier. The frequency dependence of the phase spectroscopy-derived measurement of phase velocity which yielded the best agreement with the Kramers–Kronig predicted dispersion determines the correct phase sheet offset for the phase spectroscopy data.

C. Narrow-band phase velocity measurements

Narrow-band phase velocity measurements were performed in pulse-echo mode on each of the samples as a ''gold standard'' against which to compare the broadband estimates of phase velocity. The same 3.5-MHz transducer (Panametrics V382; Waltham, MA) used for the broadband measurements was used to generate the narrowband ultrasonic signals. A pulse/function generator (Hewlett-Packard HP8116A; Palo Alto, CA) created a sinusoidal tone burst which was fed into a diplexer (RITEC, Inc.; Warwick, RI) prior to exciting the transducer. The received signals were sent through a pulser/receiver (Panametrics 5800; Waltham, MA) in external pulser mode so that variable attenuation and 40 dB of gain could be applied prior to sending the signals to an oscilloscope (Tektronix 2440; Beaverton, OR). The received signals were digitized to 8 bits and stored off-line for

analysis. A pulse generator (Hewlett-Packard HP 8112A; Palo Alto, CA) was used as a master clock for the system.

All the narrow-band measurements of interest relied on time domain correlations of tone bursts. The number of cycles chosen for each tone burst varied with frequency in order to minimize correlation errors associated with the digitization rate and transients which occur at the beginning and end of the tone burst. For a 1024-point record with a temporal resolution of 4 ns/point, the tone burst lengths were chosen as follows: 6 cycles at 2 MHz, 9 cycles at 3 MHz, 12 cycles at 4 MHz, and 15 cycles at 5 MHz. These tone burst lengths prohibited signals which had traveled different paths from overlapping. A 4-in.-long piece of polished stainless steel was chosen as a reflector for these measurements to prevent echoes from the back surface of the reflector from interfering with the signals of interest.

The narrow-band data acquired paralleled the broadband measurements described in Sec. IV B except that reference, through-sample, and front wall time domain signals were acquired for 2-, 3-, 4-, and 5-MHz tone bursts individually. To minimize errors associated with measuring Δz, all the narrow-band data were acquired in a manner such that the transducer was translated vertically only once throughout the entire data acquisition process (see Fig. 1).

The technique used to measure the host velocity for the narrow-band measurements was essentially the same as that used for the broadband data described in Sec. IV B. The only differences were that a 3-MHz tone burst consisting of nine cycles was used for the narrow-band case, and the data were acquired at a sampling rate of 250 MHz which yielded a temporal resolution of 4 ns/point. The velocity measured at 3 MHz was chosen to be representative of the host velocity at all frequencies in the narrow-band study. An average host velocity of 1488 ± 1 m/s was used in the narrowband calculations of sample thickness and phase velocity. The 1-m/s difference between the host velocity measured for the broadband and narrow-band data sets is likely due to slight temperature variations.

The sample thickness and phase velocity were measured ultrasonically at each site in a 12-site region of interest from time-of-flight information from a reference trace [Fig. 1(a)], a through-sample trace [Fig. 1(b)], and a front wall trace [Fig. 1(c)].[47–49] The reference and through-sample rf traces were averaged temporally 256 times while the front wall signals were averaged 512 times prior to digitization to reduce random noise. For a given frequency, the sample thickness at each site in the region of interest was determined using Eq. (6), and subsequently the phase velocity was calculated from

$$v_s(\omega) = v_h \cdot \left[\frac{2 l_s}{v_h \cdot (t_{\text{samp}}(\omega) - t_{fw}) + 2\Delta z} \right]. \tag{7}$$

An average phase velocity and standard deviation were calculated at each frequency by averaging the measurements at each of the 12 sites in the region of interest.

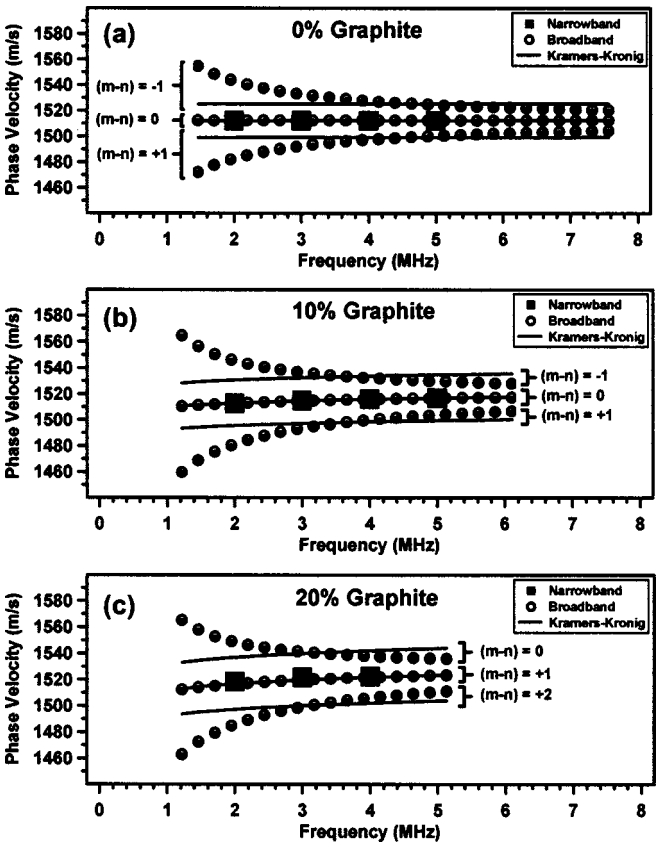


FIG. 2. Comparison of phase velocity results in three gelatin-based phantoms obtained from narrow-band tone burst measurements (filled squares), broadband phase spectroscopy measurements with three different phase sheet offsets (open circles), and the Kramers–Kronig predictions based on broadband attenuation measurements (lines).

V. RESULTS

Figure 2(a) contains a plot of the phase velocity in the 0% graphite phantom measured with narrow-band and broadband techniques. The ''gold standard'' narrow-band time-of-flight phase velocity measurements are reported in the filled squares at 2, 3, 4, and 5 MHz. The broadband data were reduced using two independent techniques. One type of analysis determined the broadband phase velocity using the phase spectroscopy approach (which suffers from the phase sheet ambiguity described in Sec. II). The graph shows broadband phase velocity curves (open circles) for phase sheet offsets corresponding to values of $(m-n)$ equal to -1, 0, and $+1$. The other broadband analysis technique calculated the dispersion from the measured attenuation coefficient using the Kramers–Kronig relation (lines). The absolute level of the Kramers–Kronig predictions was set by the phase spectroscopy-derived phase velocities at band center. These velocities and the other parameters used in the Kramers–Kronig predictions are provided in Table I. The shape of the phase spectroscopy-derived phase velocity curve which yields the best agreement with the Kramers–Kronig predicted dispersion corresponds to the correct phase sheet offset. From Fig. 2(a) it can be clearly seen that the correct phase sheet offset for the 0% graphite phantom corresponds to $(m-n)=0$. The broadband phase velocity data with a phase sheet offset of zero also agree well with the narrow-band velocity measurements.

TABLE I. Parameters used to calculate the Kramers–Kronig predictions of dispersion. The attenuation coefficient at band center, $\alpha(f_c)$ where $f_c = \omega_c/2\pi$, and the slope of attenuation, β, were obtained from linear fits about band center to the average attenuation coefficient data. The quantity $(m-n)$ is the phase sheet offset used to generate the phase spectroscopy-derived phase velocity curve from which $v(f_c)$ was obtained.

	f_c (MHz)	$\alpha(f_c)$ (Np/cm)	β (Np/cm-MHz)	$v(f_c)$ (m/s)	$(m-n)$
0% graphite phantom	4.6	+0.035	+0.010	1512.3	0
10% graphite phantom	3.7	+0.732	+0.189	1515.5	0
20% graphite phantom	3.2	+0.997	+0.315	1520.2	+1

Figure 2(b) shows the corresponding phase velocity results for the 10% graphite phantom while Fig. 2(c) shows the results for the 20% graphite phantom. For the 10% graphite phantom, the broadband phase spectroscopy calculations of phase velocity are shown for phase sheet offsets of -1, 0, and $+1$. Again the best agreement between the Kramers–Kronig predicted dispersions and the frequency dependence of the phase spectroscopy results occurs for a phase sheet offset of zero. This is further supported by the good agreement with the narrow-band measurements.

For the 20% graphite phantom, narrow-band measurements were performed at 2, 3, and 4 MHz but not a 5 MHz because the sample was too attenuating. For this sample, the phase spectroscopy-derived broadband phase velocity results are shown for phase sheet offsets of 0, $+1$, and $+2$. In this case, a phase sheet offset of $+1$ is appropriate for the 20% graphite phantom. The good agreement of the broadband phase velocity data for $(m-n)=+1$ with the narrow-band measurements provides an independent confirmation that the correct phase sheet offset has been successfully determined. We remark that from Fig. 2 it can be observed that the method of selecting a phase sheet offset which minimizes dispersion[9,30–32] does yield results consistent with the narrow-band and broadband data presented in this study, although this approach is *ad hoc* and not well understood.

The phase velocity measurements and predictions for all three of the phantoms are plotted on a smaller vertical scale in Fig. 3(a) to permit a more detailed examination of the agreement among the narrow-band measurements (filled symbols), the broadband phase spectroscopy measurements (open symbols), and the Kramers–Kronig predictions (lines). Standard deviation error bars are shown for the measured broadband and narrow-band phase velocity data based on measurements at 12 spatially independent sites in the sample (some error bars are smaller than the symbols used in the graph).

The open symbols in Fig. 3(b) show the average amplitude attenuation coefficient for the 0%, 10%, and 20% graphite phantoms based on measurements at 12 sites in each sample. The error bars represent the standard deviations. The lines on the graph are fits to the attenuation data of the form $\alpha(\omega)=\alpha(\omega_c)+(\beta/2\pi)(\omega-\omega_c)$. The center frequencies and the fit parameters for the three phantoms are provided in Table I. The value of β obtained from the curve fit and the angular frequency at band center were used in Eq. (5) to predict the dispersion in each sample from the Kramers–Kronig relations.

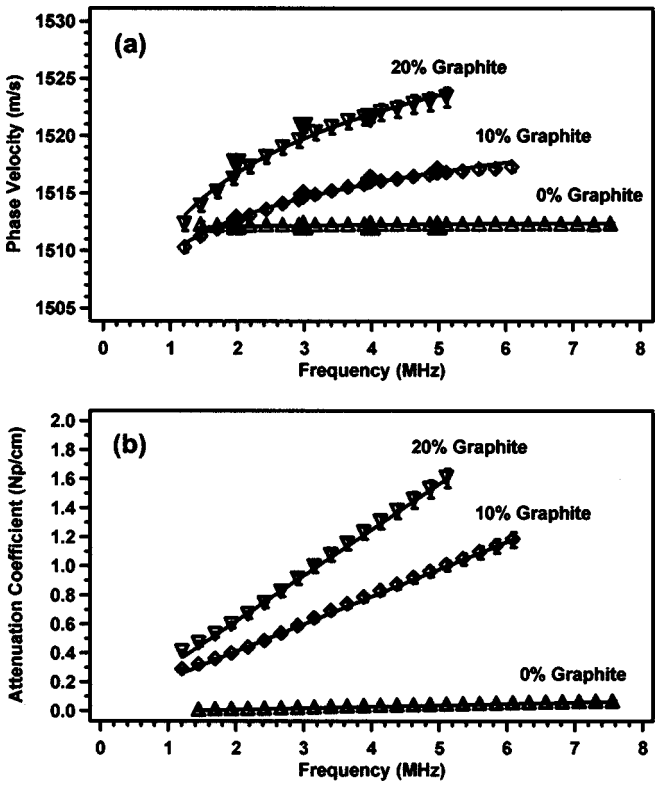


FIG. 3. (a) Illustration of the good agreement between the narrow-band phase velocity measurements (filled symbols), phase spectroscopy-derived velocity measurements (open symbols), and the Kramers–Kronig predicted dispersions (lines) for the 0%, 10%, and 20% graphite phantoms. (b) Measured broadband attenuation coefficients (open symbols) and fits (lines) of the form $\alpha(\omega)=\alpha(\omega_c)+(\beta/2\pi)(\omega-\omega_c)$ for the 0%, 10%, and 20% graphite phantoms. The slopes of attenuation obtained from the curve fits were used to predict the dispersion in the phantoms using the Kramers–Kronig relations.

VI. DISCUSSION

The phase sheet ambiguity is an inherent property of the broadband phase spectroscopy technique arising from the boundedness of the inverse tangent used to calculate the phase of the reference and through-sample signals. In practice, however, this ambiguity does not limit the usefulness of the phase spectroscopy technique. For instances where the phase velocity has been previously characterized with other techniques, that existing information can be used to infer the correct phase sheet offset. For a material whose velocity is not well characterized, techniques such as the ones discussed in this article exist which permit the correct phase sheet offset to be determined in phase spectroscopy-based measurements of phase velocity.

When the unwrapped reference and through-sample phases of symmetrized rf signals are on the same phase sheet so that $(m-n)=0$, the phase sheet offset term in Eq. (3) vanishes. This common occurrence can give the false impression that no ambiguity exists in the phase velocity calculation, and it may explain reports that the phase sheet ambiguity can be avoided if the rf signals are properly

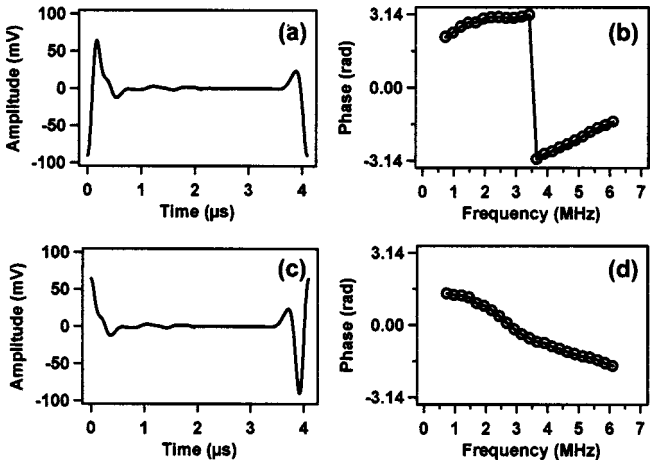


FIG. 4. Effect of the rf symmetrization criterion on phase calculation. (a) Representative rf trace for the 20% graphite phantom symmetrized by shifting the peak intensity in the rf to the beginning of the record, (b) the wrapped phase spectrum corresponding to the symmetrized rf trace in (a), (c) the same rf trace in (a) but symmetrized by shifting the peak *positive* value to the start of the record, and (d) the phase spectrum of the symmetrized rf trace in (c). Based on symmetry considerations, the likelihood of 2π discontinuities occurring within the bandwidth is minimized if the largest *positive* voltage in the rf trace is shifted to the beginning of the record.

symmetrized.[33,34] In fact, regardless of the value of the phase sheet offset, the phase sheet ambiguity always exists due to the boundedness of the inverse tangent used to calculate the phase spectra.

The phase sheet offset for a particular phase velocity calculation can depend on both the rf symmetrization and phase unwrapping routines employed. A common symmetrization method of shifting the peak intensity of the rf signal to the beginning of the record was performed in this study for illustrative purposes. This symmetrization choice often results in a continuous phase spectrum over the useful bandwidth; however, this is not always the case. A representative wrapped phase spectrum for the 20% graphite phantom, which exhibits a 2π discontinuity near band center, is shown in Fig. 4(b). The symmetrized rf from which the phase was calculated is plotted in Fig. 4(a). Because this through-sample signal exhibited symmetry somewhat similar to that of an inverted cosine, the phase at band center was close to $\pm\pi$, causing the phase to wrap near 3.5 MHz. An alternative symmetrization approach, which minimizes the likelihood of 2π discontinuities occurring in the phase spectrum by setting the phase at band center near zero, is to shift the largest *positive* peak of the rf signal to the start of the digitized record. The rf signal symmetrized in this manner is depicted in Fig. 4(c). The corresponding phase spectrum shown in Fig. 4(d) exhibits no discontinuities throughout the bandwidth.

Often symmetrization of the rf signal is sufficient to remove any discontinuities from the phase spectrum. However, for media which exhibit more marked dispersion or measurements over a sufficiently large frequency range, phase jumps larger than 2π may still occur within the bandwidth regardless of the symmetrization choice, necessitating the use of a phase unwrapping algorithm. Conventional phase unwrapping techniques unwrap phase from knowledge of the principal values of phase or from integration of the phase derivative. More sophisticated approaches, such as the adaptive integration technique proposed by Tribolet,[27] combine both approaches. All of these techniques require the phase to be known at some frequency in order to eliminate the phase ambiguity. Tribolet's approach defines the phase at zero frequency, whereas the phase unwrapping algorithm employed in this study sets the phase at a user-defined frequency within the bandwidth of the experiment where reliable phase estimates are expected. In both cases, the unwrapped phase spectra are still ambiguous to an integer multiple of 2π because the frequency at which the phase was defined is itself ambiguous to a multiple of 2π. This may not be readily apparent in the adaptive integration method because a comparison of phases (which could be on different phase sheets) was not considered.

Although the approach of using Kramers–Kronig predicted dispersions in conjunction with phase spectroscopy-derived measurements of phase velocity to eliminate the phase sheet ambiguity is valid for causal systems, it is often only practical if analytic expressions of the Kramers–Kronig dispersion relations are available for the attenuation model under investigation. The dispersion relations used in this study [Eqs. (4) and (5)] are applicable for media with attenuation coefficients that vary linear-with-frequency. (We remark, however, that recent reports[10,12,13] suggest that the dispersion in bone from 200 to 800 kHz may not be described adequately by these dispersion relations despite the approximately linear-with-frequency attenuation coefficient exhibited by bone.) Other work from this laboratory has formulated and experimentally validated general Kramers–Kronig relations for media exhibiting frequency power law attenuation coefficients,[39] an attenuation model applicable to many biological and man-made materials. For media with more complicated attenuation coefficients, the Kramers–Kronig relations can be evaluated numerically to predict the dispersion. The dispersion of Albunex has been obtained from this alternative approach.[42]

There are a number of advantages of using the Kramers–Kronig predicted dispersion in conjunction with the phase spectroscopy measurements to eliminate the phase sheet ambiguity in phase spectroscopy. The Kramers–Kronig predicted dispersion can be calculated from the same rf signals acquired for phase spectroscopy, so no additional data need be acquired. Additionally, this approach relies on a comparison of the frequency dependence of the phase velocity, providing a more robust metric than, for example, comparing the calculated broadband phase velocity to a narrowband calculation of phase velocity at a single frequency. Finally, because it is based on the causal nature of the experimental system, this approach definitively eliminates the phase sheet ambiguity rather than appealing to an *ad hoc* or empirical determination of the phase sheet offset.

VII. SUMMARY

We have introduced a method based on the causal consistency of the Kramers–Kronig relations which eliminates the phase sheet ambiguity present in phase velocity measurements obtained from the broadband phase spectroscopy tech-

nique. The Kramers–Kronig predicted dispersion is compared with the frequency dependence of the phase velocity measured with the phase spectroscopy technique to determine the correct phase sheet offset. This method was validated experimentally by comparing the broadband phase velocity measurements with a corresponding set of narrowband tone burst measurements on three graphite-gelatin phantoms exhibiting different phase velocity and attenuation characteristics. The good agreement between the two independent data sets suggests that the combination of Kramers–Kronig predicted dispersion with phase spectroscopy-based phase velocity measurements can successfully eliminate the phase sheet ambiguity in phase velocity measurements derived from frequency domain techniques.

ACKNOWLEDGMENTS

The authors wish to express their appreciation to former members of this laboratory: Dr. J. Mobley, Dr. M. S. Hughes, and Dr. G. H. Brandenburger. Their early intellectual contributions to broadband phase spectroscopy and Kramers–Kronig relationships made this study possible. This work was supported in part by NIH Grant No. HL40302.

APPENDIX: SUMMARY OF BROADBAND SPECTROSCOPY FORMALISM IN THE CONTEXT OF A PULSE-ECHO, IMMERSION EXPERIMENT

The broadband spectroscopy technique is a frequency domain method for calculating the phase velocity and attenuation coefficient in a sample from knowledge of the frequency response of a reference and through-sample signal. The following development is appropriate for a pulse-echo, immersion experiment where the sample is placed in contact with the reflector.

We begin by stating expressions for the frequency responses of the two signals of interest. The frequency response of the reference signal [Fig. 1(a)] can be written as

$$\begin{aligned}\tilde{U}_{\mathrm{ref}}(x=2L,\omega) &= [r^p_{h\to r}\cdot A(\omega)\cdot \exp(-\alpha_h(\omega)\cdot 2L)]\\ &\quad\cdot \exp(i(k_h(\omega)\cdot 2L+\zeta(\omega)))\\ &= |\tilde{U}_{\mathrm{ref}}(x=2L,\omega)|\cdot \exp(i\varphi_{\mathrm{ref}}(\omega)),\end{aligned} \quad \text{(A1)}$$

where $r^p_{h\to r}$ is the pressure amplitude reflection coefficient at the interface between the host medium (subscript ''h'') and the reflector (subscript ''r''), $A(\omega)$ and $\zeta(\omega)$ describe the transfer function of the transducer and associated electronics, $\alpha_h(\omega)$ is the amplitude attenuation coefficient in the host medium, L is the distance between the transducer and the reflector, and $k_h(\omega)=\omega/v_h$ is the wave number in the host medium. Although, in general, reflection coefficients are complex quantities, in this study they are all approximately real because $\alpha/k \ll 1$ for the samples investigated. The frequency response of the through-sample signal [Fig. 1(b)] received at the transducer is expressed as

$$\begin{aligned}\tilde{U}_{\mathrm{samp}}(x=2L,\omega) &= [(t^p_{h\to s}\cdot r^p_{s\to r}\cdot t^p_{s\to h})\cdot A(\omega)\\ &\quad\cdot \exp(-\alpha_h(\omega)\cdot 2(L-l_s)-\alpha_s(\omega)\cdot 2l_s)]\\ &\quad\cdot \exp(i(k_h(\omega)\cdot 2(L-l_s)+k_s(\omega)\cdot 2l_s\\ &\quad+\xi(\omega)))\\ &= |\tilde{U}_{\mathrm{samp}}(x=2L,\omega)|\cdot \exp(i\varphi_{\mathrm{samp}}(\omega)),\end{aligned} \quad \text{(A2)}$$

where $t^p_{i\to j}$ is the pressure amplitude transmission coefficient from medium i to medium j (the subscript ''s'' refers to the sample), l_s is the sample thickness, $\alpha_s(\omega)$ is the attenuation coefficient in the sample, and $k_s(\omega)=\omega/v_s(\omega)$ is the wave number in the sample.

The phase velocity in the sample is obtained from the phases of the reference and through-sample frequency responses in Eqs. (A1) and (A2), respectively,

$$v_s(\omega)=v_h\cdot\left[\frac{2\omega l_s}{2\omega l_s+v_h(\varphi_{\mathrm{samp}}(\omega)-\varphi_{\mathrm{ref}}(\omega))}\right], \quad \text{(A3)}$$

where v_h is the host velocity, $\varphi_{\mathrm{samp}}(\omega)$ is the phase of the through-sample signal, and $\varphi_{\mathrm{ref}}(\omega)$ is the phase associated with the reference trace. Equation (A3) is the starting point for the discussion on the source of the phase sheet ambiguity in phase spectroscopy.

The amplitude attenuation coefficient in the sample is obtained from the magnitudes of the reference and through-sample frequency responses in Eqs. (A1) and (A2), respectively,

$$\begin{aligned}\alpha_s(\omega) &= \alpha_h(\omega)+\frac{1}{2l_s}\cdot\ln\left[\frac{t^p_{h\to s}\cdot r^p_{s\to r}\cdot t^p_{s\to h}}{r^p_{h\to r}}\right]\\ &\quad+\frac{1}{2l_s}\cdot\ln\left[\frac{|\tilde{U}_{\mathrm{ref}}(\omega)|}{|\tilde{U}_{\mathrm{samp}}(\omega)|}\right].\end{aligned} \quad \text{(A4)}$$

Because the attenuation coefficient of the host medium (i.e., water) is typically much less than that of the sample, and the reflection and transmission coefficients were approximately unity, the attenuation coefficient was estimated from

$$\alpha_s(\omega)=\frac{1}{2l_s}\cdot\ln\left[\frac{|\tilde{U}_{\mathrm{ref}}(\omega)|}{|\tilde{U}_{\mathrm{samp}}(\omega)|}\right] \quad \text{(A5)}$$

for the samples in this investigation.

[1] W. Sachse and Y. H. Pao, ''On the determination of phase and group velocities of dispersive waves in solids,'' J. Appl. Phys. **49**, 4320–4327 (1978).

[2] H. Hachiya, S. Ohtsuki, M. Tanaka, and F. Dunn, ''Determination of sound speed in biological tissues based on frequency-analysis of pulse response,'' J. Acoust. Soc. Am. **92**, 1564–1568 (1992).

[3] J. E. Perrin, G. Berger, A. Salesses, and M. Agneray, ''Methods of tissue characterization with ultrasonic spectroscopy, analysis of the results obtained, prospects for clinical application in a short time,'' C. R. Seances Soc. Biol. Fil. **173**, 345–358 (1979).

[4] F. G. Sommer, P. Stetson, H. S. Chen, R. A. Stern, D. J. Rachlin, and A. Macovski, ''Prospects for ultrasonic spectroscopy and spectral imaging of abdominal tissues,'' J. Ultrasound Med. **12**, 83–90 (1993).

[5] H. Heynemann, K. P. Richter, and R. Millner, ''Ultrasonic spectroscopic examinations of pathologically altered testes,'' Z. Exp. Chir. **15**, 103–107 (1982).

[6] A. F. van der Steen, C. L. de Korte, and J. M. Thijssen, ''Ultrasonic spectroscopy of the porcine eye lens,'' Ultrasound Med. Biol. **20**, 967–974 (1994).
[7] P. K. Choi, J. R. Bae, and K. Takagi, ''Ultrasonic spectroscopy in bovine serum albumin solutions,'' J. Acoust. Soc. Am. **87**, 874–881 (1990).
[8] O. V. Moskalenko, ''Possibilities of ultrasonic spectroscopy of the blood in the diagnosis of early postoperative inflammatory complications in patients with stomach cancer,'' Klin Khir **3**, 38–39 (1998).
[9] J. Mobley, M. S. Hughes, J. N. Marsh, C. S. Hall, G. H. Brandenburger, and J. G. Miller, ''Broadband measurements of phase velocity in Albunex suspensions,'' J. Acoust. Soc. Am. **103**, 2145–2153 (1998).
[10] P. H. F. Nicholson, G. Lowet, C. M. Langton, J. Dequeker, and G. van der Perre, ''A comparison of time-domain and frequency-domain approaches to ultrasonic velocity measurement in trabecular bone,'' Phys. Med. Biol. **41**, 2421–2435 (1996).
[11] J. Wu and F. Cubberley, ''Measurement of velocity and attenuation of shear waves in bovine compact bone using ultrasonic spectroscopy,'' Ultrasound Med. Biol. **23**, 129–134 (1997).
[12] P. Droin, G. Berger, and P. Laugier, ''Velocity dispersion of acoustic waves in cancellous bone,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**, 581–592 (1998).
[13] K. A. Wear, ''Measurements of phase velocity and group velocity in human calcaneus,'' Ultrasound Med. Biol. **26**, 641–646 (2000).
[14] H. F. Wang, W. H. Jiang, and W. W. Cao, ''Characterization of lead zirconate titanate piezoceramic using high frequency ultrasonic spectroscopy,'' J. Appl. Phys. **85**, 8083–8091 (1999).
[15] I. Alig and D. Lellinger, ''Ultrasonic spectroscopy for measurement of phase velocity and attenuation at high frequencies in polymer systems,'' Acustica **76**, 273–282 (1992).
[16] D. Zellouf, Y. Jayet, N. Saint-Pierre, J. Tatibouet, and J. C. Baboux, ''Ultrasonic spectroscopy in polymeric materials. Application of the Kramers–Kronig relations,'' J. Appl. Phys. **80**, 2728–2732 (1996).
[17] J. Wu, ''Determination of velocity and attenuation of shear waves using ultrasonic spectroscopy,'' J. Acoust. Soc. Am. **99**, 2871–2875 (1996).
[18] J. Mobley, C. S. Hall, J. N. Marsh, M. S. Hughes, K. R. Waters, G. H. Brandenburger, and J. G. Miller, ''Measurements and predictions of the phase velocity and attenuation coefficient in suspensions of elastic microspheres,'' J. Acoust. Soc. Am. **106**, 652–659 (1999).
[19] P. He, ''Experimental verification of models for determining dispersion from attenuation,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **46**, 706–714 (1999).
[20] R. A. Kline and H. Saghatchi, ''Grain size and porosity effects in simulated geological samples,'' Proceedings of IEEE Ultrasonics Symposium **85CH2209-5**, 1043–1046 (1985).
[21] N. Brown, B. Castagnede, W. Lauriks, and M. Melon, ''Experimental study of the dispersion of ultrasonic waves in porous materials,'' C. R. Acad. Sci. Ser. II **319**, 393–399 (1994).
[22] H. Jeong and D. K. Hsu, ''Experimental analysis of porosity induced ultrasonic attenuation and velocity change in carbon composites,'' Ultrasonics **33**, 195–203 (1995).
[23] D. J. Roth, J. D. Kiser, S. M. Swickard, S. A. Szatmary, and D. P. Kerwin, ''Quantitative mapping of pore fraction variations in silicon nitride using an ultrasonic contact scan technique,'' Res. Nondestruct. Eval. **6**, 125–168 (1995).
[24] A. Wanner, ''Elastic modulus measurements of extremely porous ceramic materials by ultrasonic phase spectroscopy,'' Mater. Sci. Eng., A **248**, 35–43 (1998).
[25] S. I. Rokhlin, D. K. Lewis, K. F. Graff, and L. Adler, ''Real-time study of frequency dependence of attenuation and velocity of ultrasonic waves during the curing reaction of epoxy resin,'' J. Acoust. Soc. Am. **79**, 1786–1793 (1986).
[26] M. Matsukawa and I. Nagai, ''Ultrasonic characterization of a polymerizing epoxy resin with imbalanced stoichiometry,'' J. Acoust. Soc. Am. **99**, 2110–2115 (1996).
[27] J. M. Tribolet, ''A new phase unwrapping algorithm,'' IEEE Trans. Acoust., Speech, Signal Process. **25**, 170–177 (1977).
[28] W. A. Verhoef, M. J. T. M. Cloostermans, and J. M. Thijssen, ''Diffraction and dispersion effects on the estimation of ultrasound attenuation and velocity in biological tissues,'' IEEE Trans. Biomed. Eng. **32**, 521–529 (1985).
[29] V. N. Gupta, P. K. Bhagat, and A. M. Fried, ''Estimating ultrasound propagation velocity in tissues from unwrapped phase spectra,'' Ultrason. Imaging **2**, 223–231 (1980).
[30] M. S. Hughes, ''Applications of local approximations to the Kramers–Kronig relations in ultrasonics,'' Ph.D. dissertation, Washington University, St. Louis, MO, 1987.
[31] J. Mobley, ''Ultrasonic dispersion in suspensions and solids: A study of fundamental dynamics and the Kramers–Kronig relations,'' Ph.D. dissertation, Washington University, Saint Louis, MO, 1998.
[32] K. R. Waters, ''On the application of the generalized Kramers–Kronig dispersion relations to ultrasonic propagation,'' Ph.D. dissertation, Washington University, Saint Louis, MO, 2000.
[33] P. He, ''Direct measurement of ultrasonic dispersion using a broadband transmission technique,'' Ultrasonics **37**, 67–70 (1999).
[34] P. He, ''Determination of ultrasonic parameters based on attenuation and dispersion measurements,'' Ultrason. Imaging **20**, 275–287 (1998).
[35] E. O. Brigham, *The Fast Fourier Transform* (Prentice-Hall, Englewood Cliffs, NJ, 1974).
[36] J. B. Bronzan, G. L. Kane, and U. P. Sukhatme, ''Obtaining real parts of scattering amplitudes directly from cross section data using derivative analyticity relations,'' Phys. Lett. **49B**, 272–276 (1974).
[37] M. O'Donnell, E. T. Jaynes, and J. G. Miller, ''General relationships between ultrasonic attenuation and dispersion,'' J. Acoust. Soc. Am. **63**, 1935–1937 (1978).
[38] M. O'Donnell, E. T. Jaynes, and J. G. Miller, ''Kramers–Kronig relationship between ultrasonic attenuation and phase velocity,'' J. Acoust. Soc. Am. **69**, 696–701 (1981).
[39] K. R. Waters, M. S. Hughes, J. Mobley, G. H. Brandenburger, and J. G. Miller, ''On the applicability of Kramers–Kronig relations for ultrasonic attenuation obeying a frequency power law,'' J. Acoust. Soc. Am. **108**, 556–563 (2000).
[40] T. L. Szabo, ''Causal theories and data for acoustic attenuation obeying a frequency power law,'' J. Acoust. Soc. Am. **97**, 14–24 (1995).
[41] P. He, ''Simulation of ultrasound pulse propagation in lossy media obeying a frequency power law,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**, 114–125 (1998).
[42] J. Mobley, K. R. Waters, M. S. Hughes, C. S. Hall, J. N. Marsh, G. H. Brandenburger, and J. G. Miller, ''Kramers–Kronig relations applied to finite bandwidth data from suspensions of encapsulated microbubbles,'' J. Acoust. Soc. Am. **108**, 2091–2106 (2000).
[43] M. S. Hughes, S. M. Handley, and J. G. Miller, ''Nearly local Kramers–Kronig relations applied to porous epoxy,'' Proc. of IEEE Ultrasonics Symposium **87CH2492-7**, 1041–1045 (1987).
[44] C. C. Lee, M. Lahham, and B. G. Martin, ''Experimental verification of the Kramers–Kronig relationship for acoustic waves,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **37**, 286–294 (1990).
[45] M. A. Fink and J. F. Cardoso, ''Diffraction effects in pulse-echo measurement,'' IEEE Trans. Sonics Ultrason. **31**, 313–329 (1984).
[46] B. Zeqiri, ''Validation of a diffraction correction model for through-transmission substitution measurements of ultrasonic absorption and phase velocity,'' J. Acoust. Soc. Am. **99**, 996–1001 (1996).
[47] B. D. Sollish, ''A device for measuring ultrasonic propagation velocity in tissue,'' National Bureau of Standards, Special Publication 525 (1979), pp. 53–56.
[48] I. Y. Kuo, B. Hete, and K. K. Shung, ''A novel method for the measurement of acoustic speed,'' J. Acoust. Soc. Am. **88**, 1679–1682 (1990).
[49] D. K. Hsu and M. S. Hughes, ''Simultaneous ultrasonic velocity and sample thickness measurement and application in composites,'' J. Acoust. Soc. Am. **92**, 669–675 (1992).

Simulations of the thermo-acoustic lens effect during focused ultrasound surgery

Ibrahim M. Hallaj[a)]
Radiology Department, Brigham & Women's Hospital, Harvard Medical School, 221 Longwood Avenue, Boston, Massachusetts 02115

Robin O. Cleveland
Department of Aerospace and Mechanical Engineering, Boston University, 110 Cummington Street, Boston, Massachusetts 02215

Kullervo Hynynen
Radiology Department, Brigham & Women's Hospital, Harvard Medical School, 221 Longwood Avenue, Boston, Massachusetts 02115

(Received 3 March 2000; revised 6 July 2000; accepted 5 February 2001)

Laboratory measurements of soft tissue properties show a dependence of background propagation properties on temperature. For typical focused ultrasound surgery (FUS) applications, only the slow variations in tissue background parameters need to be accounted for when computing the outcome of a FUS sonication. The cumulative effect of slowly varying sound speed has been referred to in the literature as a thermal lens, or a thermo-acoustic lens because of its beam-distorting properties. An algorithm to solve the coupled acoustic-thermal problem is described, and numerical results are presented to illustrate the effects of dynamic sound-speed profiles in layered tissues undergoing FUS. The results of simulations in liver with and without a fat layer indicate that the thermal-acoustic interaction results in more complex dynamics in FUS than a simple model will predict. Both the size and the position of the lesions predicted from the simulations are affected by the thermo-acoustic lens effect. However, the overall effect from short sonications at high power from sharply focused single element sources (F-no. from 0.8 to 1.3) around 1 MHz similar to those used in clinical setups is found to be small. [DOI: 10.1121/1.1360239]

PACS numbers: 43.80.Gx, 43.80.Sh, 43.25.Cb, 43.25.Ed [FD]

I. INTRODUCTION

Minimally invasive tumor coagulation using focused ultrasound devices holds great promise in treating patients without the need for traditional surgical procedures.[1–3] The heating of soft tissues for therapeutic purposes has been demonstrated experimentally[4–7] and explained theoretically.[6,8–10]

Modeling of FUS systems is important for proper selection of device operating parameters and geometries. Numerical models provide a means of exploring thermal therapy treatment planning as well as system design. Models have been presented describing the behavior of focused ultrasonic beams in biological tissue[11,12] and the thermal effect of such beams.[13–16] Experiments in phantoms and *in vivo*[17–20] have been conducted to verify these models. Soft tissue is commonly modeled as a lossy fluid media,[15,8] and the main effects of ultrasound on soft tissues can be described assuming compressional waves in fluids.

Some interesting phenomena have been observed in biomedical acoustics which suggest revisiting the traditional models when dealing with time-varying tissues. One such phenomenon is the thermo-acoustic lensing which occurs in heated tissue, where a gradual movement in the position of the focus and a change in the local index of refraction occurs. This phenomenon may be partially explained as resulting from the temperature-dependence of the sound speed in tissue.[21–23] The acoustic intensity field generated from focused sources results in spatially inhomogeneous temperature fields in otherwise homogeneous tissue samples. Modeling such behavior requires using inhomogeneous equations to describe the acoustic and the thermal problems, and a mechanism for feedback between them.

In this paper an acoustic model incorporating inhomogeneity, absorption, and nonlinearity is used in conjunction with the bioheat transfer equation (BHTE) to predict acoustic and temperature fields in layered tissues with temperature-dependent sound speeds. The present paper deals with the coupling of the acoustic and the thermal problems via the sound speed of layered tissue in thermal therapy. While experimental data for tissue parameters as a function of temperature and thermal dose remain scarce, the data published for sound speed in fat and liver as a function of temperature are used to guide the time-domain simulations presented.

II. MODEL DESCRIPTION

A. The acoustic model

A second-order wave equation for finite-amplitude propagation in a thermo-viscous fluid is used as a model for

[a)]Author to whom correspondence should be addressed, currently with Wolf, Greenfield & Sacks, P.C., 600 Atlantic Ave., Boston, MA 02210; electronic mail: ibrahim@alumni.washington.edu

TABLE I. Base values of acoustic parameters used in the study (37 °C).

Material	ρ (kg m^{-3})	c_0 (m s^{-1})	α_0 (Np m^{-1} MHz^{-1})	β_0
Water	1000	1500	2.88×10^{-4}	3.5
Liver	1050	1596	4.5	6.0
Fat	910	1430	9.0	10.5

TABLE II. Base values of thermal parameters used in the study (37 °C).

Material	k (W m^{-1} K^{-1})	C (J kg^{-1} K^{-1})	W (kg m^{-3} s^{-1})	T_a (°C)
Water	0.60	4180	0	37
Liver	0.50	3700	0.5	37
Fat	0.50	3700	0.5	37

ultrasound propagation in this study. The equation is a generalized Westervelt equation, described in Hamilton and Morfey,[24]

$$\nabla^2 p - \frac{1}{c_0^2}\frac{\partial^2 p}{\partial t^2} - \frac{1}{\rho}\nabla p\cdot\nabla\rho + \frac{\delta}{\rho c_0^4}\frac{\partial^3 p}{\partial t^3} + \frac{\beta}{\rho c_0^4}\frac{\partial^2 p^2}{\partial t^2} = 0. \tag{1}$$

The acoustic pressure p was calculated from the wave equation, given the material properties of the medium. The background density and sound speed of the medium are ρ and c_0, respectively. The nonlinearity coefficient, β, is related to the nonlinearity parameter, B/A, by $\beta = 1 + (B/2A)$,[25] and the acoustic diffusivity, δ, accounts for both thermal and viscous losses in the fluid, and is related to the absorption coefficient by $\delta = 2\alpha c_0^3/\omega^2$, where α is the acoustic absorption coefficient and ω is the angular frequency. It is assumed that the time rates of change of the background parameters in Eq. (1) are slow enough so that the extra terms in the wave equation for time-varying media are negligible. This slow time-varying background assumption has been shown to be reasonable for typical FUS applications.[26]

For this study, the numerical solution of the wave equation in polar cylindrical coordinates is accomplished using the finite-difference time-domain (FDTD) method as described in Ref. 27. The partial derivatives in Eq. (1) are discretized to second-order accuracy in time and fourth-order accuracy in space with absorbing boundary conditions at the computational domain's edges. Table I gives the baseline values used for the acoustic parameters in Eq. (1). The data used in this study for the tissue parameters can be found in the literature, for example, in Goss *et al.*[28]

B. The thermal model

The temperature increase in soft tissues can be modeled by a diffusion type equation. The Pennes bioheat transfer equation (BHTE)[13] is used to describe the thermal effects of ultrasound in tissue:

$$\frac{\partial T}{\partial t} = \frac{k_t}{\rho C_t}\nabla^2 T - \frac{W_b C_b}{\rho C_t}(T - T_a) + \frac{Q}{\rho C_t}, \tag{2}$$

where T is the temperature in the tissue, k_t is the thermal conductivity of the tissue, C_t and C_b are the heat capacities of tissue and blood, respectively, W_b is the perfusion, T_a is the ambient (arterial) temperature (37 °C), and Q is the heat deposition source term due to the acoustic field. The heat source term was calculated as described by Pierce,[29]

$$Q = \frac{\delta}{\rho_0 c_0^4}\left\langle\left(\frac{\partial p}{\partial t}\right)^2\right\rangle. \tag{3}$$

Q was time-averaged over an acoustic period by numerical integration for use in the bioheat equation. Table II gives the reference values used for the parameters in Eq. (2).

The BHTE was solved in cylindrical coordinates using the FDTD method to the same accuracy as the wave equation. Constant flux boundary conditions were assumed for the edges of the computational domain to simulate an extended thermal region beyond the computational boundaries of the simulations, as constant temperature boundary conditions can create artifacts at the edges of the domain.

The temperature field was used to calculate thermal dose according to the formula

$$t_{43} = \int_{t=t_0}^{t=t_{\text{final}}} R^{(43-T(t))}\,dt. \tag{4}$$

after Sapareto and Dewey.[30] Here t_{43} is the thermal dose equivalent time at 43 °C, and t_0 and t_{final} are the starting and ending times of the sonication, and R is 0.5 if $T \geqslant 43$ °C, and 0.25 if $T < 43$ °C. Models commonly consider 240 min at 43 °C the threshold for lesion formation in soft tissue. This assumption has been tested against measured lesion sizes and found to be reasonable in soft tissues.[6]

C. Tissue response model

The data for sound speed as a function of temperature were compiled from measurements by Bamber and Hill.[31] These data were adapted for the present study by using one data point per 5 °C according to the trends apparent from the published values. The results are shown in Fig. 1. The polynomials used in this study to fit the sound speeds in fat and liver to the experimental data are

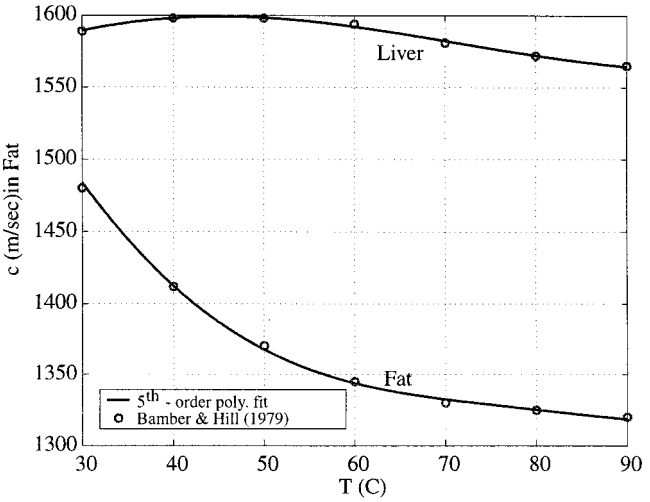


FIG. 1. Temperature dependence of sound speed in soft tissue taken from published laboratory measurements (symbols) and the corresponding polynomial fits to the data (solid).

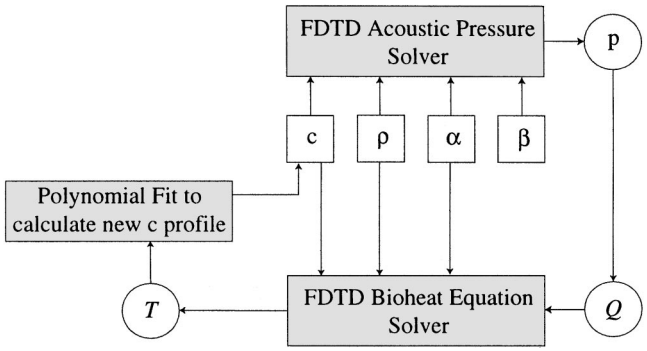


FIG. 2. Flowchart showing the iterative method for coupling the pressure and temperature calculations in the time-varying tissue simulations.

$$c_{\text{fat}}(T) = 1746 - 3.931T - 0.4128T^2 + 1.137\times 10^{-2}T^3 - 1.101\times 10^{-4}T^4 + 3.731\times 10^{-7}T^5 \text{ m/s}, \quad (5)$$

$$c_{\text{liver}}(T) = 1529 + 1.686T + 6.113\times 10^{-2}T^2 - 2.297\times 10^{-3}T^3 + 2.266\times 10^{-5}T^4 - 7.179\times 10^{-8}T^5 \text{ m/s}. \quad (6)$$

The polynomial expressions are only valid inside the interpolation region, 30 °C to 90 °C.

These data imply that during the course of a single focused ultrasound treatment of finite duration the sound speed will change, and this change is considered in subsequent calculations of the acoustic and thermal fields.

D. Acousto-thermal coupling

The flowchart in Fig. 2 shows how the acoustic and the thermal solvers were coupled for the present computations via the heating term $Q(\mathbf{x},t)$ in Eq. (3). Periodically the thermal solver was made to calculate and output the updated background sound speed of the tissue to data files. The sound speeds were calculated using the polynomial fits described above at each point in space on the computational domain. The updated $c(\mathbf{x},t)$ profile was then used as an input to the acoustic solver to calculate a new pressure field in an iterative fashion. Thus the pressure field, which drives the heating, was in turn affected by the temperature field, and so forth. The tissue parameters and CW pressure field were updated for the simulations presented here every 2 s unless otherwise noted. This update period was determined to be sufficient for the FUS problems based on a convergence test in which the periodicity of updating was varied.

III. DESCRIPTION OF THE SIMULATIONS

The basic layout for the simulations is shown in Fig. 3 and involves a single element spherical section bowl transducer in water with axis of symmetry, *z*. The bowl has a radius of curvature which varied between 3.2 and 5.2 cm, an aperture diameter of 4 cm, and frequency of 1.0 MHz unless otherwise noted. *F*-numbers between 0.8 and 1.3 were used in the simulations to mimic the focused source conditions used in FUS devices. A simulation was also carried out for a 1.5 MHz source with an *F*-number of 1.0 for comparison. Another simulation spanning 60 s was done to assess the

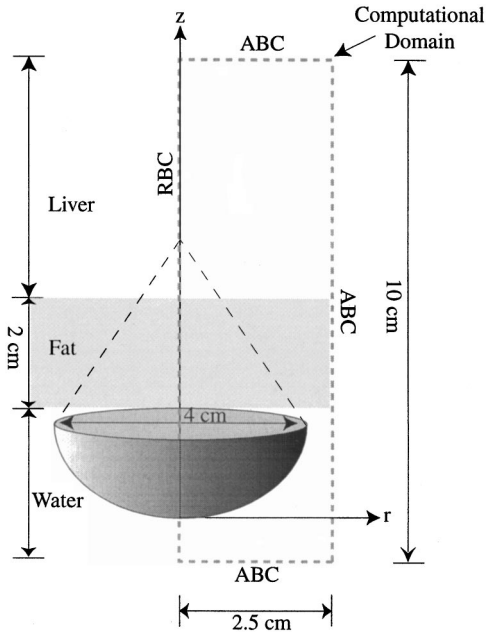


FIG. 3. Configuration used for the simulations showing the relative locations of the focused source and the water, fat, and liver layers. ABC denotes absorbing boundary conditions, and RBC denotes reflected symmetry boundary conditions. Not to scale.

effect of longer sonications in which prefocal heating can be more important than the brief high-power sonications. All sonications were adjusted for source pressure so that the final peak temperature at the focus was 80 ± 1 °C. This was based on the typical desired heating pattern for individual FUS sonications in clinical situations, chosen to avoid tissue boiling.

The calculations were carried out on a uniform grid with spatial discretization of 0.1 mm along both the radial and the axial directions. The computational domain spans a 10 cm (axial) by 2.5 cm (radial) area. However, due to the axial symmetry of the problem, the results are correct for a rotation of this 2-D space about the axis. The acoustic simulations used a time step of 10 ns, while the thermal simulations used a time step of 0.1 s. The reason for the disparity being that the acoustic FDTD code needed to resolve much shorter time scales associated with an acoustic period while the thermal FDTD code was used to compute heat transfer phenomena at much longer time scales.

The acoustic pressure, intensity, temperature, and sound-speed scalar fields were tracked and stored at 2 s intervals for a total run time of 10 s in the following simulations unless otherwise noted. The acoustic code was run until a steady-state CW field was obtained on the computational domain (about 100 μs). This technique of using absorbing boundary conditions to obtain CW results from transient codes provided good convergence and has been shown to be successful for acoustic problems.[32] The BHTE code was then run using the most up-to-date values calculated for sound speed of 2 s, then the acoustic code was run again until steady state was reached, and so on. The results provide a sequence of pressure and temperature fields, as well as the evolving sound-speed profiles of the inhomogeneous tissues.

The source was simulated to be in water at 37 °C, and

TABLE III. Simulation parameters, including source powers, source pressures, and focal intensities which were required to achieve 80 °C peak temperature at the end of each CW sonication.

f (Mhz)	F-no.	Duration (s)	cm Fat	Power (W)	p_{source} (MPa)	I_{focus} (W m^{-2})
1.0	0.8	10	0	66.1	0.38	3183
1.0	0.8	10	2	78.2	0.41	3764
1.0	1.0	10	0	82.1	0.43	2650
1.0	1.0	10	2	98.1	0.47	3166
1.5	1.0	10	2	65.2	0.38	4749
1.0	1.0	60	2	60.5	0.37	1951
1.0	1.3	10	2	128.1	0.55	2565
1.0	1.3	10	3.5	140.1	0.57	2565

projecting a beam into soft (liver) tissue either directly adjacent to the water or separated from the water by a fat layer. Each of the layers was taken to be initially homogeneous and at 37 °C. Upon insonation by the acoustic field the tissue layers' sound speeds were assumed to evolve in space and time according to Eqs. (5) and (6). The water temperature and properties were kept constant, assuming that a regulated supply of cooling water was available. Table III shows the values of the pressures, powers, and intensities used for the simulations. For each case, the source pressure was arrived at by an iterative process using interpolation such that an 80 °C peak temperature was achieved at the focus at the end of the sonication, as stated earlier. The parameters and dimensions in Table III were considered representative of possible clinical therapeutic ultrasound treatment systems and tissue samples.

Separate simulations were conducted which did not take into account the acousto-thermal lensing for comparison. In these simulations the water and tissue layers were sonicated for the same duration with the same sources and powers, but the sound speeds were never altered from their baseline values during the heating.

IV. RESULTS

A. Short (10 s) CW sonications at 1.0 MHz, *F* no.=0.8, 1.0

Simulation results are presented as axial profiles of peak pressure, temperature, and sound speed evolving over the 10 s sonications. The profiles in Figs. 4–7 are plotted at 2 s intervals to illustrate the dynamics of the problem for the 1 MHz, F no.$=0.8$ and F no.$=1.0$ cases with and without a 2 cm thick fat layer.

Several main observations can be made from the peak pressure slices: First, the location of the peak pressure (the acoustic focus) drifts slightly in time toward the source. The position of the peak pressure does not drift by more than 2 mm for any of the 1 MHz, 10 s runs, and is only a fraction of a millimeter for the cases with no fat layer. The drift is due to the refraction of the rays as they propagate through the thermo-acoustic lens. The rate of drift appears to decrease over the course of a simulation because the temperature rise levels off during a sonication as conduction and perfusion act to limit the temperature rise. Standing waves were seen in the water and in the fat layer. No standing waves were seen in the liver tissue, as it was assumed to be unbounded using the absorbing boundary conditions. From the temperature data we can see the drift of the hot spot toward the transducer as time progresses, mirroring the change in the acoustic pressure field.

The sound-speed variation was most remarkable in the fat layer, where the temperature dependence was most pronounced. The sound speed in fat underwent a change from its baseline value of 1430 m/s down to 1330 m/s near the focus after 10 s of heating. The main observations from the sound-speed profiles from the simulations are the fact that generally the sound speeds in tissue and liver (above 45 °C) drop near the focus as the temperature increases. A slight increase was seen at the periphery of the focal zone as the liver was heated only slightly, and was on the $c(T)$ region of the curve in which sound speed climbs to a local maxima near 45 °C. Also noted was the way in which the sound speed in the fat

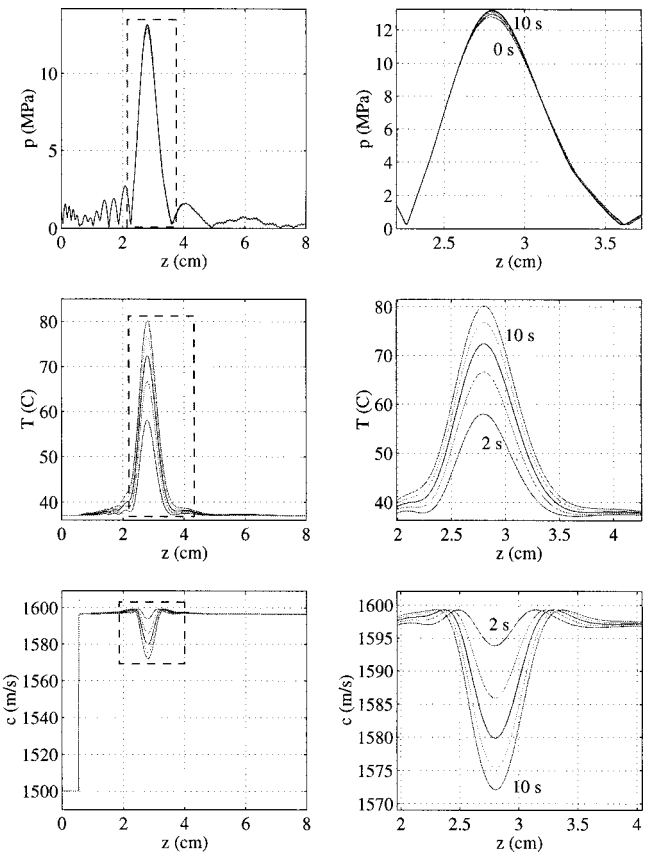


FIG. 4. The evolution of (a) p, (b) T, and (c) c_0 for the 1.0 MHz, F-no. $=0.8$ case with no fat layer, having 10 s duration. The curves on the right show enlarged version of the curves on the left in the vicinity of the focus (dotted outlines). Source power$=66.1$ W.

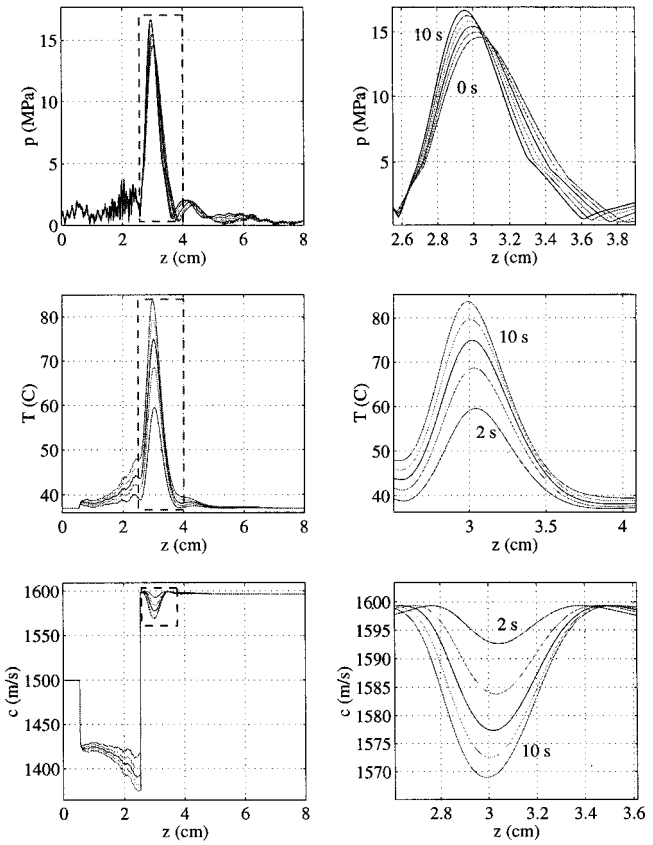


FIG. 5. The evolution of (a) p, (b) T, and (c) c_0 for the 1.0 MHz, F-no. $=0.8$ case with a 2 cm fat layer, having 10 s duration. The curves on the right show enlarged version of the curves on the left in the vicinity of the focus (dotted outlines). Source power$=78.2$ W.

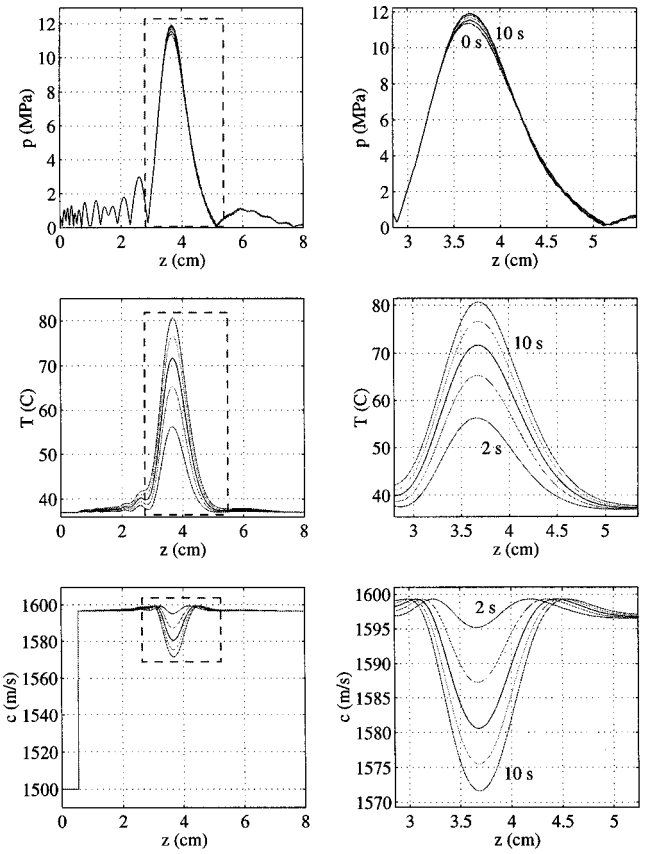


FIG. 6. The evolution of (a) p, (b) T, and (c) c_0 for the 1.0 MHz, F-no. $=1.0$ case with no fat layer, having 10 s duration. The curves on the right show enlarged version of the curves on the left in the vicinity of the focus (dotted outlines). Source power$=82.1$ W.

layer changed readily with the relatively mild heating, according to the polynomial interpolation given in Eq. (6).

B. Short sonications at higher frequency (1.5 MHz)

Simulations using a 1.5 MHz source with similar geometry to the previously described $F1.0$ source were carried out. This source resulted in very little thermo-acoustic lensing effect. We attribute this primarily to the higher focusing ability of high-frequency sources. In this case the fraction of energy heating the prefocal region and fat was smaller than for the 1.0 MHz runs of similar geometry due to the tighter focus. Even with a 2 cm fat layer in the prefocal region, this source resulted in a focal shift of about 0.5 mm only (Fig. 8).

C. Short sonications at lower gain (F no.$=1.3$)

Simulations were also carried out using a 1.0 MHz source similar to those described previously, but having a longer radius of curvature (5.2 cm) for a F-number of 1.3. In these simulations both a 2 cm thick fat layer as well as a thicker 3.5 cm fat layer could be used. The results from these simulations are shown in Figs. 9 and 10. We see that for thicker fat layers more thermo-acoustic lensing occurs, especially when the focus is near the liver–fat interface. This is because when the focus is near the liver–fat interface, more acoustic energy is deposited in the fat, especially near the interface, and increased heating of the prefocal fat tissue occurs.

Some other data from the 3.5 cm fat layer simulation are shown for their illustrative value. Figure 11(a) presents a cross-axis view of the temperature near the focus at 10 s for runs with and without thermal lensing included. The thermal lens effect contributed an extra 10% to the peak temperature at the focus in this case. The hot spot's position and peak temperature were tracked and are presented in Fig. 11(b), which shows how the location of the maximum temperature moved toward the source in time. Finally, in Fig. 12, the 1 min and 240 min outlines of the calculated thermal dose at 43 °C are presented. The 1 min dose lines are presented to show the onset of multiple lesion development can occur at locations other than the intended focus in some cases, this being enhanced by the thermal lensing effect. The lesions formed with thermal lensing included were less elongated, more peaked, and were closer to the transducer than simulations without thermal lensing predicted.

D. Long (60 s) CW sonications

A set of simulations was carried out for the 1.0 MHz, F no.$=1.0$ transducer, but at a lower power (60.5 W), such that without thermal lensing feedback the peak temperature after 60 s reached 80 °C. This resulted in more noticeable thermal lensing than the brief sonications described in the last sections. The comparison made was between a simulation in which the thermo-acoustic lens effect was modeled, and one in which the background sound speed of the tissue

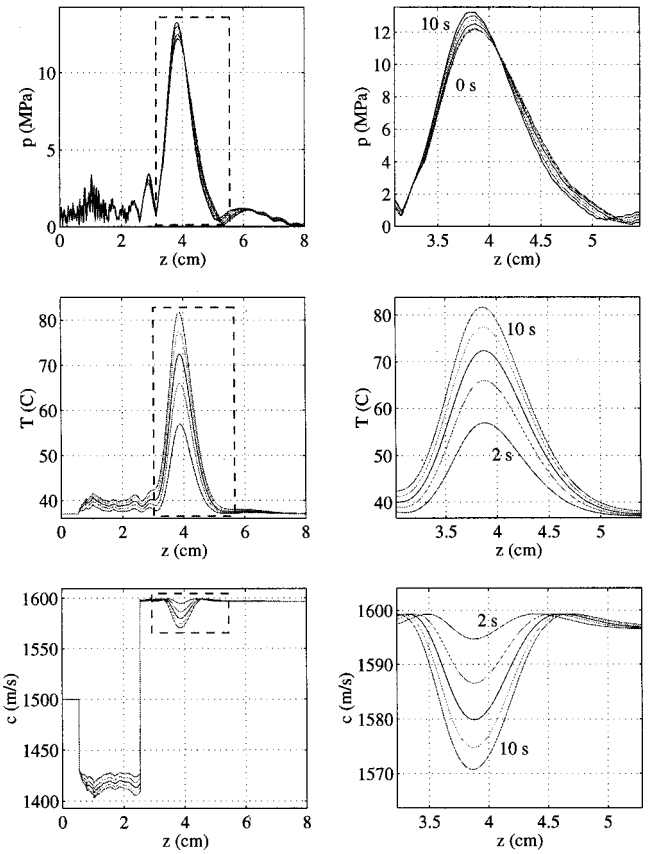


FIG. 7. The evolution of (a) p, (b) T, and (c) c_0 for the 1.0 MHz, F-no. $=1.0$ case with a 2 cm fat layer, having 10 s duration. The curves on the right show enlarged version of the curves on the left in the vicinity of the focus (dotted outlines). Source power$=98.1$ W.

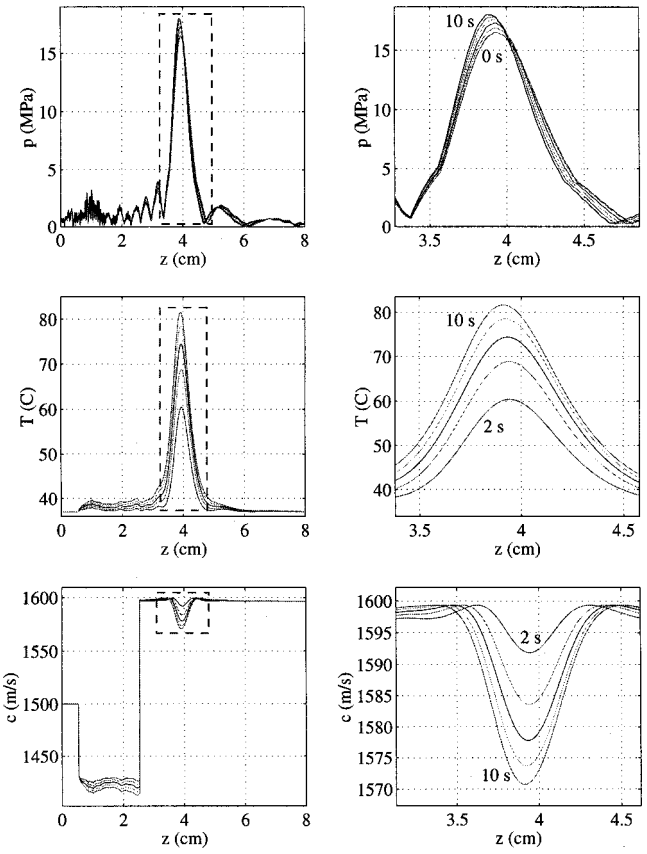


FIG. 8. The evolution of (a) p, (b) T, and (c) c_0 for the 1.5 MHz, F-no. $=1.0$ case with a 2 cm fat layer, having 10 s duration. The curves on the right show enlarged version of the curves on the left in the vicinity of the focus (dotted outlines). Source power$=65.2$ W.

layers was held constant at its baseline value. Figure 13 shows the axial peak pressure, temperature, and sound speed. We see that the peak pressure is shifted by about 2 mm toward the source when acousto-thermal lensing occurs. The magnitude of the peak pressure also increases as the shape of the focal spot is changed. We also note an increase in peak temperature from a 60 s sonication in the presence of thermo-acoustic lensing. The location of the pressure and temperature maxima shifted about 2 mm overall.

Figure 14(a) shows the contours around the 240 min equivalent doses for the 60 s runs. Note that for this predictor of lesioning, very little movement or size change can be seen between simulations with and without the thermal lensing phenomenon. This is due to the similar temperature history for the sonications with and without thermal lensing included. The lesion predicted by the 240 min contours showed a slight (1 mm) movement toward the transducer and a slight change in shape. In Fig. 14(b) we compare the 240 min dose contours for the case without thermal lensing described above with a case where 60 s of cooling followed the 60 s of heating. The cooling stage made only a small contribution to the total volume of the predicted lesion. Finally, in Fig. 14(c) the temperature history of the geometric focus is shown for the case with no thermal lensing for a 60 s sonication followed by 60 s of cooling.

V. DISCUSSION

Simulations of acoustic pressure and temperature fields for focused ultrasound devices were performed showing the effect of varying sound speeds in tissues as a function of their temperature. Thermo-acoustic lensing effects were observed, and lead to a small migration of the focal hot spot during an FUS treatment. In addition, the peak temperatures were higher than those predicted by static models under similar conditions. The effects were due to the enhanced refraction of acoustic rays through the inhomogeneous media layers. The magnitude of the peak pressure and the peak temperature were also found to gradually increase for a steady CW source pressure as the thermo-acoustic lens developed. The maximum shift in the position of the focal spot for single brief (10 s) sonications was under 2 mm toward the source. The maximum difference in predicted peak temperature compared to the same simulations run without thermo-acoustic lensing was about 4 °C, or about a 10% difference in the temperature rise.

Specific effects noted were the effect of including a fat layer in front of the liver tissue, increasing the frequency by 50%, increasing the radius of curvature, and running a longer, lower power simulation. The main factors in producing thermo-acoustic lensing are the amount of prefocal heating which occurs along the propagation path, and the response of the medium along this path to heating. The higher F-number and the slow long sonications resulted in more

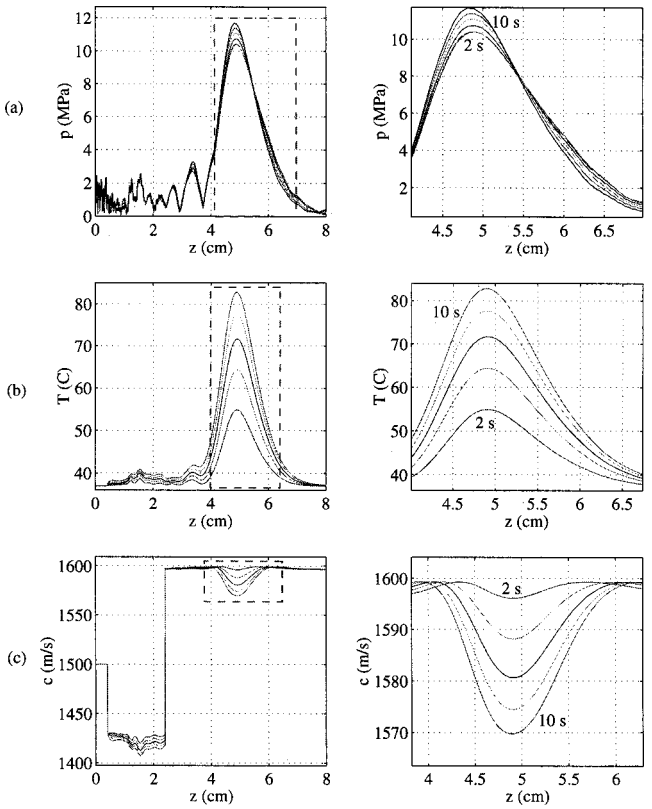


FIG. 9. The evolution of (a) p, (b) T, and (c) c_0 for the 1.0 MHz, F-no. $=1.3$ case with a 2 cm fat layer, having 10 s duration. The curves on the right show enlarged version of the curves on the left in the vicinity of the focus (dotted outlines). Source power$=128.1$ W.

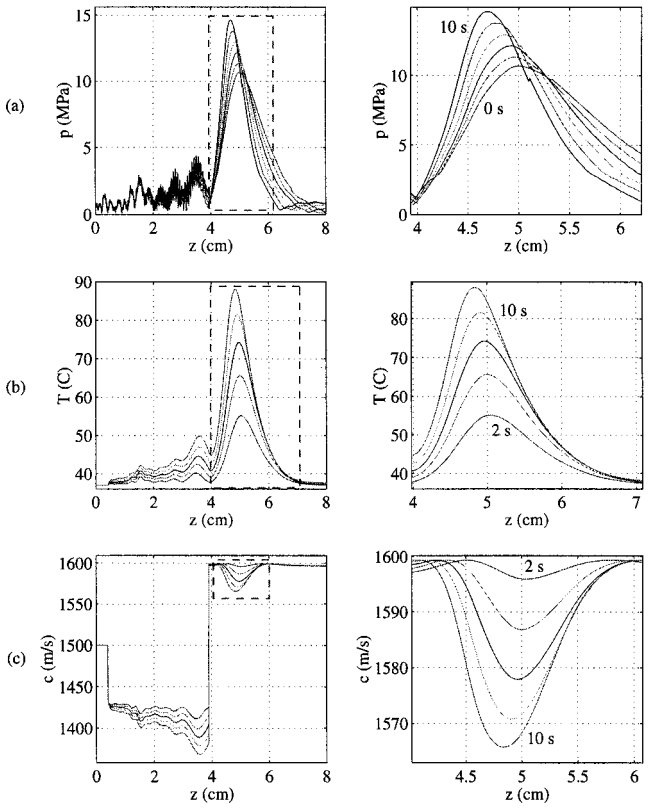


FIG. 10. The evolution of (a) p, (b) T, and (c) c_0 for the 1.0 MHz, F-no. $=1.3$ case with a 3.5 cm fat layer, having 10 s duration. The curves on the right show enlarged version of the curves on the left in the vicinity of the focus (dotted outlines). Source power$=140.1$ W.

pronounced thermal lensing. This effect was enhanced by the presence of the fat layer. In these simulations the fat acted to exaggerate the lensing because of the rapid descent in its sound speed as a function of temperature. Without the fat layer for short sonications almost no noticeable lensing was produced by the liver tissue alone (fractions of a mm). The increased frequency or decreased F-number resulted in reduced near-field heating and thermal lensing. We expect that sources with large focal spots that require longer sonications to achieve lesioning will result in more pronounced thermo-acoustic lensing. The same can be expected for phased array systems,[33,7] which induce large lesions, for they will subject the prefocal tissue to more integrated thermal dose and heating over the duration of a FUS treatment.

Our model assumed that the dependence of sound speed on temperature was explicit, and used previously published data for the sound speed in fat and liver as a function of temperature to construct an interpolation polynomial which was periodically used to update the tissue sound-speed profiles. This model served as a starting point for investigating the phenomena of slowly varying tissue, and more work is needed to understand the mechanisms of change in tissue. Other time-varying material parameters can be modeled and inserted into the calculation loop. Data for variation of absorption coefficient are available for a limited set of conditions, but due to variations in measurement techniques, quantitative interpretation of this data remains unclear. Generally speaking, the absorption coefficient has been observed to increase with thermal dose and temperature.[34] A preliminary study showed that the increase in local absorption near the focal region accelerated the heating rate and accompanying thermo-acoustic lens effect, and the increase was dependent on the temperature rise and the functional dependence of

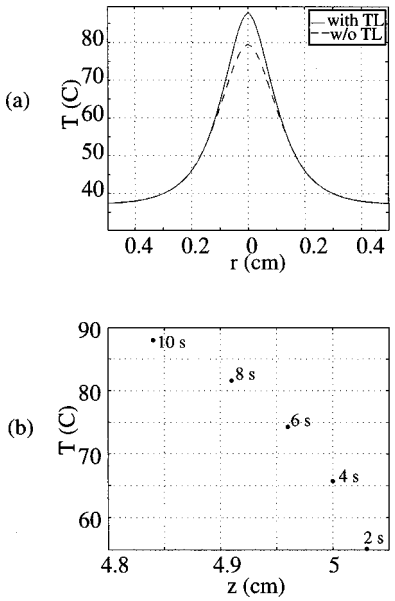


FIG. 11. (a) Slices of the peak temperature in the focal planes at 10 s for the $f=1.0$ MHz F-no.$=1.3$ case with 3.5 cm of fat with and without thermal lensing effect. (b) Movement of the peak temperature tracked at 2 s intervals for the $f=1.0$ MHz F-no.$=1.3$ case with 3.5 cm of fat and thermal lensing included.

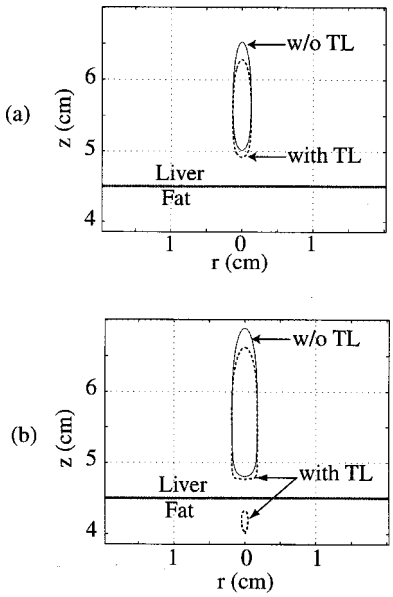


FIG. 12. Calculated thermal dose contours comparing the results from the $f=1.0$ MHz F-no.$=1.3$ with 3.5 cm of fat simulations with and without thermal lensing included. (a) The $t_{43}=240$ min dose outlines, (b) the $t_{43}=1$ min dose outlines.

absorption on temperature.[26] Equation (3) helps to explain how an acceleration of the thermo-acoustic interaction arises if α increases with temperature, because as the tissue is heated α would increase, leading to a higher heating rate,

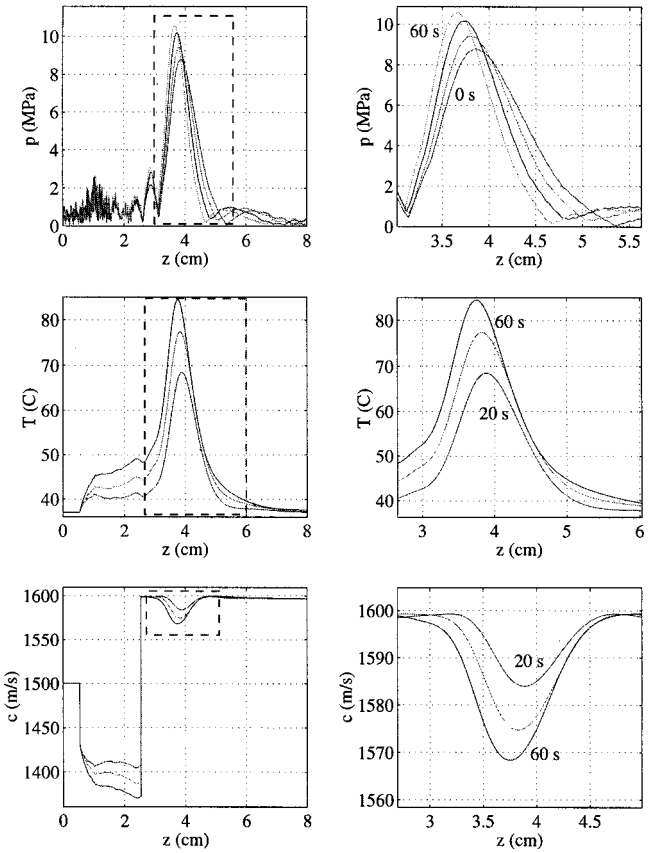


FIG. 13. The evolution of (a) p, (b) T, and (c) c_0 for the 1.0 MHz, F-no. $=1.0$ case with a 2 cm fat layer, having 60 s duration. The curves on the right show enlarged version of the curves on the left in the vicinity of the focus (dotted outlines). Source power$=60.5$ W.

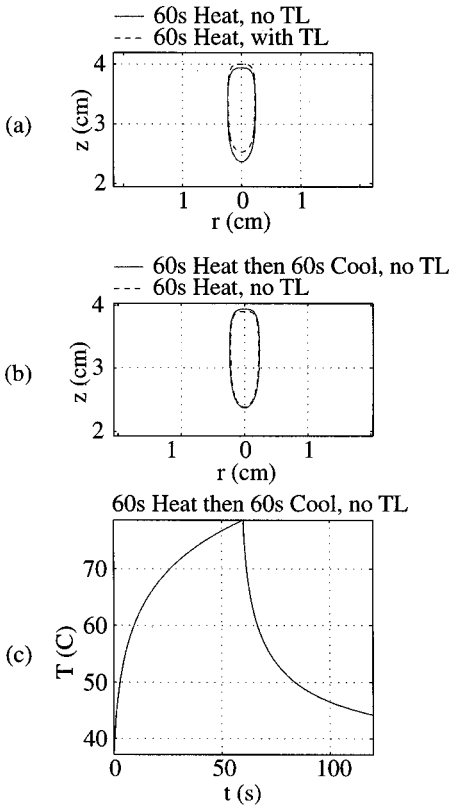


FIG. 14. The outlines of the calculated $t_{43}=240$ min dose curves (a) with and without thermal lensing, (b) with and without subsequent cooling, (c) the temperature at the focus, for the long sonication case with $f=1$ MHz, F-no.$=1.0$ at 60.5 W.

and faster rising temperatures. However, because the effect would be most significant in a local area near the focus only, it is not expected that this effect would significantly contribute to thermal lensing. Data for other properties, such as the temperature-dependence of the nonlinearity coefficient and the density, are not readily available for most tissues.

In conclusion, the dependence of sound speed on tissue temperature appears to have a minor impact on the coagulated tissue volume or its location for brief sonications from single-focus sources having F-numbers around 1.0 and frequencies around 1.0 MHz, as commonly used in clinical ultrasound surgery systems. This result may not apply to multiple-focus arrays or scanned systems, which produce much higher near-field thermal doses. In such systems, the prefocal regions experience the effects of repeated over lapping sonications or scans, while the smaller focal zones are only heated by their respective individual sonications. Since the thermal lens effect is a cumulative path effect of propagation from the source to the focus, it is most significant for propagation through a long path of heated tissue, such as the prefocal tissue in scanned or multifocal array systems.

ACKNOWLEDGMENT

This work was supported in part by the National Cancer Institute (Grant No. R01:CA46627).

[1] F. J. Fry, ''Intense focused ultrasound in medicine,'' Eur. Urol. **23**(Suppl I), 2–7 (1993).
[2] N. Sanghvi and R. Hawes, ''High intensity focused ultrasound,'' Exper. & Investigational Endoscopy **4**, 383–395 (1994).
[3] C. Hill and G. ter Haar, ''Review article: High intensity focused ultrasound- potential for cancer treatment,'' Br. J. Radiol. **68**, 1296–1303 (1995).
[4] K. Hynynen, R. Roemer, E. Moros, C. Johnson, and D. Anhalt, ''The effect of scanning speed on temperature and equivalent thermal exposure distributions during ultrasound hyperthermia *in vivo*,'' IEEE Trans. Microwave Theory Tech. **34**, 552–559 (1986).
[5] G. ter Haar, ''Ultrasound focal beam surgery,'' Ultrasound Med. Biol. **3**, 1089–1100 (1995).
[6] C. Damianou, K. Hynynen, and X. Fan, ''Evaluation of accuracy of a theoretical model for predicting the necrosed tissue volume during focused ultrasound surgery,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**, 182–187 (1995).
[7] D. Daum, N. Smith, R. King, and K. Hynynen, ''*In vivo* demonstration of noninvasive thermal surgery of the liver and kidney using an ultrasonic phased array,'' Ultrasound Med. Biol. **25**, 1087–1098 (1999).
[8] J. Wu and G. Du, ''Temperature elevation in tissues generated by finite-amplitude tone bursts of ultrasound,'' J. Acoust. Soc. Am. **88**, 562–1577 (1990).
[9] C. Hill, I. Rivens, M. Vaughan, and G. ter Haar, ''Lesion development in focused ultrasound surgery: A general model,'' Ultrasound Med. Biol. **20**, 259–260 (1994).
[10] X. Fan and K. Hynynen, ''Ultrasound surgery using multiple sonications—Treatment time considerations,'' Ultrasound Med. Biol. **22**, 471–481 (1996).
[11] X. Fan and K. Hynynen, ''The effect of wave reflection and refraction at soft tissue interfaces during ultrasound hyperthermia treatments,'' J. Acoust. Soc. Am. **91**, 1727–1736 (1992).
[12] H. Wan, P. VanBaren, E. Ebbini, and C. Cain, ''Ultrasound surgery: comparison of strategies using phased array systems,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **43**, 1085–1098 (1996).
[13] H. Pennes, ''Analysis of tissue and arterial blood temperature in the resting human forearm,'' J. Appl. Phys. **1**, 93–122 (1948).
[14] P. Lele and A. Pierce, ''The thermal hypothesis of the mechanism of ultrasonic focal destruction in organized tissues,'' Proceedings of the workshop on Interaction of Ultrasound and Biological Tissues, Bureau of Radiological Health, FDA 73-8008, pp. 121–128 (1973).
[15] W. Nyborg, ''Heat generation by ultrasound in a relaxing medium,'' J. Acoust. Soc. Am. **70**, 310–312 (1981).
[16] M. Curley, ''Soft tissue temperature rise caused by scanned, diagnostic ultrasound,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **40**, 59–66 (1993).
[17] J. Wu, J. Chase, and T. Holzapfel, ''Temperature rise in a tissue-mimicking material generated by unfocused and focused ultrasonic transducers,'' Ultrasound Med. Biol. **5**, 495–512 (1992).
[18] K. Hynynen, ''Demonstration of enhanced temperature elevation due to nonlinear propagation of focussed ultrasound in dog's thigh *in vivo*,'' Ultrasound Med. Biol. **13**, 85–91 (1987).
[19] M. Kolios, A. Worthington, M. Sherar, and J. Hunt, ''Experimental evaluation of two simple thermal models using transient temperature analysis,'' Phys. Med. Biol. **43**, 3325–3340 (1992).
[20] L. Chen, G. ter Haar, D. Robertson, J. Bensted, and C. Hill, ''Histological study of normal and tumor-bearing liver treated with focused ultrasound,'' Ultrasound Med. Biol. **25**, 847–856 (1999).
[21] C. Le Floch, M. Tanter, and M. Fink, ''Self-defocusing in ultrasonic hyperthermia: Experiment and simulation,'' Appl. Phys. Lett. **74**, 3062–3064 (1999).
[22] R. Seip, P. VanBaren, C. Cain, and E. Ebbini, ''Noninvasive real-time multipoint temperature control for ultrasound phased array treatments,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **43**, 1063–1073 (1996).
[23] J. Lu, H. Ying, Z. Sun, M. Motamedi, B. Bell, and L. Sheppard, ''*In vitro* measurement of speed of sound during coagulate tissue heating,'' in Proceedings of the IEEE Ultrasonics Symposium, pp. 1299–1302 (1996).
[24] M. F. Hamilton and C. L. Morfey, ''Model equations,'' in *Nonlinear Acoustics*, edited by M. F. Hamilton and D. T. Blackstock (Academic, San Diego, 1998), Chap. 3.
[25] R. T. Beyer, ''The Parameter B/A'' in *Nonlinear Acoustics*, edited by M. F. Hamilton and D. T. Blackstock (Academic, San Diego, 1998), Chap. 2.
[26] I. Hallaj, R. Cleveland, R. Roy, and R. Holt, ''Coupled thermal-acoustic simulation results with temperature-dependent tissue parameters for therapeutic ultrasound,'' J. Acoust. Soc. Am. **104**, 1844(A) (1998).
[27] I. Hallaj and R. Cleveland, ''FDTD simulation of finite-amplitude pressure and temperature fields from biological ultrasound,'' ARLO **1**, 7–12 (1999).
[28] S. Goss, R. Johnston, and F. Dunn, ''Comprehensive compilation of empirical ultrasonic properties of mammalian tissues,'' J. Acoust. Soc. Am. **64**, 423–457 (1978).
[29] A. D. Pierce, *Acoustics, An Introduction to its Physical Principles and Applications* (AIP Press, Woodbury, NY, 1989), Chap. 10.
[30] S. Sapareto and W. Dewey, ''Thermal dose determination in cancer therapy,'' Int. J. Radiat. Oncol., Biol., Phys. **10**, 787–800 (1984).
[31] J. Bamber and C. Hill, ''Ultrasonic attenuation and propagation speed in mammalian tissues as a function of temperature,'' Ultrasound Med. Biol. **5**, 149–157 (1979).
[32] R. Schechter, K. Simmonds, R. Mignogna, and P. Delsanto, ''Use of a transient wave propagation code for 3D simulation of CW radiated transducer fields,'' Ultrasonics **37**, 89–96 (1999).
[33] S. Umemura and C. Cain, ''The sector-vortex phased array: Acoustic field synthesis for hyperthermia,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **36**, 249–257 (1989).
[34] C. Damianou, N. Sanghvi, F. Fry, and R. Maas-Moreno, ''Dependence of ultrasonic attenuation and absorption in dog soft tissues on temperature and thermal dose,'' J. Acoust. Soc. Am. **102**, 628–634 (1997).
[35] E. Carstensen, S. Becroft, W. Law, and D. Barbee, ''Finite amplitude effects on the threshold for lesion production in tissues by unfocused ultrasound,'' J. Acoust. Soc. Am. **70**, 302–309 (1981).

Vocal behavior of male sperm whales: Why do they click?

Nathalie Jaquet,[a] Stephen Dawson, and Lesley Douglas
Department of Marine Science, University of Otago, P.O. Box 56, Dunedin, New Zealand

(Received 12 August 2000; revised 22 January 2001; accepted 28 February 2001)

Off Kaikoura, New Zealand, we recorded individually identified male sperm whales (*Physeter macrocephalus*) for entire dive cycles in order to investigate vocal behavior of individual whales and to examine possible functions of sperm whale clicks. In our study, sperm whales were almost always silent at the surface. They consistently started clicking within 25 s after fluking-up and diving. During the first 10 s of clicking, interclick intervals were significantly correlated with water depths at the location of fluke-up. The first ''creak'' was produced on average 7.5 min into a dive. Interclick intervals usually decreased substantially before clicks turned into ''creaks.'' The highest click rate recorded in this study was 90.9 click/s, and clicks-within-creaks were much shorter than ''usual clicks'' (mean of 3.6 ms versus 17 to 30 ms). The number of creaks per minute of dive and the length of a dive were significantly correlated. On average, sperm whales were silent for the last 3.6 min before surfacing. Short sequences of ''surface clicks'' (3 to 8 metallic clicks with mean interclick interval of 5.5 s) were often produced at the end of a dive (in 57% of the dives), but their function remains puzzling. The results of this study suggest that usual clicks and creaks are both used for echolocation purposes, the former to gather information about acoustically reflective features and the latter to detect prey.
[DOI: 10.1121/1.1360718]

PACS numbers: 43.80.Ka, 43.80.Jz, 43.66.Gf [WA]

I. INTRODUCTION

Sperm whales (*Physeter macrocephalus*) are highly vocal cetaceans and primarily produce sharp, impulsive, broadband sounds called clicks (Backus and Schevill, 1966). Clicks can be produced with a variety of repetition rates, and have been assigned to four main categories. ''Usual clicks,'' the most commonly heard sound, have an interclick interval (ICI) of about 0.5 to 1 s; ''slow clicks'' have an ICI of about 5 to 7 s (Weilgart and Whitehead, 1988); ''creaks'' are series of very rapid clicks with up to 220 clicks per second (Gordon, 1987); and ''codas'' are short, patterned series of clicks with irregular repetition rates (Watkins and Schevill, 1977).

Codas are the only sperm whale vocalizations for which function appears to be clear, and for which there is a consensus among researchers: they are mainly produced during social interactions and are thought to play a role in communication (Whitehead and Weilgart, 1991; Pavan *et al.*, 2000). The function of slow clicks, which were described by Weilgart and Whitehead (1988) and Mullins *et al.* (1988), is still puzzling. They were only produced by males, either at particular stages of their dive cycles (Mullins *et al.*, 1988) or in particular areas and by particular individuals (Weilgart and Whitehead, 1988). That these sounds are used in several different contexts suggests that they may have several different functions.

It has often been assumed that usual clicks are produced for echolocation purposes. They closely resemble the clicks of other known echolocators (e.g., *Tursiop truncatus*, *Delphinapterus leucas*), and are mainly heard in foraging context (Whitehead and Weilgart, 1991). On the other hand, Watkins (1980) suggested that the characteristics of usual clicks (i.e., lack of directionality, regular repetition rate, long click duration and high click level) make them unsuitable for echolocation. He also argued that sperm whales are mostly silent when not in acoustic contact with other sperm whales, and thus he postulated that usual clicks are produced for communication purposes only. However, recent studies on click source level and directionality have shown that sperm whale clicks are highly directional and thus potentially suitable for echolocation (Møhl *et al.*, 2000). Because sperm whale creaks closely resemble the feeding buzzes of bottlenose dolphins and bats (Au, 1993), it has been postulated that creaks represent short-range echolocation when sperm whales are closing in on their prey (Gordon, 1987; Goold and Jones, 1995). However, sperm whale diet consists almost exclusively of squid (Kawakami, 1980), which, lacking a swim bladder, have a very low target strength (Starr and Thorne, 1998). Therefore, other researchers (Fristrup and Harbison, 1993; Frankel, personal communication) argue that sperm whales cannot detect squid by using echolocation, but rely on vision since light plays an important role in the deep-sea ecosystem.

Our lack of knowledge of the function of sperm whale clicks is partly due to the difficulty of attributing a particular vocalization to a specific individual or to a particular stage of its dive cycle. Most studies of sperm whale vocal behavior have been conducted with groups of female and immature whales (e.g., Gordon, 1987; Whitehead and Weilgart, 1991). The cohesive social organization of these groups, as well as the small distances between each individual (Whitehead, 1989), prevented the recording of vocalizations from individually identified whales. Kaikoura, New Zealand, represents a unique area to study sperm whale vocal behavior as male sperm whales are loosely aggregated (Jaquet *et al.*,

[a]Electronic mail: nathalie.jaquet@stonebow.otago.ac.nz

TABLE I. Type, timing, and sample size of vocalization at different stages of the dive cycle.

Question	Requirements	Sample size	Results
Timing of first vocalization	Exact timing of fluke-up and first vocalization	373	$\overline{x}=24.9$ s after fluke-up CV=80%
Type of first vocalization	Recording of the first vocalization	373	Surface clicks in 24.9% of the dives Usual clicks in 75.1% of the dives
Timing of first creak	Exact timing of fluke-up, excellent recording until at least the first creak	36	$\overline{x}=7.53$ min after fluke-up CV=33%
Timing of last creak	Recording of the last creak, and exact timing of surfacing	20	$\overline{x}=6.67$ min before surfacing CV=42%
Duration of creaks	Recording of very clear creaks	376	$\overline{x}=15.65$ s CV=93%
Timing of last vocalization	Recording of the last vocalization and exact timing of surfacing	47	$\overline{x}=3.64$ min before surfacing CV=34%
Type of last vocalization	Recording of the last vocalization and exact timing of surfacing	47	Surface clicks in 57.4% of the dives Creaks in 21.3% of the dives Usual clicks in 21.3% of the dices
Number of surface clicks	Recordings good enough to count the number of surface clicks	20	$\overline{x}=5.87$ surface clicks CV=30%
Interclick intervals in trains of surface clicks	Surface clicks clear enough to measure the interclick interval between them	192	$\overline{x}=5.47$ s CV=29%

2000), making possible the repeated recording of single identified individuals.

In this paper, we describe the vocal behavior of individually identified male sperm whales during entire dive cycles. This includes a description of a vocalization which resembles the slow click described by Weilgart and Whitehead (1988) but which is produced in totally different contexts. We also present data on the characteristics of creaks and of clicks-within-creaks, and discuss possible functions of sperm whales vocalizations in light of these new results.

II. METHODS

Data were collected from a 6.6 m rigid-hulled inflatable boat powered by a 90 hp outboard engine during four field seasons spanning four to ten weeks each (from 1996 to 1999). Almost all research effort was conducted within a study area of 10 by 15 nautical miles (n.miles) situated over the Kaikoura canyon and the continental slope (Jaquet *et al.*, 2000). Fluke photographs were taken at the start of the dive to allow individual identification (Arnbom, 1987). Recordings were made using an omnidirectional hydrophone (Sonatech 8178; minimum frequency response 100 Hz to 30 kHz$\pm$5 dB) lowered to 20 m. The hydrophone was first connected to a fixed gain amplifier (flat response from 0 to 45 kHz) and then to one channel of a Sony TCD-D10PROII Digital Audio Tape recorder (frequency response 20 Hz to 22 kHz$\pm$1 dB). The remaining channel was used to record a spoken commentary. Recordings were filtered using a low-pass filter at 20 kHz and digitized at 44.1 kHz (16 bit). Analyses of short click sequences were conducted using Canary™ signal processing software (v. 1.2.4; Cornell University). Interclick intervals were measured directly on the waveform from the onset of one click to the onset of the next. Click rates were analyzed using either Moby Click (Jäke, 1996) or Rainbow Click (Gillespie and Leaper, 1997).

III. RESULTS

During a dive cycle, recording quality was variable and occasionally some part of the recording had to be disregarded. Sample sizes for different type of vocalizations were thus unequal (Table I). The vast majority of sperm whale vocalizations were heard when the whales were underwater. Clicking while at the surface was very rare, and sperm whales were never heard clicking during shallow dives (i.e., when a whale sinks slowly under the surface without raising its fluke).

To investigate whether sperm whales were echolocating on the sea floor during the first few seconds of their dives and/or during the first 5 min of their descent, a coefficient of correlation was calculated between the interclick interval and the water depth at the location of the fluke-up. There was a significant correlation between interclick intervals (averaged over the first 10 s) and water depths at the dive location ($r=0.688$, $p=0.002$, $df=17$). However, there was no consistent decrease in interclick intervals during the first 5 min of their descent (Douglas, 2000).

Creaks were only emitted during the middle part of a dive and were never heard either at the beginning of the dive or when whales were at the surface (Table I). The structure of click-within-creaks was difficult to determine, as the orientation of the whales in relation to the hydrophone was not known. On a few occasions, whales appeared to be ''creaking'' on the boat (the creak could also be heard directly through the hull without the need of a hydrophone), and thus were oriented roughly toward our hydrophone. In these cases, click-within-creaks had the same basic structure as usual clicks (i.e., a multi-pulse structure). However, click-within-creaks were always considerably shorter ($\overline{x}=3.59$ ms, CV=35%, range=1.55 to 5.907 ms, $n=328$) than the average duration of usual clicks. Interclick intervals (ICI) varied considerably both within and between creaks. The shortest ICI measured during this study was 11.9 ms ($n=2103$). In general, ICI decreased consistently from the beginning of the creak and then stabilized between 20 and 30 ms (see Fig. 1 for a typical example). Creaks were usually followed by a short period of silence ($\sim$2 to 10 s). The number of creaks per min of dive and the length of a dive were significantly correlated ($r=0.788$, $df=12$, $p<0.001$).

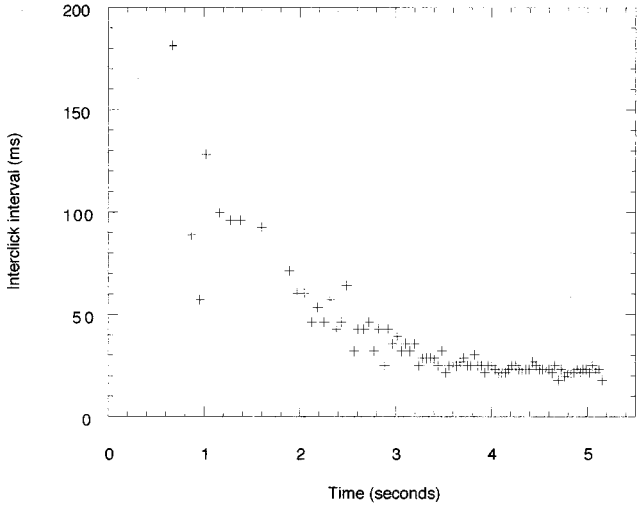


FIG. 1. Typical variation in interclick intervals within a creak.

Individuals which were performing dives longer than average were also producing proportionally more creaks.

Sperm whales were silent during the last 3 to 4 min of their dives. The majority of the dives ended with loud metallic clicks that we called ''surface clicks'' (Table I). These clicks sounded very different to usual clicks: they were characterized by an interclick interval of 5 to 6 s (Table I) and were heard at the very beginning or very end of the dive. Every individual for which we had more than two recordings ended at least one of its dive with surface clicks. The production of surface clicks was not correlated with particular depths, nor particular areas or seasons.

Surface clicks were generally produced in short trains of 3 to 8 clicks, each train having a mean duration of 24.1 s (CV=39%, $n=16$). Multiple trains of surface clicks (3 to 6 short trains) were heard in four dives out of 20. Within trains of surface clicks, interclick interval (ICI) stayed more or less constant, decreased slightly or fluctuated as the regularly-spaced teeth of a saw (Fig. 2). There were large differences in the mean ICI between each train of surface clicks, however, these differences were not related to individuals as, for example, MTR 140 had means varying from 5.5 to 7.9 s while HL 180, LSR 60, and NN 160 had mean ICI of 3.7, 3.5, and 3.5 s, respectively.

Off the 47 end of dive recordings, 27 dives ended with a surface click, 10 with a creak, and 10 with a usual click (Table I). There was a highly significant difference in the timing of surface clicks, creaks, and usual clicks (1 factor ANOVA, $p=0.0001$, $df=44$). The time interval was the longest when the last vocalization was a creak ($\overline{x}=4.86$ min, $SE=0.3$, $n=10$), slightly shorter when it was a usual click ($\overline{x}=4.18$ min, $SE=0.45$, $n=10$), and significantly shorter when it was a surface click ($\overline{x}=3.05$ min, $SE=0.18$, $n=27$). There was no correlation between the time from the last vocalization until surfacing, and the depth of the sea floor ($r=-0.012$, $df=46$, $p>0.5$). As well, there was no correlation between type of last vocalization and bottom depth at surfacing.

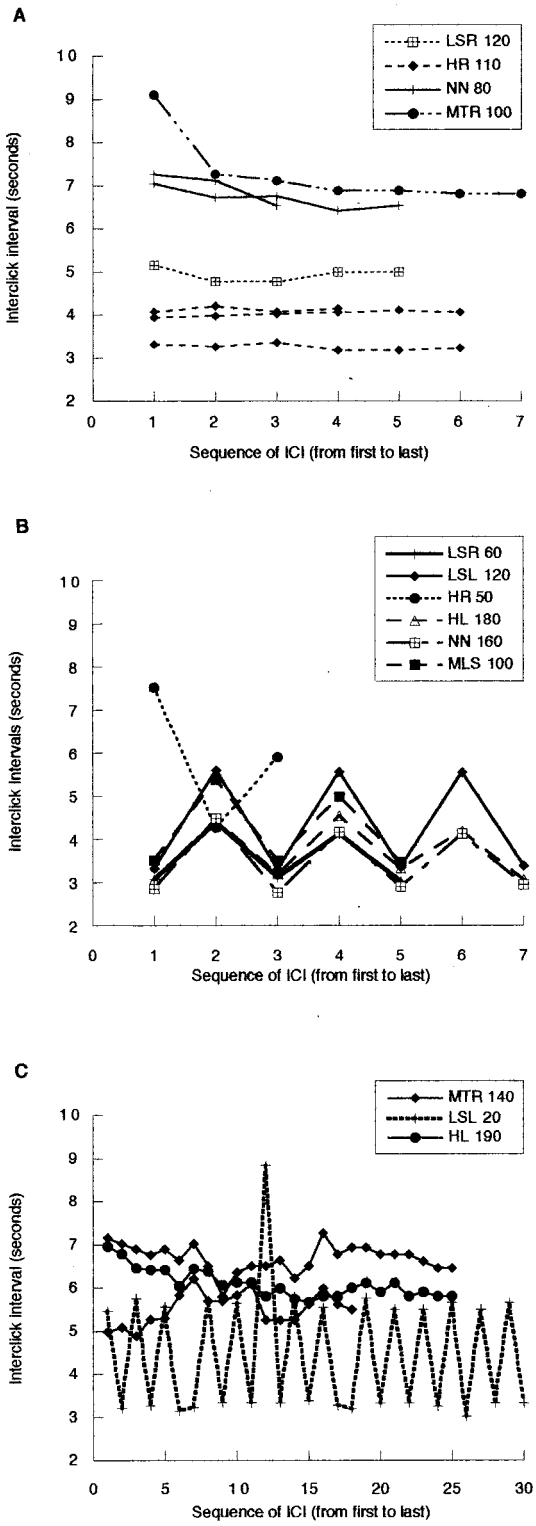


FIG. 2. Variation in interclick intervals in surface click sequences: (a) and (b) short trains; (c) long trains.

IV. DISCUSSION

If the primary function of sperm whale usual clicking is for echolocation while foraging, we would expect most of the clicking to occur during foraging dives, and very little clicking to occur at the surface. As shallow diving seem to be sperm whales usual reaction to disturbance, and does not

appear to be related to foraging (IFAW, 1996), we would also expect very little clicking while shallow diving. Our study supports this hypothesis as 97.1% of sperm whale vocalizations were heard during foraging dives. During these dives, sperm whales clicked almost continuously, being silent for only 15.5% of the time between fluke-up and surfacing. On the other hand, if the primary function of usual clicking is for communication (as postulated by Watkins, 1980), one would expect differences in vocalization rate between males and females (females being more social than males), but one would not expect whales to click exclusively during foraging dives. Our results are not consistent with Watkins' (1980) hypothesis.

In our study, sperm whales started clicking consistently about 25 s after fluke-up. Assuming a descent rate of 60 to 120 m per min (Gordon, 1987; Papastavrou *et al.*, 1989), this would indicate that, off Kaikoura, male sperm whales start clicking at a depth of 25 to 50 m. This is considerably shallower than observations from previous studies of groups of female and immature sperm whales (150 to 300 m, Gordon, 1987; Papastavrou *et al.*, 1989). It is possible that differences in bathymetric characteristics between areas induce differences in the timing of the first click, or that males start clicking earlier than females.

The significant correlation between the interclick interval averaged over the first 10 s of vocalization of a whale dive and the water depth at the fluke-up location, suggests that the first few clicks serve to detect the bottom depth. However, the absence of a consistent decrease in interclick intervals during the first 5 min of the dive suggests that sperm whales are not echolocating consistently on the seafloor during their descent.

If we assume that sperm whale clicks are used (at least on some occasions) for echolocation purposes, then the time intervals between their clicks most likely divulges some information on maximum target range. Dolphins normally space their clicks apart so that clicks and echoes do not overlap, and there is always a short time lag between the reception of the previous click echo and the emission of the next click (Au, 1993). Therefore, if sperm whales behave similarly to most known echolocators, a consistent ICI of about 1.04 s during the first click train suggests that during the beginning of the dive, sperm whales scan a maximum of 750 m ahead of themselves. Similarly, ICIs in killer whales (*Orcinus orca*) suggest that they typically scan 400 to 800 m ahead (Barrett-Lennard *et al.*, 1996).

Off Kaikoura, sperm whales were almost always silent during the last 3 to 4 min of their dives. Assuming an ascent rate similar to their descent rate (60 to 120 m per min), this result suggests that the whales stopped clicking at a depth of 480 to 180 m. The absence of any correlation between the timing of the last vocalization and the water depth at the location of surfacing suggests that sperm whales off Kaikoura are either not foraging on the sea-floor or that they forage on the way up, until depths of 200 to 500 m. These results are consistent with the suggested dive depths of sperm whales (300 to 800 m, Gordon, 1987; Papastavrou *et al.*, 1989), and with Jaquet *et al.* (2000), who suggested that sperm whales off Kaikoura were not routinely diving to the sea-floor.

The multi-pulse structure of clicks-within-creaks found in this study differs from what was found by Gordon (1987) who stated that clicks-within-creaks had a single pulse structure. Similarly, Backus and Schevill (1966) reported a single pulse structure for the one creak they analyzed. Some of the clicks we analyzed seemed also to have a single pulse structure and it is therefore possible that the multi-pulse structure may be very hard to detect if a whale is not facing toward the hydrophone. Alternatively, a whale may be able to vary the number of pulses which are produced within a click.

The average duration of clicks-within-creaks ($\overline{x} = 3.59$ ms, $n = 3.28$) was much shorter than the average duration of usual clicks (18–24 ms, Backus and Schevill, 1966; ~17 ms this study), and consequently the inter-pulse intervals (IPIs) were also much shorter. It is puzzling that sperm whales can produce clicks with shorter IPI than in usual clicks as IPI duration is supposed to be a function of the spermaceti sac length (Norris and Harvey, 1972; Goold, 1996). In our study, the maximum click rate recorded was 90.9 clicks per s, a close match with Watkins *et al.*'s (1985) maximum click rate of 90 clicks/s. These rates suggest plausible maximum target distances of 5 to 12 m (assuming instantaneous processing and click length of 3.6 ms).

The interclick interval between usual clicks tended to decrease before creaks. Within creaks, ICIs also tended to decrease from the beginning to the end of the creak. These results are consistent with Gordon's (1987) findings and with results obtained on bat echolocation behavior (Au, 1993). Creaks often started with an ICI of 50 to 60 ms and ended up with an ICI of 15 to 20 ms. These variations in ICI suggest that sperm whales moved about 33 m between the beginning and the end of a creak. The resulting speed estimate (7.4 km/h: 33 m in 16 s) is slightly higher than the speed of sperm whales averaged over several hours to several days (2.5 to 6.7 km/h; Gordon, 1987; Jaquet and Whitehead, 1999), but is plausible. Furthermore, during short periods of time, individual sperm whales have been recorded at significantly higher speeds (11 km/h, Arnbom *et al.*, 1987; 10 km/h, personal observation). However, the exact whale–target distance cannot be known as there is no information on the duration of the echo-processing time for sperm whales, and as this delay is likely to vary depending on target distance and task difficulty (Au, 1993).

Our results show that the characteristics of sperm whale creaks are very similar to the characteristics of dolphin, porpoise and bat echolocation buzzes, and therefore suggest that sperm whales are emitting creaks when closing in on prey. If creaks represent feeding events, one would expect that successful dives (i.e., dives where a large amount of food items are found) would last longer than unsuccessful ones. The significant correlation between the number of creaks per minute of dive and the dive length found in the present study supports this hypothesis.

The surface clicks heard in the present study had similar characteristics to the slow clicks described by Weilgart and Whitehead (1988) and Mullins *et al.* (1988), and the clangs described by Gordon (1987). All these vocalizations had a

long interclick interval (5 to 7 s on average) and sounded very metallic. However, the slow clicks heard on the breeding grounds (Gordon, 1987; Weilgart and Whitehead, 1988) were produced in a different context than the ones heard at higher latitudes (Mullins *et al.*, 1988; this study). On the breeding grounds, slow clicks were only produced by large mature males, and were produced in long sequences of up to 23 min (Gordon, 1987). Off the Scotian Shelf, however, slow clicks were emitted in short sequences at the very end of dives (Mullins *et al.*, 1988). Similarly, off Kaikoura, surface clicks were produced mainly at the end of a dive, in short sequences of about 24 s in duration, and were produced by males of all sizes. It appears that Mullins *et al.*'s (1988) slow clicks are equivalent to our surface clicks, and that both are used in different contexts to slow clicks in the tropics.

Surface clicks were produced by every individual for which we had multiple recordings, thus surface click production did not depend on individual differences. Although over half of the dives ended with a short sequence of surface clicks, there was no consistent pattern as to when they were emitted (depth, presence of boats in the area, presence of other whales nearby, season, location, etc.). If the main function of these clicks is echolocation, a mean ICI of 5.5 s suggests a maximum detection range of about 4 km. As the whales were on their way up at a probable depth of 180 to 360 m (assuming an ascent rate of 60 to 120 m/min) when producing these clicks, it seems unlikely that these clicks were needed for long-range echolocation.

The production of surface clicks is also difficult to explain as a form of communication. Surface clicks were produced regardless of the presence of other whales and were never produced by females (personal observation). Weilgart and Whitehead (1988) suggested that the ICI in slow clicks might allow identification of individual males but did not have enough data to support this hypothesis. It is clear that in the case of surface clicks, the ICI does not allow individual identification as a single individual can have widely varied mean ICIs.

The results of our study suggest that usual clicks and creaks are both used for echolocation purposes, the former for long-range echolocation (to gather information on acoustically reflective features), and the latter for short-range echolocation (to detect prey). However, although the primary function of these sounds appears to be echolocation, creaks may be a necessary consequence of sperm whale feeding, and thus may be used by nearby whales as a reliable signal of feeding behavior (Tyack, 1998). Furthermore, since during foraging female sperm whales coordinate their behavior without the emission of codas, it is possible that usual clicks also have a communication function. Therefore the echolocation function of usual clicks and creaks suggested in this study does not preclude some communication function as well.

ACKNOWLEDGMENTS

This study was funded by a fundraising partnership between World Wide Fund for Nature (NZ) and Telecom NZ Ltd, and by the University of Otago. Reckitt and Colman (NZ) Ltd., Ocean Electronics and Hutchwilco each assisted by sponsoring items of equipment. We are very grateful to all those who helped us collecting the data at sea. We would like to thank IFAW and Douglas Gillespie for providing ''Rainbow Click,'' and Luke Rendell for his help in using it. Nathalie Jaquet was supported in NZ by postdoctoral fellowships from Novartis, ''Musée de Bâle'' and the Roche Research Foundation (Switzerland). The manuscript was much improved by comments from Deborah McCutchen, John Goold, Keith Jensen, Christoph Richter, Bernd Würsig and two anonymous reviewers.

Arnbom, T. (**1987**). ''Individual identification of sperm whales,'' Rep. Int. Whal. Commn. **37**, 201–204.

Arnborn, T., Papastavrou, V., Weilgart, L. S., and Whitehead, H. (**1987**). ''Sperm whales react to an attack by killer whales,'' J. Mammal. **68**, 450–453.

Au, W. W. L. (**1993**). *The Sonar of Dolphins* (Springer-Verlag, 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.

Barrett-Lennard, L. G., Ford, J. K. B., and Heise, K. A. (**1996**). ''The mixed blessing of echolocation: Differences in sonar use by fish-eating and mammal-eating killer whales,'' Anim. Behav. **51**, 553–565.

Douglas, L. (**2000**). ''Click counting: An acoustic censusing method for estimating sperm whale abundance,'' M.Sc. thesis, University of Otago, Dunedin, New Zealand.

Fristrup, K., and Harbison, G. R. (**1993**). ''Vision and sperm whale foraging,'' Tenth Biennial Conference on the Biology of Marine Mammals, Galveston, Texas.

Gillespie, D., and Leaper, R. (**1997**). ''Detection of sperm whale (*Physeter macrocephalus*) clicks and discrimination of individual vocalisations,'' Eur. Res. Cet. **10**, 87–91.

Goold, J. C. (**1996**). ''Signal processing techniques for acoustic measurement of sperm whale body lengths,'' J. Acoust. Soc. Am. **100**, 3431–3441.

Goold, J. C., and Jones, S. E. (**1995**). ''Time and frequency domain characteristics of sperm whale clicks,'' J. Acoust. Soc. Am. **98**, 1279–1291.

Gordon, J. C. (**1987**). ''The behaviour and ecology of sperm whales off Sri Lanka,'' Ph.D. thesis, University of Cambridge, U.K.

IFAW (**1996**). Report of the Workshop on the Special Aspects of Watching Sperm Whales, Roseau, Commonwealth of Dominica.

Jäke, O. (**1996**). ''Acoustic censusing of sperm whales at Kaikoura, New Zealand: An inexpensive method to count clicks and whales,'' M.Sc. thesis, Otago University, Dunedin, New Zealand.

Jaquet, N., and Whitehead, H. (**1999**). ''Movements, distribution and feeding success of sperm whales in the Pacific Ocean, over scales of days and tens of kilometers,'' Aquatic Mammals **25**, 1–13.

Jaquet, N., Dawson, S., and Slooten, E. (**2000**). ''Seasonal distribution and diving behaviour of male sperm whales off Kaikoura: Foraging implications,'' Can. J. Zool. **78**, 407–419.

Kawakami, T. (**1980**). ''A review of sperm whale food,'' Scientific Reports of the Whales Research Institute **32**, 199–218.

Møhl, B., Wahlberg, M., Madsen, P. T., Miller, L. A., and Surlykke, A. (**2000**). ''Sperm whale clicks: Directionality and source level revisited,'' J. Acoust. Soc. Am. **107**, 638–648.

Mullins, J., Whitehead, H., and Weilgart, L. S. (**1988**). ''Behaviour and vocalizations of two single 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 (*Physeter macrocephalus*),'' in *Animal Orientation and Navigation*, edited by S. R. Galler, K. Schmidt-Koenig, G. J. Jacobs, and R. E. Belleville, Spec. Publ., NASA, Washington D.C., pp. 397–419.

Papastavrou, V., Smith, S. C., and Whitehead, H. (**1989**). ''Diving behaviour of the sperm whale (*Physeter macrocephalus*) off the Galápagos Islands,'' Can. J. Zool. **67**, 839–846.

Pavan, G., Hayward, T., Borsani, J., Priano, M., Manghi, M., Fossati, C., and Gordon, J. (**2000**). ''Time patterns of sperm whale codas recorded in

the Mediterranean Sea 1985–1996,'' J. Acoust. Soc. Am. **107**, 3487–3495.
Starr, R. M., and Thorne, R. E. (**1998**). ''Acoustic assessment of squid stocks,'' in *Squid Recruitment Dynamics. The Genus Illex as a Model, the Commercial Illex Species and Influences on Variability*, edited by P. G. Rodhouse (FAO Fish. Tech. Paper, Rome).
Tyack, P. L. (**1998**). ''Acoustic communication under the sea,'' in *Animal Acoustic Communication*, edited by S. L. Hopp, M. J. Owren, and C. S. Evans (Springer, Berlin), pp. 163–220.
Watkins, W. A. (**1980**). ''Acoustics and the behavior of sperm whales,'' in *Animal Sonar Systems*, edited by R. G. Busnel and J. F. Fish (Plenum, New York), pp. 283–290.
Watkins, W. A., and Schevill, W. E. (**1977**). ''Sperm whale codas,'' J. Acoust. Soc. Am. **62**, 1485–1490.
Watkins, W. A., Moore, K. E., and Tyack, P. (**1985**). ''Sperm whale acoustic behaviors in the southeast Carribean,'' Cetology **49**, 1–15.
Weilgart, L. S., and Whitehead, H. (**1988**). ''Distinctive vocalizations from mature male sperm whales (*Physeter macrocephalus*),'' Can. J. Zool. **66**, 1931–1937.
Whitehead, H. (**1989**). ''Formations of foraging sperm whales, *Physeter macrocephalus*, off the Galápagos Islands,'' Can. J. Zool. **67**, 2131–2139.
Whitehead, H., and Weilgart, L. (**1991**). ''Patterns of visually observable behaviour and vocalizations in groups of female sperm whales,'' Behaviour **118**, 275–296.

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.

Probabilistic solutions of the Helmholtz equation

B. V. Budaev[a] and D. B. Bogy[b]
Department of Mechanical Engineering, University of California, Berkeley, California 94720

(Received 2 August 2000; revised 16 February 2001; accepted 20 February 2001)

Solutions of the Schrödinger and Helmholtz equations are represented as mathematical expectations of functionals on paths in a complex space. Formulas obtained for the Helmholtz equation appear as direct improvements of the ray-method approximations to the exact solution. Presented numerical results confirm the efficiency of the suggested description of wave propagation phenomena. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1365113]

PACS numbers: 43.20.Bi, 02.70.Rr, 03.65.Ge [ANN]

One of the most efficient tools for analysis of wave propagation is the ''ray method.'' [1] This method is based on representations of wave fields in the Liouville form $\Sigma \phi_n(x) e^{iS_n(x)}$, where the eikonals $S_n(x)$ are defined by the canonical procedure and the amplitudes $\phi_n(x)$ are defined by second-order partial differential equations, which, in many important cases, may be approximated by first-order transport equations admitting explicit solutions. In general, however, equations defining amplitudes $\phi_n(x)$ have neither exact nor approximate analytic solutions.

Here, we develop an approach to problems of wave propagation combining the physical clarity of the ray method with the versatility of direct numerical methods. We start, following closely the scheme of the ray method, but instead of looking for approximate expressions for the amplitudes of the Liouville representations we obtain their exact values as the mathematical expectations of some functionals on the space of Brownian trajectories. The obtained solutions provide direct improvements of the ray-method approximations, and they are shown to admit efficient numerical evaluations.

The first probabilistic solutions of partial differential equations were obtained in the 1920s.[2] Rapid progress in the development of probabilistic methods in partial differential equations was made in the 1950s, after the publication of Feynman's[3] and Kac's[4] papers applying methods of functional integration to the Schrödinger and the diffusion equations. In Ref. 5, solutions of the Helmholtz equation were represented by Feynman's path integrals and classical ray-method formulas were rederived as asymptotic approximations of Feynman's integrals. Quite recently, new attempts have been made to apply Feynman's integral methods to acoustics[6] and electromagnetic.[7] Probabilistic methods based on Kac's solution of the diffusion equation have also been recently applied to steady flow computations[8] and these methods are increasingly used for analysis of wave propagation[9,10] in random media. In Ref. 10 the competitiveness of probabilistic methods in wave propagation is discussed, and in Refs. 9 and 10 such methods are used to study the transport of energy by waves propagating in random media. Such transport is described by first-order differential equations and a probabilistic method is developed in those papers for their analysis. Here, instead, we deal with problems in nonrandom media but employ probabilistic methods for the analysis of the second-order auxiliary equations whose first-order components are also known in ray theory as transport equations.

Consider the Cauchy problem

$$\frac{\partial \psi}{dt} = \frac{i}{2}\nabla^2 \psi + B\psi, \quad \psi|_{t=0} = F, \tag{1}$$

for the Schrödinger equation in $\mathbb{R}^N$ with an analytic potential $B(x)$ and an initial data $F(x)$.

Let $\mathbf{w}_t$ be the standard Brownian motion[11] in $\mathbb{R}^N$ and let $\mathfrak{B}_t$ be a family of operators acting on analytic functions $F(x), x \in \mathbb{C}^N$ by the formula

$$\mathfrak{B}_t F(x) = \mathbf{E}[F(\boldsymbol{\xi}_t^x) e^{\int_0^t B(\xi_s^x) ds}], \tag{2}$$

$$\boldsymbol{\xi}_t^x = x + \sqrt{i}\,\mathbf{w}_t, \tag{3}$$

where the mathematical expectation $\mathbf{E}$ corresponds to the Wiener measure on Brownian motions $\mathbf{w}_t$ in the real space $\mathbb{R}^N$.

Standard probabilistic considerations[11] show that $\mathfrak{B}_{t+s} = \mathfrak{B}_t \mathfrak{B}_s$, i.e., that the family $\mathfrak{B}_t$, $t \geqslant 0$, is a semigroup, and that the infinitesimal operator $\mathfrak{b} = d/dt|_{t=0} \mathfrak{B}_t$ of this semi-

[a]Electronic mail: budaev@cml.me.berkeley.edu
[b]Electronic mail: dbogy@cml.me.berkeley.edu

group coincides with the operator $F \to (i/2)\nabla^2 F + BF$ from the right-hand side of (1). Then, the semigroup property $\mathfrak{B}_{t+s} = \mathfrak{B}_t \mathfrak{B}_s$ implies the identity $(d/dt)\mathfrak{B}_t = \mathfrak{b}\mathfrak{B}_t$, which results in the conclusion that if the mathematical expectation in (2) exists, then

$$\psi(x) = \mathfrak{B}_t F(x) \tag{4}$$

represents the solution of the problem (1). Solution (4) is similar to solutions of the Schrödinger equation employed in Ref. 14 where the Feynman's path integral solution was modified by deformation of the paths in the complex space.

This approach can be extended to the Schrödinger equation

$$\frac{\partial \psi}{\partial t} = \frac{i}{2}\nabla^2 \psi + \mathbf{A}\cdot\boldsymbol{\nabla}\psi + B\psi, \quad \psi|_{t=0} = F, \tag{5}$$

with an analytic vector potential $\mathbf{A}$. To solve (5) introduce random motions $\boldsymbol{\xi}_t^x$ in $\mathbb{C}^N$ governed by the stochastic differential equation

$$d\boldsymbol{\xi}_t^x = \mathbf{A}(\boldsymbol{\xi}_t^x)dt + \sqrt{i}\,d\mathbf{w}_t, \quad \boldsymbol{\xi}_0^x = x, \tag{6}$$

where $\mathbf{w}_t$ is still the standard Brownian motion in $\mathbb{R}^N$. Operators $\mathfrak{B}_t$ defined by (2) with random motions $\boldsymbol{\xi}_t^x$ satisfying (6) form a semigroup whose infinitesimal operator coincides with the operator on the right-hand side of (5). Correspondingly, formula (2) with $\boldsymbol{\xi}_t^x$ from (6) solves the Cauchy problem (5).

The integration of Eq. (5) from zero to infinity with respect to the time variable shows that the mathematical expectation

$$\phi(x) = \mathbf{E}\left[\int_0^\infty F(\boldsymbol{\xi}_t^x) e^{\int_0^t B(\boldsymbol{\xi}_s^x)ds} dt\right], \tag{7}$$

averaging random walks from (6), satisfies the equation

$$\frac{i}{2}\nabla^2\phi + \mathbf{A}\cdot\boldsymbol{\nabla}\phi + B\phi + F = 0. \tag{8}$$

Formula (7) can be used to derive another class of solutions of Eq. (8), more convenient for analysis of boundary value problems. Let G be a domain in $\mathbb{R}^N$ with the boundary ∂G, and let Q be a cylinder in $\mathbb{C}^N$ constructed over G such that $\mathrm{Re}(Q) = G$ and $\mathrm{Re}(\partial Q) = \partial G$. Next, assume that $B(x)$ is extended outside Q as $B(x) = -\infty$ for $x \notin Q$, and that $F(x) = f(x)\delta_{\partial Q}(x)$, where $\delta_{\partial Q}(\cdot)$ is the Dirac's δ function with the carrier on ∂Q. Then, the evaluation of (7) leads to the conclusion that an analytic solution of the Dirichlet problem

$$\frac{i}{2}\nabla^2\phi + \mathbf{A}\cdot\boldsymbol{\nabla}\phi + B\phi = 0, \quad \phi|_{\partial G} = f, \tag{9}$$

with analytic boundary values $f(x)$ may be represented by the mathematical expectation

$$\phi(x) = \mathbf{E}[f(\boldsymbol{\xi}_\tau^x) e^{\int_0^\tau B(\boldsymbol{\xi}_s^x)ds}], \tag{10}$$

where τ is the first instant when the real component $\mathrm{Re}[\boldsymbol{\xi}_t^x]$ of the complex path $\boldsymbol{\xi}_t^x$ governed by a stochastic equation (6) hits the boundary ∂Q.

One particular case of (8) is of special interest for the theory of wave propagation, where the problem

$$\Delta^2\psi + k^2\psi = 0, \quad \psi|_{\partial G} = f, \tag{11}$$

for the Helmholtz equation is routinely split[1,12] by the substitution $\psi = \phi e^{-iS}$ to the extended transport equation

$$\frac{i}{2}\nabla^2\phi + \boldsymbol{\nabla}S\cdot\boldsymbol{\nabla}\phi + \frac{1}{2}(\nabla^2 S)\phi = 0, \quad \phi|_{\partial G} = f e^{iS}, \tag{12}$$

and the independent eikonal equation

$$(\boldsymbol{\nabla}S)^2 = k^2, \tag{13}$$

whose solution may always be constructively determined by a well-known method,[12] treating (13) as a particular case of the canonic Hamilton–Jacobi equation of classical mechanics.[13]

After the eikonal $S(x)$ is determined, Eq. (12) has the structure of (9) with $\mathbf{A} = \boldsymbol{\nabla}S$ and $B = \frac{1}{2}\mathrm{div}(A)$. Correspondingly, the solution of (12) can be immediately derived from (7) and it has the form

$$\phi(x) = \mathbf{E}[f(\boldsymbol{\xi}_\tau^x) e^{iS(\boldsymbol{\xi}_\tau^x) + 1/2\int_0^\tau \nabla^2 S(\boldsymbol{\xi}_s^x)ds}], \tag{14}$$

where

$$d\boldsymbol{\xi}_t^x = \boldsymbol{\nabla}S(\boldsymbol{\xi}_t^x)dt + \sqrt{i}\,d\mathbf{w}_t, \quad \boldsymbol{\xi}_0^x = x. \tag{15}$$

Due to a special relationship between the exponent in (14) and the drift term $\boldsymbol{\nabla}S\,dt$ of the stochastic equation (15), solution (14) can be further transformed to the alternative form

$$\phi(x) = \mathbf{E}\left[f(\boldsymbol{\xi}_\tau^x)\sqrt{\frac{J(\boldsymbol{\xi}_\tau^x)}{J(\boldsymbol{\xi}_0^x)}} \times \exp\left(iS(\boldsymbol{\xi}_\tau^x) - \int_0^\tau \frac{i}{4}\nabla^2(\ln J)ds + \frac{\sqrt{i}}{2}\boldsymbol{\nabla}(\ln J)\cdot d\mathbf{w}_s \right)\right], \tag{16}$$

where the exponent contains Itô's[11] stochastic integration and $J(x)$ is the geometrical divergence of the vector field $\mathbf{A} = \boldsymbol{\nabla}S$ introduced[1,12] as the quantity satisfying the Liouville identity

$$\mathrm{div}[\mathbf{A}(x)] = \frac{d}{d\tau}\bigg|_{\tau=0} \ln[J(x + \mathbf{A}(x)\tau)], \quad \mathbf{A} = \boldsymbol{\nabla}S. \tag{17}$$

Solution (16) is convenient for the analysis of wave propagation phenomena because (17) is obviously related to the well-known solution

$$\phi_0(x) = f(\boldsymbol{\xi}_\tau^x)\sqrt{\frac{J(\boldsymbol{\xi}_\tau^x)}{J(\boldsymbol{\xi}_0^x)}} e^{iS(\boldsymbol{\xi}_\tau^x)}, \quad \text{where } (d/dt)\boldsymbol{\xi}_t^x = \boldsymbol{\nabla}S(\boldsymbol{\xi}_t^x), \quad \boldsymbol{\xi}_\tau^x \in \partial G, \tag{18}$$

of the transport equation

$$2\boldsymbol{\nabla}S\cdot\boldsymbol{\nabla}\phi_0 + (\nabla^2 S)\phi = 0, \quad \phi|_{\partial G} = f e^{iS}, \tag{19}$$

considered in the ray theory as an approximation to (12).

Probabilistic solutions (14), (16), may also be used for direct numerical simulation. For example, the mathematical expectation in (14) may be approximated by the average

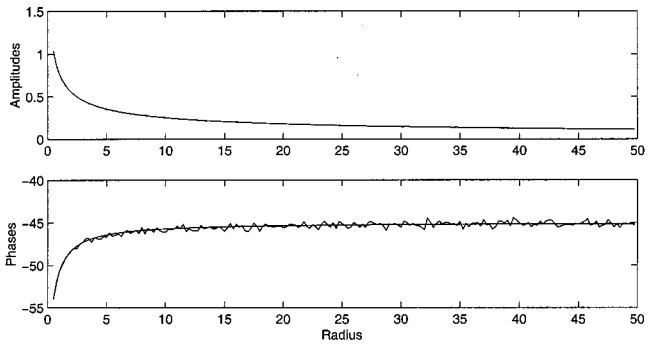


FIG. 1. Simulation of $e^{-ir}H_0^{(1)}(r)$ by random walks.

$$\phi(x)\approx\frac{1}{M}\sum_{m=1}^{M} f(\boldsymbol{\xi}_{mn_*}) \times\exp\left(iS(\boldsymbol{\xi}_{mn_*})-\frac{1}{2}\sum_{n=0}^{n_*-1}\nabla^2 S(\boldsymbol{\xi}_{mn})(\Delta\mathbf{w}_n)^2\right), \quad (20)$$

computed over the discrete trajectories

$$x=\boldsymbol{\xi}_{m0}\to\boldsymbol{\xi}_{m1}\to\boldsymbol{\xi}_{m2}\to\cdots\to\boldsymbol{\xi}_{mn_*},\quad \mathrm{Re}(\boldsymbol{\xi}_{mn_*})\in\partial G, \quad (21)$$

whose nodes are consequently computed by the formulas

$$\boldsymbol{\xi}_{m(n+1)}=\boldsymbol{\xi}_{mn}+\boldsymbol{\nabla}(S(\boldsymbol{\xi}_{mn}))(\Delta\mathbf{w}_n)^2+\Delta\mathbf{w}_n, \quad (22)$$

where $\Delta\mathbf{w}_n$ are random vectors each of whose Cartesian coordinates take one of two equally possible values $\pm\epsilon$, with sufficiently small ϵ.

To get an indication of the numerical efficiency of the obtained probabilistic formulas (14), (16), consider a particular two-dimensional case of (12) with $k=1$ and with the vector field $\mathbf{A}=\mathbf{x}/\|\mathbf{x}\|$ generated by the eikonal $S(x)=\|x\|=r$.

Equation (12) corresponding to this case reduces to the form $(i/2)\nabla^2\phi-(\partial/\partial r)\phi-(1/2r)\phi=0$, which has an obvious solution $\phi(r,\theta)=e^{-ir}H_0^{(1)}(r)$, defined in an exterior domain $r\geqslant R>0$. Since the vector field $\mathbf{A}(r,\theta)=\mathbf{x}/\|\mathbf{x}\|$ and its geometric divergence $J(r,\theta)=r$, are both analytic in the domain $\mathrm{Re}(r)\geqslant R$, the probabilistic formula (14) can be used for simulation of the function $e^{-ir}H_0^{(1)}(r)$ considered as the solution of the Dirichlet problem

$$\frac{i}{2}\nabla^2\phi-\frac{\partial}{\partial r}\phi-\frac{1}{2r}\phi=0,\qquad \phi(R,\theta)=e^{-iR}H_0^{(1)}(R),\quad r\geqslant R. \quad (23)$$

Figure 1 shows the results of the numerical simulation of the function $e^{-ir}H_0^{(1)}(r)$. This function is computed twice: as a Hankel function and as the solution of the Dirichlet problem (23) given by the probabilistic formula (14) approximated by (20)–(22). We also carried out computation of this function based on the solution (16), but the results were practically identical to the ones based on (14).

The presented results correspond to the radius $R=0.5$. The mathematical expectations from (14) estimated by the averaging included 2500 independent discrete Brownian walks with Cartesian increments $\epsilon=0.07$. The amplitudes of the computed functions are displayed on the first diagram, and the second diagram displays their phases. The computations were very stable and despite the use of a rather rough discretization and the simplest algorithms, the relative error was maintained below the low 1.5% level over the entire trial interval $0.5<r<50$.

The probabilistic approach to analysis of the Helmholtz equation may be extended far beyond the simple example considered here. In future papers we consider application of this method to problems with different boundary conditions including Neumann and impedance boundary conditions widely used in acoustics. We have also applied this method to a scattering problem on a semitransparent object and obtained a unified representation of the transmitted and backscattered waves.

ACKNOWLEDGMENTS

This research was supported by NSF Grant No. 9730779 and by the William S. Floyd, Jr. Distinguished Professorship in Engineering held by D. Bogy.

[1] J. B. Keller, ''A geometric theory of diffraction,'' in *Calculus of Variations and its Applications* (McGraw Hill, New York, 1958), pp. 27–52.
[2] H. P. Philips and M. Wiener, ''Nets and Dirichlet problem,'' J. Math. Phys. **2**, 105–124 (1923).
[3] R. P. Feynman, ''Space-time approach to non-relativistic quantum mechanics,'' Rev. Mod. Phys. **20**, 367–387 (1948).
[4] M. Kac, ''On distribution of certain Wiener functionals,'' Trans. Am. Math. Soc. **65**, No. 1, 1–13 (1949).
[5] J. B. Keller and D. W. McLaughlin, ''The Feynman integral,'' Am. Math. Monthly, May, 451–465 (1975).
[6] R. B. Schlottmann, ''A path integral formulation of acoustic wave propagation,'' Geophys. J. Int. **137**, 353–363 (1999).
[7] R. D. Nevels, J. A. Miller, and R. E. Miller, ''A path integral time-domain method for electromagnetic scattering,'' IEEE Trans. Antennas Propag. **48**, No. 4, 565–573 (2000).
[8] F. Y. Hunt, J. F. Douglas, and J. Bernal, ''Probabilistic computation of Poiseuille flow velocity fields,'' J. Math. Phys. **36**, No. 5, 2386–2401 (1995).
[9] G. Bal, J. B. Keller, G. Papanicolaou, and L. Ryzhik, ''Transport theory for acoustic waves with reflection and transmission at interfaces,'' Wave Motion **30**, 303–327 (1999).
[10] G. Bal, G. Papanicolaou, and L. Ryzhik, ''Probabilistic theory of transport processes with polarization,'' SIAM (Soc. Ind. Appl. Math.) J. Appl. Math. **60**, No. 5, 1639–1666 (2000).
[11] E. B. Dynkin, *Markov Processes* (Springer, Berlin, 1965).
[12] V. P. Maslov and M. V. Fedoriuk, *Semi-classical Approximation in Quantum Mechanics* (Reidel Dordrecht, Boston, 1981).
[13] V. I. Arnold, *Mathematical Methods of Classical Mechanics* (Springer, Berlin, 1978).
[14] J. Chang and W. H. Miller, ''Monte Carlo path integration in real time via complex coordinates,'' J. Chem. Phys. **50**, 1648–1652 (1987).

Improved description of shock wave evolution in media with frequency power law dependent attenuation

Kirk D. Wallace, Mark R. Holland, and James G. Miller
Laboratory for Ultrasonics, Department of Physics, Washington University, Saint Louis, Missouri 63130-4899

(Received 28 September 2000; accepted for publication 27 February 2001)

A recent article by Kashcheeva *et al.* considers the effect of dispersion on waveform distortion and shock propagation [Acoust. Phys. **46**, 170–177 (2000)]. The authors determine the dispersion from knowledge of the host material's attenuation properties and a nearly local form of the Kramers–Krönig relations published by our laboratory in 1978. Recent work by our laboratory and others has shown that this original nearly local approximation overestimates the actual dispersion when the frequency power law describing the attenuation deviates from unity. The present letter illustrates the impact of using the exact (nonlocal) Kramers–Krönig relations on shock propagation in media with attenuation obeying a frequency power law ranging from 0–2. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1367250]

PACS numbers: 43.25.Cb [MFH]

I. INTRODUCTION

The frequency dependence of a medium's phase velocity (dispersion) can noticeably alter the evolution of a finite amplitude plane wave. A recent article by Kashcheeva *et al.*[1] considered the combined effects of dispersion and attenuation on waveform distortion and shock evolution.[2] The attenuation was assumed to be described by a power law function of the angular frequency ω,

$$\alpha(\omega)=\alpha_0\left(\frac{\omega}{\omega_0}\right)^y. \tag{1}$$

The symbol α_0, expressed in units of inverse length, is the attenuation at the fundamental frequency ω_0, and y is referred to as the frequency power law. The authors utilized a nearly local approximate form of the Kramers–Krönig relations, previously published by our laboratory,[3,4] to determine the dispersion from knowledge of the host material's attenuation properties. Recent work by our laboratory[5] and others[6,7] has shown that this original nearly local approximation overestimates the actual dispersion when the frequency power law describing the attenuation deviates from unity. The present letter illustrates the effect of using the exact (nonlocal) Kramers–Krönig relations, instead of the original nearly local approximation, on the description of a finite amplitude waveform propagating in a variety of attenuation media with frequency power laws spanning the range from 0–2.

As mentioned above, the frequency dependence of the phase velocity (dispersion) can be calculated from a detailed knowledge of a material's frequency dependent attenuation properties and the appropriate Kramers–Krönig transformation. Our laboratory published an approximate nearly local form of the Kramers–Krönig relations in 1978, which yielded good agreement between measurements of dispersion and predictions from attenuation experiments on tissue specimens.[3,4] The range of applicability of this original nearly local approximation of the Kramers–Krönig relations was later shown to be limited to cases where the attenuation is approximately linear with frequency. Szabo pointed out inconsistencies between the predictions of the approximate nearly local forms and experimental data from liquids with attenuation proportional to frequency power laws greater than one.[8] More recent experimental work from our lab[5] and others[6,7] has measured and successfully matched the ultrasonic dispersion, in a variety of attenuating media exhibiting frequency power laws from 1–2, with predictions of the exact (nonlocal) form of Kramers–Krönig relations published by Horton.[9,10] In this letter we use the exact (nonlocal) Kramers–Krönig relations instead of the original nearly local approximation to improve the description of a finite amplitude waveform propagating in a lossy, dispersive medium. Because the exact (nonlocal) Kramers–Krönig relation corrects the tendency of the original nearly local form to overestimate the actual amount of dispersion, it will be shown that the impact of dispersion on nonlinear plane-wave distortion is less than previously reported.[1]

II. METHODS

The numerical method employed is based on a frequency domain solution to the one-dimensional Burgers equation.[11] The method is modified to incorporate the asymptotic behavior of the harmonic components in the shocked waveform. This so-called ''spectral modification approach'' appears to have been first described by Khokhlova and Sapozhnikov at the 128th Meeting of the Acoustical Society in 1994[12] and subsequently published with Pishchal'nikov in 1996.[2] With the modified spectral approach, the evolving wave profile is regarded as a sum of smoothly varying components and an exact sawtooth waveform. The sawtooth waveform is characterized by an amplitude and a discontinuity location, θ_s.

The recent publication by Kashcheeva *et al.* extends the spectral modification approach to consider the propagation of shocked waves in media with attenuation described by a frequency power law.[1] An alternative time-domain method, incorporating an evolution equation equivalent to the integral

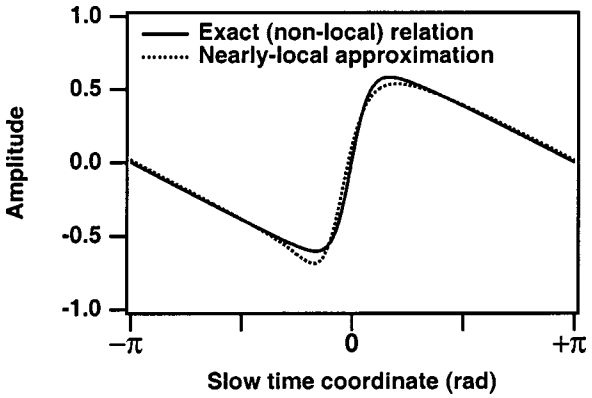


FIG. 1. Effect of the choice of Kramers–Krönig dispersion relationship for a simulated plane wave propagating through a medium with attenuation proportional to the square of the frequency ($y=2$). For this case, the original nearly local approximation (dashed curve) predicts a nonzero dispersion and thus yields a result different from the exact nonlocal solution (solid curve). The waveforms correspond to dimensionless simulation parameters of attenuation normalized distance $z=0.3$ and nonlinearity-to-dissipation ratio $\Gamma=10$.

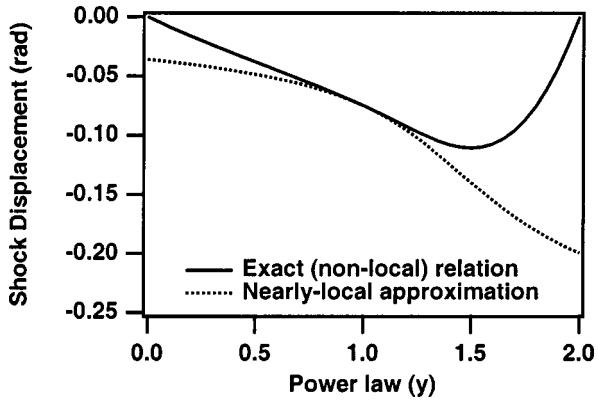


FIG. 2. Impact of Kramers–Krönig predicted dispersion on relative shock location for an intense plane wave propagating through a medium with attenuation properties given by a frequency power law, y. The original nearly local approximation (dashed curve) agrees with the exact, nonlocal solution (solid curve) for power law values close to one and overestimates the shock displacement elsewhere. The curves correspond to dimensionless simulation parameters of attenuation normalized distance $z=0.3$ and nonlinearity-to-dissipation ratio $\Gamma=10$.

dispersion relations, is outlined in a recent proceedings paper.[13] Kashcheeva *et al.* report results for an initial large amplitude sinusoid propagating in media with attenuation power laws of 1.0, 1.2, 1.42, 1.5, and 2.0, and employ the original nearly local approximate form of the Kramers–Krönig relations to determine the dispersion.[1] In the current study, we compare the impact of using the exact (nonlocal) Kramers–Krönig relations on finite amplitude wave propagation in media with an attenuation power law between 0–2.

To validate our approach, we coded three separate implementations of the modified spectral approach as outlined in Kashcheeva *et al.*, and achieved agreement with Figs. 1, 2(a), 3, and 4 in the original work[14] with all three sets of code. Two of the simulations were run by propagating and mixing a user selected maximum number of harmonic components (typically 50) with a fixed step size. One of these two was implemented with a graphical interface in IGOR Pro (Wavemetrics, Inc., Lake Oswego, OR) and the second was coded with a command line interface in Mathematica (Wolfram Research, Inc., Champaign, IL). The third implementation was also coded in Mathematica but utilized the built-in *NDSolve* function to find a numerical solution to the system of nonlinear differential equations (one for each harmonic considered, 50) and boundary conditions. In all of the cases, the dispersion was calculated from the assumed power law form of the attenuation with both the original nearly local approximate form of the Kramers–Krönig relationship [Eq. (6) in Ref. 3, Eq. (8) in Ref. 1] and the exact (nonlocal) form [Eqs. (6), (10), and (13) in Ref. 5]. All simulations were performed on an Apple Macintosh G3/400 PowerPC computer.

III. RESULTS

In the current study the attenuation frequency power law was varied from 0.0 to 2.0 in steps of 0.05. All other material parameters were held constant, with values selected to coincide with choices made by Kashcheeva *et al.* The initial waveform was a sinusoid with a frequency of 1.0 MHz. The attenuation was 0.038 cm^{-1} and the small amplitude phase velocity was 1546 m/s at the 1 MHz fundamental frequency. The nonlinear parameter B/A was set to 7.0. The initial sinusoid had an amplitude of 4.5 MPa which, when combined with the above material properties, corresponds to a value of 10 for the dimensionless nonlinearity-to-dissipation ratio (Gol'dberg number, Γ). This selection of simulation parameters is representative of an experiment that could be performed with butylene glycol, a liquid that has been previously considered as a tissue mimicking material.[15]

Figure 1 shows the distorted waveforms including effects of dispersion determined from both the nearly local approximation and the exact (nonlocal) form of the Kramers–Krönig relations for the case of the attenuation proportional to the square of the frequency ($y=2$). The curves depict a snapshot of the evolving progressive wave at a distance $x=7.9$ cm from the source. This propagation distance corresponds to an attenuation normalized length of $z=0.3$ ($z=x\cdot\alpha_0$) in the Kashcheeva *et al.* manuscript. The two ''slow time'' waveforms are somewhat similar but exhibit several distinctive features. The differences in the waveforms are of a comparable nature and magnitude to differences observed by Averkiou and colleagues[13] for shocked waveforms generated in media with several attenuation power law behaviors (y). Comparing one simulated shocked waveform ($y=1.1$) to another ($y=2.0$) in their second figure these authors state that ''the smooth parts of the waveform are close for different values'' of the attenuation power law whereas ''the structure of the shock front is quite different.''

Another notable distinction between the two waveforms in Fig. 1 of the current work is the relative location of the shock front within one period of the slow time wave. The exact (nonlocal) form of the Kramers–Krönig relations correctly predicts zero dispersion for the attenuation proportional to frequency squared case ($y=2$). This is apparent in the steep shock front passing through the origin exactly halfway between the beginning and the end of the cycle. The

original nearly local approximate form of the Kramers–Krönig relations predicts a nonzero dispersion for $y=2$. The resulting shock front is observed to be shifted to the left in the slow time reference frame (which corresponds to forward in time in the real time reference frame).

Figure 2 depicts the relative slow time shock displacement as a function of attenuation power law y for both the nearly local approximate form and the (nonlocal) exact form of the Kramers–Krönig relations. The location of the sawtooth component discontinuity in the slow time coordinate system can be determined from the phase angle between the Nth and the $(N-1)$th harmonic [as given by Eq. (12) in Ref. 1]:

$$\theta_s = \arg\left(\frac{C_N}{C_{N-1}}\right), \tag{2}$$

where C_N is the complex amplitude of the Nth harmonic. Without dispersion, the angle between any two successive harmonics would be a constant $-\pi/2$. This constant phase progression arises from a multiplicative value of $-i$ in the solution of the system of coupled, nonlinear differential equations. In the presence of dispersion, this angle will depart from the $-\pi/2$ value and the shock front will be displaced from the center of the distorted time-domain waveform.

IV. CONCLUDING REMARKS

For the specific range of simulation parameters considered, the largest discrepancy between the exact (nonlocal) Kramers–Krönig relation and the original nearly local approximate form occurs when the attenuation is proportional to the square of the frequency. The exact (nonlocal) relationship specifies that there is zero dispersion for this case whereas the nearly local approximation incorrectly predicts an anomalous dispersion of approximately 3% for the frequency range from 1–50 MHz (and approximately 1% for 1–20 MHz). Because of the increased velocity attributed to the higher harmonic components the time domain waveform distorts asymmetrically about the shock front and the relative location of the shock is displaced. The maximal difference in the shock displacement of 0.2 radians corresponds to an overestimation of 1 m/s (approximately 0.06%) for the average velocity of the shock front when using the original nearly local Kramers–Krönig approximation.

Numerical methods are commonly employed to model the combined effects of nonlinearity, attenuation, dispersion, and diffraction on a propagating finite amplitude wave. One might initially think that the effects of dispersion could be ignored because the total change in phase velocity over a given experimental bandwidth is modest. However, as a finite amplitude wave propagates and progressively distorts, tens or even hundreds of integer harmonic frequencies may be required to accurately describe the resulting waveform. Each of these higher harmonics may have a different small signal velocity as determined by the host medium's dispersion. Because of the range of frequencies inherent in the large number of harmonics, changes in phase velocity may need to be considered even for materials exhibiting only modest dispersion.

ACKNOWLEDGMENT

This work was supported by in part by NIH Grant HL40302.

[1] S. S. Kashcheeva, O. A. Sapozhnikov, V. A. Khokhlova, M. A. Averkiou, and L. A. Crum, ''Nonlinear distortion and attenuation of intense acoustic waves in lossy media obeying a frequency power law,'' Akust. Zh. **46**, 211–219 (2000) [English translation: Acoust. Phys. **46**, 170–177 (2000)].
[2] Yu. A. Pishchal'nikov, O. A. Sapozhnikov, and V. A. Khokhlova, ''A modification of the spectral description of nonlinear acoustic waves with discontinuities,'' Akust. Zh. **42**, 412–417 (1996) [English translation: Acoust. Phys. **42**, 362–367 (1996)].
[3] M. O'Donnell, E. T. Jaynes, and J. G. Miller, ''General relationships between ultrasonic attenuation and dispersion,'' J. Acoust. Soc. Am. **63**, 1935–1937 (1978).
[4] M. O'Donnell, E. T. Jaynes, and J. G. Miller, ''Kramers–Kronig relationship between ultrasonic attenuation and phase velocity,'' J. Acoust. Soc. Am. **60**, 696–701 (1981).
[5] Kendall R. Waters, Michael S. Hughes, Joel Mobley, Gary H. Brandenburger, and James G. Miller, ''On the applicability of Kramers–Kronig relations for ultrasonic attenuation obeying a frequency power law,'' J. Acoust. Soc. Am. **108**, 556–563 (2000).
[6] Ping He, ''Simulation of ultrasound pulse propagation in lossy media obeying a frequency power law,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**, 114–125 (1998).
[7] Thomas L. Szabo, ''Causal theories and data for acoustic attenuation obeying a frequency power law,'' J. Acoust. Soc. Am. **97**, 14–24 (1995).
[8] Thomas L. Szabo, ''Time domain wave equations for lossy media obeying a frequency power law,'' J. Acoust. Soc. Am. **96**, 491–500 (1994).
[9] C. W. Horton, ''Dispersion relations in sediments and sea water,'' J. Acoust. Soc. Am. **55**, 547–549 (1974).
[10] C. W. Horton, ''Comment on 'Kramers–Kronig relationship between ultrasonic attenuation and phase velocity', '' J. Acoust. Soc. Am. **70**, 1182 (1981).
[11] Jerry H. Ginsberg and Mark F. Hamilton, ''Computational Methods,'' in *Nonlinear Acoustics*, edited by Mark F. Hamilton and David T. Blackstock (Academic, San Diego, CA, 1998), pp. 309–342.
[12] V. A. Khokhlova and O. A. Sapozhnikov, ''Modificaiton of the spectral method for describing nonlinear acoustic waves containing shocks,'' J. Acoust. Soc. Am. **96**, 3321 (1994).
[13] M. A. Averkiou, L. A. Crum, V. A. Khokhlova, and O. V. Rudenko, ''Nonlinear waveform distortion and energy attenuation of intense acoustic waves in biological tissue,'' presented at the 14th International Symposium on Nonlinear Acoustics, Nanjing, China (1996).
[14] We note the presence of minor typographical errors in Eqs. (9), (10), and (13) in both the original Russian and the English translation of Ref. 1 above.
[15] B. Granz, ''Measurement of shock wave properties after the passage through a tissue mimicking material,'' Proc.-IEEE Ultrason. Symp. **CH3468-6**, 1847–1851 (1994).

Sound diffraction by an underwater ridge with soft finite impedance

Ji-Xun Zhou and Xue-Zhen Zhang
Institute of Acoustics, The Chinese Academy of Sciences, Beijing 100080, China
and School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332

James S. Martin and Yves H. Berthelot
School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332

(Received 1 October 1999; revised 28 July 2000; accepted 13 February 2001)

Laboratory experiments have been conducted in a water tank to investigate sound propagation over an underwater ridge with soft, finite impedance, very low density (0.12 kg/m^3) and sound speed (250 m/s), and very high attenuation (2–3 dB/cm in the frequency range of interest of 50–85 kHz). The diffracted acoustic pressure was measured along the ridge surface and along a vertical axis behind the ridge. Good agreement was found between predictions from the theory of matched asymptotic expansions (MAE) and the experimental data. The diffracted sound field is quite sensitive to the acoustic attenuation of diffraction material, thus showing potential as an inversion technique to characterize soft material attenuation.
[DOI: 10.1121/1.1362686]

PACS numbers: 43.30.Es, 43.30.Dr, 43.30.Zk [SAC-B]

I. INTRODUCTION

Many features of sound propagation over complex topography can be understood by studying sound propagation over grounds of simple shapes. Indeed, sound propagation over a simple convex impedance surface has been the object of both theoretical and experimental investigations in the field of outdoor sound propagation in recent years.[1–6] These results should be applicable to high-frequency, short-range sound propagation in the ocean over an undulating seabed, including seabottom ridges and hills. For instance, there is useful analogy between the two physical problems shown in Fig. 1, which is taken from Ref. 6: curved ray path propagation in an inhomogeneous medium above a flat boundary, and diffraction and creeping waves around a curved surface in a homogeneous medium. The relationship between the problems in Figs. 1(a) and (b) has been discussed.[1,2,6,7] In a deep ocean, the sound velocity increases with water depth, and the curved ray paths are similar to the ones shown in Fig. 1(a). (For example, see Fig. 2 of Ref. 8.) Thus, laboratory scaled experiments on sound propagation over an underwater ridge may offer some physical insight to acoustic interactions with a seabed.

In this letter, experimental results obtained in a water tank to study sound propagation over an underwater ridge with soft finite impedance are compared with predictions from the MAE theory of Fock and Pierce.[7,9] The letter is organized as follows. First, we briefly introduce the theoretical expressions for the ridge diffraction, derived from the MAE method. The experimental setup in a water tank is reported in Sec. III. Particular attention is given to the choice of material used in the experiment to represent the soft, curved, boundary. In the fourth section, experimental data are compared with theoretical predictions along the ridge surface, and along a vertical axis behind the ridge. Some conclusions and discussions are given in the last part.

II. THEORY

The MAE method was used by Fock[9] in the 1940s and 1950s to analyze radio wave propagation around the earth. Pierce and Main applied the method to sound diffraction problems.[2] For clarity, the main results are given again here.

The two-dimensional prototype problem is shown in Fig. 2. Let a harmonic plane wave $P_i e^{i(kx-\omega t)}$ be incident upon a smooth, convex, ridge with radius of curvature R such that $kR \gg 1$, where $k=\omega/c$ is the wave number in the fluid of density ρ and sound speed c. The diffracting boundary has an acoustic impedance Z_b. The insertion loss at a field point behind the ridge is defined as

$$IL = -20\log_{10}|p/p_0| = -20\log_{10}|G|, \tag{1}$$

where p is the sound pressure at the field point with the diffraction ridge, and p_0 is the sound pressure in the free field at the same point, without the ridge. The general expression for G is given by

$$G(\xi,\eta,q) = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty}\left[v(\alpha-\eta) - \frac{v'(\alpha)-qv(\alpha)}{w_1'(\alpha)-qw_1(\alpha)} \times w_1(\alpha-\eta)\right]e^{i\alpha\xi}\,d\alpha, \tag{2}$$

where the normalized parabolic coordinates (ξ, η) are defined by

$$\xi=(u/R)(kR/2)^{1/3}, \quad \eta=(2v/R)(kR/2)^{2/3},$$

with

$$u=2^{-1/2}\{[(2Ry+R^2)^2+4R^2x^2]^{1/2}-(2Ry+R^2)\}^{1/2},$$

$$v=R[(x/u)-1],$$

and the nondimensional scaled admittance q is defined by

$$q=i(kR/2)^{1/3}(\rho c/Z_b).$$

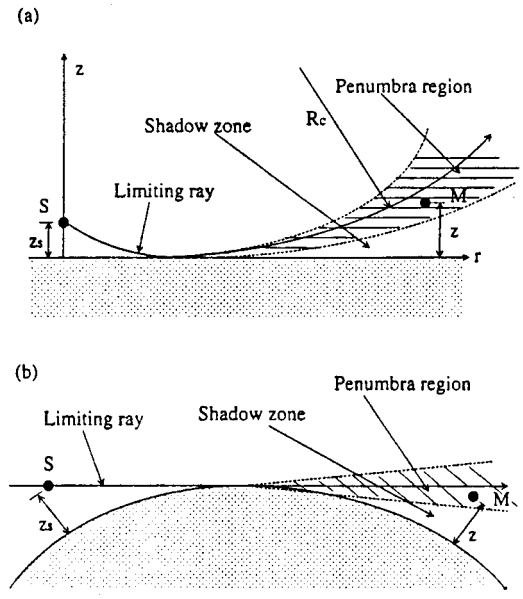


FIG. 1. (a) Propagation over a flat boundary in an inhomogeneous, upper refractive atmosphere. (b) Propagation over a convex surface in a homogeneous medium.

The axis $v=0$, or $\eta=0$, defines the curved boundary. In Eq. (2), the prime denotes a derivative with respect to the argument and the functions v and w_1 are the Fock functions which are directly related to the Airy function of complex argument, $\mathrm{Ai}(z)$, by

$$v(\alpha-\eta)=\sqrt{\pi}\,\mathrm{Ai}(\alpha-\eta),$$

$$w_1(\alpha-\eta)=2\sqrt{\pi}e^{i\pi/6}\mathrm{Ai}(e^{i2\pi/3}(\alpha-\eta)).$$

On the surface of the ridge, Eq. (2) reduces to

$$G(\xi,0,q)=\frac{1}{\sqrt{\pi}}\int_0^\infty \frac{e^{-is\xi/2}e^{-s\xi\sqrt{3}/2}}{w_2'(s)-e^{i2\pi/3}qw_2(s)}\,ds + \frac{1}{\sqrt{\pi}}\int_0^\infty \frac{e^{is\xi}}{w_1'(s)-qw_1(s)}\,ds, \tag{3}$$

where Fock function w_2 is defined as

$$w_2(z)=2\sqrt{\pi}e^{-i\pi/6}\mathrm{Ai}(ze^{-i2\pi/3}).$$

Behind the ridge, G can approximately be decomposed into two parts: a rigid knife-edge diffraction (Φ) superposed on what Fock calls a ''background'' (Ψ):

$$G=\Phi+\Psi, \tag{4}$$

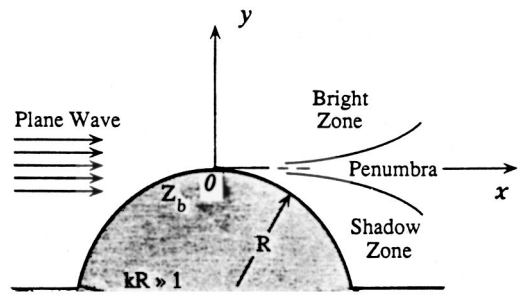


FIG. 2. Geometry of the 2-D diffraction problem.

where

$$\Psi=-\frac{1}{\sqrt{\pi}}\int_0^\infty e^{is\xi e^{i2\pi/3}}\frac{v'(s)-qe^{i2\pi/3}v(s)}{w_2'(s)-qe^{i2\pi/3}w_2(s)} \times w_2(s-\eta e^{-i2\pi/3})\,ds - \frac{1}{\sqrt{\pi}}\int_0^\infty e^{is\xi}\frac{v'(s)-qv(s)}{w_1'(s)-qw_1(s)}w_1(s-\eta)\,ds \tag{5}$$

and

$$\Phi=[e^{-i\xi^3/3}e^{i\xi\eta}\mathrm{H}(Y)-(e^{i\pi/4}/\sqrt{2})e^{i(2/3)\eta^{3/2}}A_D(Y)], \tag{6}$$

where $\mathrm{H}(Y)$ is the Heaviside step function and $A_D(Y)$ is the knife-edge diffraction integral defined in Ref. 7. The argument Y represents a dimensionless height which is positive in the bright zone and negative in the shadow zone. It is defined as

$$Y=-(2/\pi)^{1/2}\eta^{1/4}(\xi-\eta^{1/2}). \tag{7}$$

All above expressions involve integrals with integrands that can be expressed in terms of Airy and Fock functions of complex argument. An algorithm capable of calculating these function to high precision for arbitrary complex argument has already been developed.[2] Based on the above set of equations, a computer program has been written to calculate the sound propagation over a diffraction ridge.

III. EXPERIMENT

It is well known[7,10] that complicated diffraction problems where the boundary or the wave front is not flat can be greatly simplified by assuming that the diffracting surface is locally reacting. In fact, the theoretical results of Sec. II were derived under the assumption that the diffraction ridge had such an impedance boundary condition. Thus, for meaningful comparison between experimental data and theoretical predictions, the diffraction material used in the experiment must be carefully chosen to satisfy this approximation.

In general, the boundary impedance can be expressed as

$$Z_b(\theta)=\rho_b c_b^*/\cos\theta,$$

where ρ_b is the density of the material, c_b^* is its complex sound speed to reflect the dissipative nature of the material, and θ is the refraction angle in the lower diffraction medium. If the sound velocity in the diffraction body is much less than in water, the refraction angle approaches zero and the impedance of the boundary does not depend on the angle of incidence, thus satisfying the requirement of local reaction.[10] Thus, a light and soft closed-cell foam material with very low bulk sound speed was chosen to cover the diffraction ridge. Measurements made independently at the Institute of Acoustics in Beijing[11] on the same material indicate that its density ρ_b is about 0.12 g/cm^3 and the bulk sound speed c_b is about 250 m/s, which is indeed much smaller than the speed in the water. The complex impedance of the diffracting surface is $Z_b=\rho_b c_b^*$, where c_b^* is the complex sound speed in the material, whose imaginary part depends directly on the attenuation in the material, α_b, that is inversed from experimental data in Sec. IV. The cylindrical ridge was made by

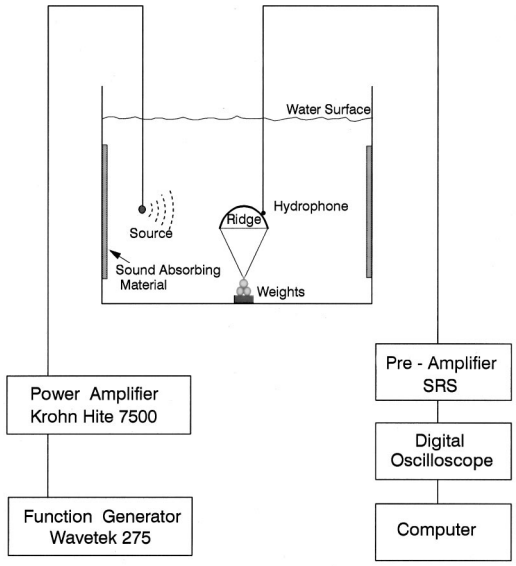


FIG. 3. Experimental arrangement.

bending a 3.4-mm-thick aluminum sheet metal to a radius of curvature of 36 cm. The length of the ridge was 61 cm. The surface was covered with a 2.5-cm-thick closed-cell foam material for modeling a soft, locally reacting, finite impedance surface.

Laboratory measurements were conducted in a 3780-liter cylindrical tank. The diagram in Fig. 3 shows the experimental arrangement. A gated sine wave (tone burst of 5 cycles at 50 and 85 kHz) from a function generator was input to a power amplifier to drive piezoelectric ceramic spherical source with 19-mm diameter. The source was placed far away from the apex of the ridge at the depth of the critical grazing ray. The signal was received by a hydrophone, input to a preamplifier (with built-in bandpass filters), and displayed on a digital oscilloscope. The data were analyzed by a computer. The hydrophone captured the signals at various field points behind the ridge and at a reference location in front of the ridge. The ratio of the measured signal amplitude at the field point to that of a reference signal, corrected by the spreading loss and the water absorption, is interpreted as the insertion loss *IL* defined by Eq. (1).

IV. RESULTS

Figures 4 and 5 show the measured insertion loss as a function of the dimensionless arclength defined by

$$\xi=(s/R)(kR/2)^{1/3},$$

where s is the actual arclength. The solid lines are the theoretical results based on Eq. (3). The surface impedance Z_b used to calculate the admittance q is calculated using the values of density $\rho_b=0.12\ \mathrm{g/cm^3}$ and the bulk sound speed $c_b=250\ \mathrm{m/s}$. The imaginary part of the sound speed was determined by a best fit technique between the experimental data and the predicted values. The value of α_b was thus determined by an inverse technique. At 50 kHz, sound attenuation was found to be $\alpha_b=2.32\ \mathrm{dB/cm}$, while at 85 kHz it was found to be $\alpha_b=3.04\ \mathrm{dB/cm}$. It is interesting to note that the predicted values of the insertion loss on the diffracting surface are quite sensitive to the attenuation in the material, thus providing a good method to determine, albeit indirectly, the value of attenuation α_b. In contrast, they were not very sensitive to the value of sound speed c_b because of the very low values of ρ_b and c_b. The experimental data and the MAE predictions are in good agreement, even very deep in the shadow zone, thus confirming that the MAE theory can be used to predict sound diffraction by curved surfaces in underwater applications.

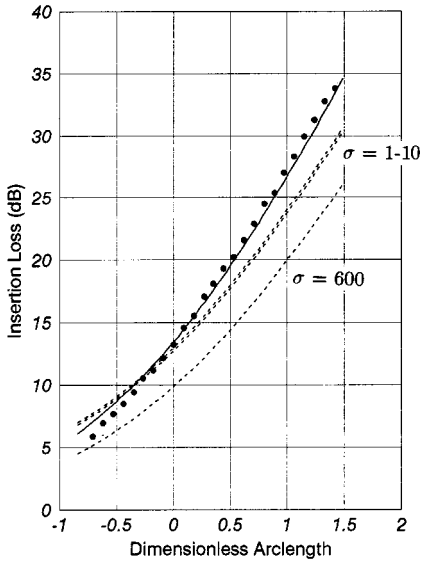


FIG. 4. Insertion loss on the diffracting surface at $f=50$ kHz. Dots: experimental data; solid line: prediction from the MAE theory with Z_b calculated from $\rho_b=0.12\ \mathrm{g/cm^3}$, $c_b=250\ \mathrm{m/s}$, and $\alpha_b=2.32\ \mathrm{dB/cm}$; dotted lines: MAE theory with Z_b calculated from the Delany–Bazley model[12] with $\sigma=1$, 10 and 600 cgs-rayls/cm.

As an aside, it was also noted that, as expected, the Delany and Bazley model[12] used to model fibrous absorbent

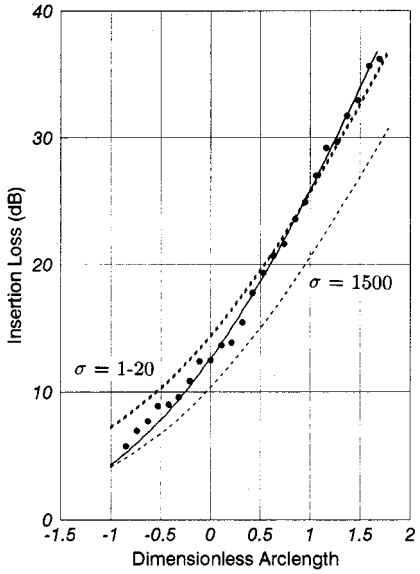


FIG. 5. Insertion loss on the diffracting surface at $f=85$ kHz. Dots: experimental data; solid line: prediction from the MAE theory with Z_b calculated from $\rho_b=0.12\ \mathrm{g/cm^3}$, $c_b=250\ \mathrm{m/s}$, and $\alpha_b=3.04\ \mathrm{dB/cm}$; dotted lines: MAE theory with Z_b calculated from the Delany–Bazley model[12] with $\sigma=1$, 20 and 1500 cgs-rayls/cm.

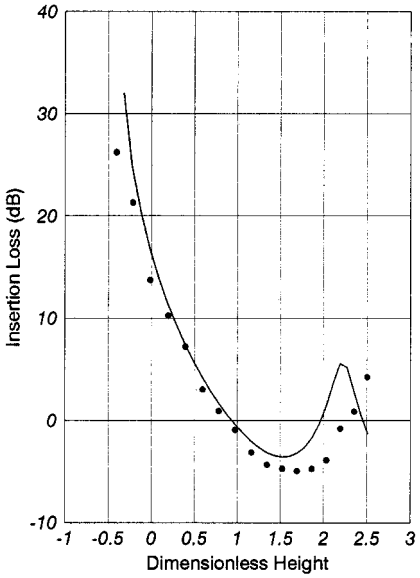


FIG. 6. Insertion loss along a vertical axis at a distance $x=15$ cm behind the apex, from the deep shadow to the bright zone, at $f=50$ kHz. Dots: experimental data; solid line: prediction from the MAE theory with Z_b calculated from $\rho_b=0.12$ g/cm^3, $c_b=250$ m/s, and $\alpha_b=2.32$ dB/cm.

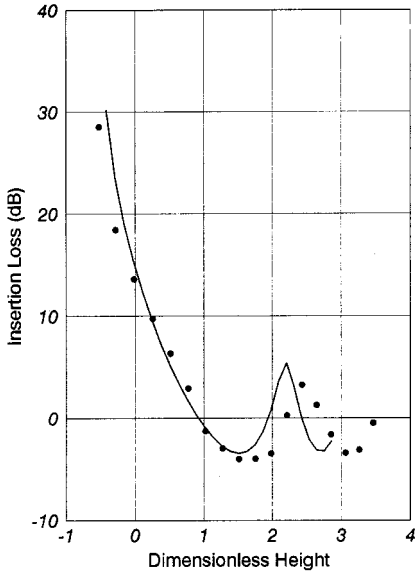


FIG. 7. Insertion loss along a vertical axis at a distance $x=15$ cm behind the apex, from the deep shadow to the bright zone, at $f=85$ kHz. Dots: experimental data; solid line: prediction from the MAE theory with Z_b calculated from $\rho_b=0.12$ g/cm^3, $c_b=250$ m/s, and $\alpha_b=3.04$ dB/cm.

materials could not be used to fit the data of Figs. 4 and 5. In this model, the acoustic impedance is given by $Z_b=\rho c[1+0.0571(\rho f/\sigma)^{-0.75}+i0.087(\rho f/\sigma)^{-0.73}]$, where ρc is the impedance of upper (air) medium, f is the frequency in Hz, and σ is the effective flow resistivity in mks-Rayls/m. The three dotted lines in Fig. 4 represent predictions using values of $\sigma=1$, 10, and 600 cgs Rayls/cm. The three dotted lines in Fig. 5 represent predictions using $\sigma=1$, 20, and 1500 cgs Rayls/cm.

Figures 6 and 7 show the measured insertion loss as a function of dimensionless height behind the ridge at a distance $x=15$ cm behind the apex. The circles are the experimental results. The solid lines are the theoretical predictions by Eqs. (4)–(6). The surface impedance of the diffraction ridge is calculated from the values of density $\rho_b=0.12$ g/cm^3, sound speed $c_b=250$ m/s, and attenuation $\alpha_b=2.32$ dB/cm at 50 kHz, and $\alpha_b=3.04$ dB/cm at 85 kHz as found from the best fit described previously for the case of the sound field measured on the surface. Again, the MAE theory predicts very well the sound field between the deep shadow and the penumbra (i.e., up to a dimensionless height of unity). It is not expected to be particularly accurate in the bright zone above the ridge, a result that is confirmed by the poor predictions at dimensionless heights greater than unity in Figs. 6 and 7.

V. CONCLUSION

The sound field around an underwater convex surface covered with a soft, closed cell material of finite impedance is well predicted by the MAE theory of Fock and Pierce. The insertion loss is very sensitive to the acoustic attenuation coefficient of the diffraction material. This fact may be used to measure the attenuation coefficient of soft underwater materials by a simple inversion technique. The research may also be used to simulate high-frequency acoustic interactions with undulating seabed, or to design underwater barriers.[13]

ACKNOWLEDGMENTS

This work was supported by ONR and CAS. The authors wish to thank Dr. Jeff Simmen for helpful discussions and comments. Yun S. Chase is appreciated for his help with the experiment.

[1] A. Berry and G. A. Daigle, ''Controlled experiments of the diffraction of sound by a curved surface,'' J. Acoust. Soc. Am. **83**, 2047–2058 (1988).
[2] A. D. Pierce and G. L. Main, ''Computational algorithms for the matched asymptotic expansion solution of high frequency acoustic wave diffraction's by curved surfaces of finite impedance,'' in *Advances in Computer Methods for Partial Differential Equations—6* (IMACS, Rutgers Univ., New Brunswick, NJ, 1987), pp. 187–194.
[3] J. X. Zhou and Y. B. Berthelot, ''Sound propagation over a large smooth ridge of finite impedance,'' in 14th ICA Proceedings, K2-4 (Beijing, China, 1992).
[4] Y. H. Berthelot and J. X. Zhou, ''Scale model experiments on the validity of the matched asymptotic expansions theory for sound diffraction by curved surface of finite impedance,'' J. Acoust. Soc. Am. **93**, 605–608 (1993).
[5] Y. H. Berthelot, ''A note on the acoustic penumbra behind a curved surface,'' J. Acoust. Soc. Am. **99**, 2428–2429 (1996).
[6] K. M. Li, Q. Wang, and K. Attenborough, ''Sound propagation over impedance surface,'' J. Acoust. Soc. Am. **104**, 2683–2691 (1998).
[7] A. D. Pierce, *Acoustics: An Introduction to its Physical Principles and Applications* (McGraw–Hill, New York, 1981).
[8] H. Medwin and J. C. Novarini, ''Modified sound refraction near a rough ocean bottom,'' J. Acoust. Soc. Am. **76**, 1791–1796 (1984).
[9] V. A. Fock, *Electromagnetic Diffraction and Propagation Problems* (Pergamon, New York, 1965).
[10] L. M. Brekhovskikh, *Waves in Layered Media* (Academic, New York, 1980), 2nd ed.
[11] Y. J. Wang and T. G. Zhang (private communication, 1996).
[12] M. E. Delany and E. N. Bazley, Appl. Acoust. **3**, 105–116 (1970).
[13] K. Kuramoto, S. Yamaguchi, K. Oimatsu, and S. Kuwahara, ''Sound attenuation by barriers in underwater,'' J. Acoust. Soc. Jpn. **15**, 13–16 (1994).

Vector potential in spherical polar coordinates

F. Ahmad and Ahsan-ul-Haq
Department of Mathematics, Quaid-I-Azam University, Islamabad, Pakistan

(Received 1 July 2000; accepted for publication 27 February 2001)

We comment on a recent paper [J. Acoust. Soc. Am. **104**, 2836–2844 (1998)] and point out certain errors in the mathematical formulation of the problem. © *2001 Acoustical Society of America.*
[DOI: 10.1121/1.1367249]

PACS numbers: 43.40.Ey, 43.20.Gp [ANN]

I. INTRODUCTION

Recently Sessarego *et al.*[1] have discussed scattering by an elastic sphere embedded in an elastic medium. They have represented the displacement $\mathbf{u}$ as

$$\mathbf{u}=\text{grad}\,\phi+\text{curl}\,\boldsymbol{\psi}, \tag{1}$$

where ϕ and $\boldsymbol{\psi}$ are, respectively, the scalar and vector potentials and satisfy the equations

$$\nabla^2\phi+k_L^2\phi=0, \tag{2}$$

$$\nabla^2\boldsymbol{\psi}+k_T^2\boldsymbol{\psi}=0. \tag{3}$$

We use the same notation as in Ref. 1. Since the incident wave is a P-wave the azimuthal symmetry allows the vector potential to be of the form $\boldsymbol{\psi}=(0,0,\psi)$. In Ref. 1 it is claimed that the functions ϕ and ψ satisfy the following equations in spherical polar coordinates (r,θ,φ),

$$\nabla^2_{r\theta}\phi+k_L^2\phi=0, \tag{4}$$

$$\nabla^2_{r\theta}\psi+k_T^2\psi=0, \tag{5}$$

where

$$\nabla^2_{r\theta}=\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right)+\frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right). \tag{6}$$

In this letter we wish to point out that Eq. (5) is incorrect. Since Eqs. (2) and (3) and the assumption $\boldsymbol{\psi}=(0,0,\psi)$ appear to lead *naturally* to Eqs. (4) and (5), it is all the more important to draw the attention of workers in elasticity to a feature of curvilinear coordinates, which is easy to overlook.

II. CORRECT EQUATIONS

Although the same symbol ∇^2 appears in both Eqs. (2) and (3), it has a different representation when each equation is written in an *invariant vector form*. Whereas Eq. (2) has the representation

$$\text{div grad}\,(\phi)+k_L^2\phi=0, \tag{7}$$

on the other hand, Eq. (3) is represented as

$$\text{grad div}\,(\boldsymbol{\psi})-\text{curl curl}(\boldsymbol{\psi})+k_T^2\boldsymbol{\psi}=0. \tag{8}$$

In spherical polar coordinates, Eq. (8) leads to three coupled equations involving all three components of $\boldsymbol{\psi}$ (Ref. 2, p. 76). Even when $\boldsymbol{\psi}$ possesses only a single component, Eq. (8) becomes

$$\nabla^2_{r\theta}\psi-\frac{1}{r^2\sin^2\theta}\psi+k_T^2\psi=0. \tag{9}$$

From Eq. (1) it follows that

$$u_r=\frac{\partial\phi}{\partial r}+\frac{1}{r\sin\theta}\left[\frac{\partial}{\partial\theta}(\psi\sin\theta)\right], \tag{10a}$$

$$u_\theta=\frac{1}{r}\frac{\partial\phi}{\partial\theta}-\frac{1}{r}\left[\frac{\partial}{\partial r}(r\psi)\right], \tag{10b}$$

$$u_\varphi=0. \tag{10c}$$

It is well known[3] that the solution of Eq. (9) can be represented as

$$\psi_1=\sum_{n=0}^{\infty} i^n(2n+1)a_nR_n(k_Tr)\frac{d}{d\theta}P_n(\cos\theta), \tag{11}$$

where R_n stands for the spherical Bessel function j_n or $h_n^{(1)}$ and a_n are arbitrary constants. Solutions of Eqs. (4) and (5) are, respectively,

$$\phi=\sum_{n=0}^{\infty} i^n(2n+1)b_nR_n(k_Lr)P_n(\cos\theta), \tag{12}$$

and

$$\psi_2=\sum_{n=0}^{\infty} i^n(2n+1)c_nR_n(k_Tr)P_n(\cos\theta). \tag{13}$$

The boundary conditions require that the displacement vector as well as the components σ_{rr}, $\sigma_{r\theta}$, and $\sigma_{r\varphi}$ of the stress tensor be continuous at the interface $r=a$. To satisfy them it is *essential* that when ϕ and ψ are substituted in Eqs. (10), the angular dependence factors out in each term of the series on the right. When we substitute in Eq. (10a) ψ_1 and ϕ, as given by Eqs. (11) and (12), we obtain

$$u_r=\sum_{n=0}^{\infty} i^n(2n+1)\left[k_Lb_nR_n'(k_Lr)-a_n\frac{n(n+1)}{r}R_n(k_Tr)\right]P_n(\cos\theta). \tag{14}$$

On the other hand, when we put ϕ and ψ_2 in Eq. (10a) the following expression results:

$$u_r = \sum_{n=0}^{\infty} i^n (2n+1) \left[k_L b_n R_n'(k_L r) P_n(\cos\theta) + \frac{c_n}{r} R_n(k_T r) \left\{ \frac{d}{d\theta} P_n(\cos\theta) + \cot\theta P_n(\cos\theta) \right\} \right]. \tag{15}$$

It is clear that Eq. (14) appears in the proper form as required by the boundary conditions whereas Eq. (15) does not.

ACKNOWLEDGMENTS

Financial support for this work was provided by the Quaid-I-Azam University Research Fund. Also, we are grateful to Professor Uberall for a suggestion to improve the present letter.

[1] J. P. Sessarego, J. Sageloli, and H. Uberall, ''Scattering by an elastic sphere embedded in an elastic isotropic medium,'' J. Acoust. Soc. Am. **104**, 2836–2844 (1998).

[2] J. D. Achenbach, *Wave Propagation in Elastic Solids* (North-Holland, Amsterdam, 1980).

[3] G. F. C. Duff and D. Naylor, *Differential Equations of Applied Mathematics* (Wiley, New York, 1968).

The effect of speechreading on masked detection thresholds for filtered speech

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

(Received 30 October 2000; accepted for publication 13 February 2001)

Detection thresholds for spoken sentences in steady-state noise are reduced by 1–3 dB when synchronized video images of movements of the lips and other surface features of the face are provided. An earlier study [K. W. Grant and P. F. Seitz, J. Acoust. Soc. Am. **108**, 1197–1208 (2000)], showed that the amount of masked threshold reduction, or *bimodal coherence masking protection* (BCMP), was related to the degree of correlation between the rms amplitude envelope of the target sentence and the area of lip opening, especially in the mid-to-high frequencies typically associated with the second (*F*2) and third (*F*3) speech formants. In the present study, these results are extended by manipulating the cross-modality correlation through bandpass filtering. Two filter conditions were tested corresponding roughly to the first and second speech formants: *F*1 (100–800 Hz) and *F*2 (800–2200 Hz). Results for *F*2-filtered target sentences were comparable to those of unfiltered speech, yielding a BCMP of roughly 2–3 dB. Results for *F*1-filtered target sentences showed a significantly smaller BCMP of approximately 0.7 dB. These results suggest that the magnitude of the BCMP depends on both the spectral and temporal properties of the target speech signal. [DOI: 10.1121/1.1362687]

PACS numbers: 43.66.Dc, 43.66.Mk [SPB]

I. INTRODUCTION

Auditory–visual speech recognition is the most robust form of speech communication used by humans or machines, especially in noisy environments. Further understanding of the processes involved in human auditory–visual speech recognition will likely assist in the development of hearing aids that are more resistant to environmental noise and reverberation, as well as automatic speech recognition systems that are capable of achieving high rates of accuracy with speech spoken conversationally. Advancements in these two areas would have benefits for hearing-impaired individuals as well as normal-hearing persons who need to communicate in noisy or reverberant environments. This study addresses some fundamental questions regarding how vision and audition interact during speech recognition, and seeks to further understand the mechanisms by which watching a talker's lips and face during speech production helps to reduce interference from background noise. For automatic speech recognition, being able to segregate the speech signal from other extraneous sound sources, including speech from other talkers, is an essential first step in decoding the target speech signal into meaningful units. Understanding the perceptual mechanisms that allow human observers to accomplish this task through the use of speechreading should be of obvious benefit in transferring this knowledge to machine applications.

Recently, we have demonstrated that watching the movement of the lips and face can also improve the *detection* of speech (Grant and Seitz, 2000). Depending on the specific sentence, the improvement in detection thresholds can be between 1–3 dB. For all three sentences tested, a significant difference in thresholds was obtained between auditory–visual and auditory conditions, with greater sensitivity demonstrated for the auditory–visual conditions. There were also significant differences in thresholds across sentences. While the overall amount of masking protection due to visual speech cues may seem relatively small, a 1–3 dB improvement in masked thresholds can nevertheless make a substantial difference for speech understanding.

To account for their results, Grant and Seitz (2000) assumed that subjects have the ability to correlate the visible movements of the speech articulators (e.g., variation in the area of lip opening during speech production) and the acoustic speech envelope computed over time intervals corresponding roughly to the average duration of a syllable (333 ms). When this correlation is high (e.g., greater than 0.9) and the amplitude envelope is at a maximum relative to other peaks in the speech sample, there will be a positive effect of speechreading on detection thresholds. Grant and Seitz called this effect of speech detection thresholds *bimodal coherence masking protection* (BCMP), adapted from the earlier work of Gordon (1997a, b). The label BCMP is used to denote the fact that the information from one modality (in this case, visual) partially protects the target speech signal from the deleterious effects of noise. In the present study, the effects of speechreading on speech detection were explored further by manipulating the degree of correlation between the peak amplitude locations in the speech waveform and the lip-area function. This was accomplished by filtering target sentences using bandpass filters with different center frequencies and bandwidths. The purpose of the present experiment was to test the hypothesis that speechreading aids auditory detection of spoken sentences when the amplitude envelope peaks of the sentence coincides with a temporal location of high cross-modal correlation between area of lip opening and amplitude envelope. Since the correlation between area of mouth opening and speech amplitude envelope

a)Electronic mail: grant@tidalwave.net

tends to be the greatest for mid- to high-frequency speech signals (Grant and Seitz, 2000), the magnitude of the BCMP would be greater for speech targets containing mid-frequency energy than for speech targets containing low-frequency energy.

II. METHODS

A. Subjects

Six normally hearing subjects (mean age = 37.7 years) participated in the study. Subjects were screened to assure pure-tone air-conduction thresholds bilaterally of ≤20 dB HL at audiometric test frequencies 0.25–4.0 kHz and ≤30 dB HL at 6.0 kHz (ANSI, 1989). All subjects were native English speakers with normal or corrected-to-normal vision (visual acuity equal to or better than 20/30 as measured with a Snellen chart). Eligible subjects were paid $10.00 per hour to compensate them for their participation.

B. Stimuli

Speech materials consisted of video-recorded spoken sentences from the IEEE/Harvard sentence corpus (IEEE, 1969). The visual portion of each sentence was transferred to an optical disk recorder (Panasonic TQ-3031F). The audio portion of each sentence was digitized (16-bit A/D, 20-kHz sampling rate), filtered (8.5 kHz), normalized in level, and stored on a personal computer. Two sentences and their variants (as described below) were used. These were ''Both brothers wear the same size,'' and ''Watch the log float in the wide river.'' In the previous study (Grant and Seitz, 2000), these two sentences (denoted here as sentences 2 and 3 to be consistent with the previous study) provided approximately 1 and 2.5 dB of BCMP, respectively.

The target sentences were digitally bandpass-filtered using FIR filters with greater than 100 dB/oct attenuation outside the passband. Two bandpass filters were applied to each target sentence. One filter was centered primarily on the *F*1 speech region (100–800 Hz) whereas the second filter was centered primarily on the *F*2 speech region (800–2200 Hz). After filtering, the target audio sentences were scaled in amplitude so that the average rms levels were equivalent.

C. Procedure

Subjects were tested binaurally under headphones (Beyer Dynamic DT770) in a sound-treated both using an adaptive two-interval forced-choice (2IFC) tracking procedure. Masked thresholds for detecting speech were obtained under both auditory alone (A) and auditory–visual (AV) conditions. Each test block consisted of multiple interleaved tracks corresponding to the two different filtered target sentences and different test conditions (auditory and auditory–visual). The masking noise consisted of a white noise low-pass filtered at 8.5 kHz, its duration equal to the target sentence plus a random amount (additional 200–800 ms). The target sentence was temporally centered in the noise. For the AV conditions, video speech information of the talker saying the target sentence was presented in both observation intervals. Thus, the identical visual speech information was provided during both observation intervals. The filtered audio signals were realigned with the corresponding unprocessed audio signals on the optical disk to account for any audio delays that were introduced by the filtering process. Video signals were displayed on a 19-in. color monitor (SONY PVM 2030) positioned approximately 5 ft from the subject.

The subject's task was to identify the interval containing the target auditory sentence. The speech signal level was held constant at approximately 50 dB SPL. The intensity of the noise masker varied independently for each track according to a 3-up, 1-down adaptive tracking procedure targeting the 79% point on the psychometric function (Levitt, 1971). Each track was controlled independently and selected randomly on each trial. The initial step size of the digital noise attenuator (TDT PA4) was 3 dB during the first three reversals in the direction of the track. At that point the step size changed to 1 dB for additional six track reversals. Threshold estimates were computed as the mean of the noise levels obtained on the last six reversal points. Final thresholds for each of the target sentences in each of the conditions were the average of three separate estimates. If the standard error of the three estimates exceeded 1 dB, a fourth estimate was obtained and the final threshold value was the average of all four estimates.

Because there were two target sentences, two filter conditions (*F*1 and *F*2), and two modalities (A and AV), a total of eight interleaved tracks were run on each test block. Each test block took approximately 1.5 h to complete, allowing for frequent rest periods.

III. RESULTS

The mean speech-detection threshold, expressed in terms of the speech-to-noise ratio, was −22.36 dB under auditory–visual conditions and −21.13 dB under auditory conditions demonstrating a small but consistent BCMP. Figure 1 shows the magnitude of the BCMP for each target sentence as well as the average BCMP across sentences (conditions labeled AV_{WB} and AV_{O}, will be described later). As observed in the figure, both *F*1- and *F*2-filtered sentences resulted in a BCMP, but the magnitude of the BCMP for *F*2-filtered sentences was larger than that for the *F*1-filtered sentences (average $AV_{F1} = 0.76$, average $AV_{F2} = 1.70$) and the magnitude of the *F*2-filtered BCMP for sentence 3 was greater than that for sentence 2 ($AV_{F2} = 0.98$ dB for sentence 2 and 2.42 for sentence 3). Multiple *t*-tests with Bonferroni-adjusted probabilities were conducted to test whether the magnitude of the BCMP for each sentence and each filter condition was significantly greater than zero (i.e., no difference in masked threshold for auditory and auditory–visual conditions). Results showed that only the *F*2-filtered targets yielded significant BCMPs ($t = 4.45$, $p = 0.027$ and $t = 5.9$, $p = 0.008$ for sentence 2 and 3, respectively).

To further analyze these effects, a repeated measures ANOVA (Analysis of Variance) with BCMP as the dependent variable and filter condition and sentence as factors revealed a significant effect for filter [$F(1,5) = 6.9$, $p = 0.047$]. Neither sentence nor the interaction between sentence and

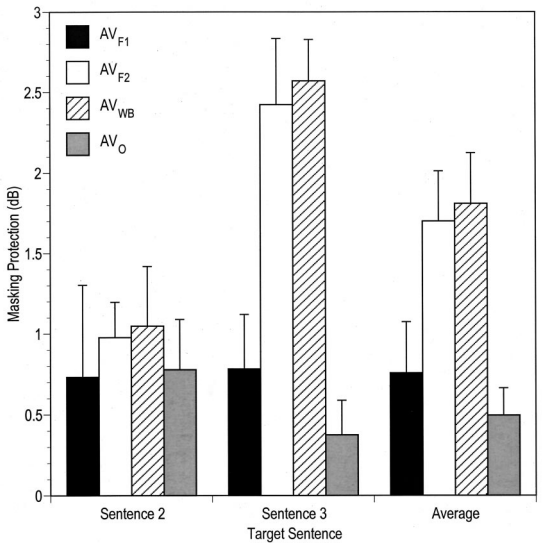


FIG. 1. Difference in auditory and auditory–visual masked detection thresholds (masking protection) for spoken filtered sentences. AV_{F1}: auditory–visual presentation of speech filtered between 100–800 Hz; AV_{F2}: auditory–visual presentation of speech filtered between 800–2200 Hz; AV_{WB}: auditory–visual presentation of wideband speech (100–8500 Hz); AV_O: auditory presentation of wideband speech preceded by visual orthography. Error bars show 1 standard deviation.

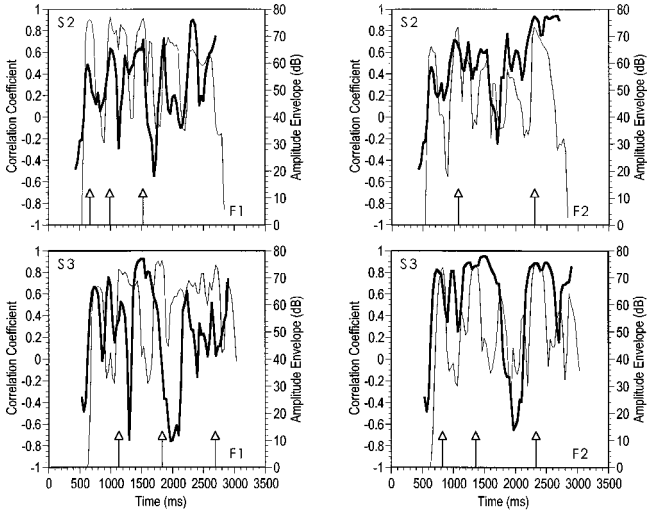


FIG. 2. Local correlations (bold line) between area of mouth opening and speech amplitude envelope (see text for explanation). For each panel, the left axis shows the Pearson product–moment correlation whereas the right axis shows the rms energy of the speech target signal averaged over successive 33-ms rectangular windows (thin line). Arrows show the temporal locations of the most prominent amplitude peaks in each target sentence (presumably the temporal regions most likely responsible for the detection response). The top two panels are for sentence S2, ''Both brothers wear the same size.'' The bottom two panels are for the sentence S3, ''Watch the log float in the wide river.'' The left two panels are for *F*1-filtered targets. The right two panels are for *F*2-filtered targets.

filter was significant, although the difference across sentences approached significance ($p = 0.077$).

Figure 1 also compares the present BCMP results with those reported earlier by Grant and Seitz (2000) obtained by the same six subjects for wideband target sentences (AV_{WB}) and for sentences presented auditorily with the aid of an orthographic display (AV_O) informing the subject of the exact text of the target sentence presented on each trial. The orthographic display was presented just prior to each test trial and lasted for 0.5 s. Knowing the text of the target sentence resulted in a small BCMP (approximately 0.5 dB), probably due to a slight reduction in informational masking and stimulus uncertainty (Watson and Kelly, 1981). The BCMP for low-frequency speech (*F*1 filtering) was only slightly greater than that for orthography, and like orthography, appears to be roughly independent of the target sentence. The BCMP for wideband speech and for *F*2-filtered speech signals were nearly identical, suggesting that when presented with wideband speech targets in conjunction with speechreading, the listener/observer extracts the cross-modality coherence between variations in visible facial kinematics and the acoustic amplitude envelope derived mostly from mid-frequency spectral channels.[1]

Grant and Seitz (2000) suggested BCMP may be at least partially explained by the degree of correlation between the temporal envelope of the target sentence and the kinematic variation of the area of mouth opening during the production of the target sentence. Correlations between amplitude envelope and area functions for these target sentences known to produce a significant BCMP were previously observed to be only about 0.5. However, Grant and Seitz also noted that the detection of speech only requires a very brief moment of the target signal to be audible, and that correlations computed over shorter time frames approximating that of a syllable may be more relevant to explanations of BCMP. Figure 2 shows these local correlations (thick line) along with the speech amplitude envelope (thin line) for the four target signals. The top two panels show the results for sentence 2 whereas the bottom two panels show results for sentence 3. On the left side of the figure are *F*1-filtered targets and on the right are *F*2-filtered targets. According to Grant and Seitz (2000), the relevant information in each panel necessary to explain the presence of significant BCMP is the temporal location of the greatest amplitude envelope peaks comprising the top 2% of the amplitude range (indicated by arrows) along with the local correlation at these moments in time. Note that for both sentences, the correlations are higher in the vicinity of amplitude envelope peaks for the *F*2-filtered targets than for *F*1-filtered targets, and that for sentence 3 in particular, all *F*2-filtered peak-amplitude locations are associated with relatively high correlations. Focusing on the six peak locations identified for the *F*1-filtered targets, the average correlation between lip-area function and acoustic amplitude envelope was 0.34 (s.d. = 0.32), whereas the average correlation between lip-area function and acoustic amplitude envelope correlation for the five peak locations identified for the *F*2-filtered targets was 0.82 (s.d. = 0.11).

IV. DISCUSSION AND CONCLUSIONS

The data presented in this study support previous findings that speechreading provides useful information that can be used by cue listeners to the temporal and spectral locations of high energy regions in speech signals. Essentially, watching the variations in the movement of the mouth during speech production informs listeners both *when* in time and *where* in the spectrum to expect signal energy. By focusing

the attention of the listener to specific spectro-temporal locations in the speech waveform, the ability to hear speech in noisy backgrounds is improved.

Correlation data (Grant and Seitz, 2000) comparing the acoustic envelope from various spectral regions of speech to variations in the area of lip opening during speech production show that amplitude envelopes from the *F*2 speech region appear to have greater coherence with visible oral kinematics than envelopes derived from other spectral regions (e.g., *F*1, *F*3, and wideband). It is hypothesized that this coherence is due primarily to the association between changes in the place of constriction in the front cavity of the vocal tract and rapid changes in *F*2 formant frequency (Stevens, 1998). Furthermore, the fact that there is greater modulation in *F*2 frequency than in *F*1 frequency with changes in place of articulation suggests that the acoustic energy observed at the output of a filter with a fixed bandwidth centered in the *F*2 speech region (as used in the present study) might also be highly modulated. Thus, there appears to be a natural linkage between front cavity area and lip shape and the acoustic fluctuations in energy associated with frequencies in the *F*2 region. Moreover, this information appears to be at least partially available through speechreading and can be used to improve speech detection thresholds in noise.[2]

For automatic speech recognition (ASR) and noise-reduction algorithms for speech enhancement, these findings offer new and potentially interesting possibilities. For example, it may be possible to construct a temporal filter based on variations in inter-lip distance or area of mouth opening that estimates the amplitude envelope of mid-frequency speech bands. Such a time-varying filter could then be used to process the acoustic environment so as to enhance probable speech signals and reduce extraneous acoustic signals that are unrelated to the visual speech dynamics. Recent experiments in ASR have demonstrated that combining mouth shape with acoustic data can improve recognition performance with noisy speech (Girin *et al.*, 2001, in press).

The integration of visual and acoustic speech information allows for robust and reliable speech recognition that is greatly resistant to noise and reverberation. Bisensory integration appears to proceed rather automatically and with a fairly high degree of efficiency (Braida, 1991; Grant and Seitz, 1998; Massaro and Cohen, 2000), especially for normal-hearing subjects. The question of just where in the speech perception process auditory–visual integration takes place has been a topic of much discussion. The present data, as with the previous study by Grant and Seitz (2000), demonstrate that physical correspondence between visible speech kinematics and acoustic modulations of speech output provides an opportunity for auditory and visual speech data to merge at the level of signal detection. Whether this represents an extremely early phase of auditory–visual integration is unclear, since it is possible that improvements in masked thresholds, as evidenced by significant BCMP, may occur for more central reasons (Watson and Kelly, 1981). However, it does bring into question whether it is possible to model auditory–visual speech processing strictly in terms of independent auditory and visual speech processes.

ACKNOWLEDGMENTS

This work was supported by Grant Number DC00792 from the NIDCD and the Department of Clinical Investigation, Walter Reed Army Medical Center under Work Unit #2590-99. I wish to thank Sid Bacon, Andrew Faulkner, and an anonymous reviewer for their helpful comments. All subjects participating in this research provided written informed consent prior to beginning the study. The opinions or assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the Department of the Army or the Department of Defense.

[1]One caveat in interpreting these results is that the *F*1 and *F*2 filters had different bandwidths. Thus, there were significant differences in spectrum level across the two filter conditions. However, these differences in spectrum level probably have little or no impact on the current results. First, the peak amplitude levels for *F*1 and *F*2 target sentences were quite similar (within 2 dB) after the filter outputs were scaled to have the same overall rms level. Second, the primary comparison across filter conditions is made using BCMP measures, which are themselves direct comparisons between auditory and auditory–visual threshold values within a given sentence and a given filter condition.

[2]It has been argued here and elsewhere (Grant and Seitz, 2000) that BCMP is attributed primarily to the correlation between local peaks in the speech amplitude envelope and visible lip shape information. Another possibility is that there is an association between speechreading and rapid changes in *F*2 frequency often observed at boundaries between consonants and vowels. These sudden changes in *F*2 frequency, as opposed to local amplitude peaks, may be the cue underlying the observed BCMP. However, it is important to remember that at detection threshold, so little of the speech signal is available due to the very poor signal-to-noise ratio, that it seems more plausible to consider amplitude peaks, rather than *F*2 frequency, as the primary cue.

American National Standards Institute (**1989**). ''Specifications for Audiometers,'' ANSI S3.6-1989 (American National Standards Institute, New York).

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

Girin, L., Schwartz, J. L., and Feng, G. (**2001**). ''Audio–visual enhancement of speech in noise,'' J. Acoust. Soc. Am. (in press).

Gordon, P. C. (**1997a**). ''Coherence masking protection in brief noise complexes: Effects of temporal patterns,'' J. Acoust. Soc. Am. **102**, 2276–2283.

Gordon, P. C. (**1997b**). ''Coherence masking protection in speech sounds: The role of formant synchrony,'' Percept. Psychophys. **59**, 232–242.

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

Grant, K. W., and Seitz, P. F. (**2000**). ''The use of visible speech cues for improving auditory detection of spoken sentences,'' J. Acoust. Soc. Am. **108**, 1197–1208.

IEEE (**1969**). ''IEEE recommended practice for speech quality measurements,'' Institute of Electrical and Electronic Engineers, New York.

Levitt, H. (**1971**). ''Transformed up–down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.

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

Stevens, K. N. (**1998**). *Acoustic Phonetics* (MIT Press, Cambridge, MA).

Sumby, W. H., and Pollack, I. (**1954**). ''Visual contribution to speech intelligibility in noise,'' J. Acoust. Soc. Am. **26**, 212–215.

Watson, C. S., and Kelly, W. J. (**1981**). ''The role of stimulus uncertainty in the discrimination of auditory patterns,'' in *Auditory and Visual Pattern Recognition*, edited by D. J. Getty and J. H. Howard, Jr. (Lawrence Erlbaum Associates, Hillsdale, NJ), pp. 37–59.

Evaluation of speech intelligibility with the coordinate response measure

Douglas S. Brungart[a)]
Air Force Research Laboratory, Human Effectiveness Directorate, Wright-Patterson AFB, Ohio 45433-7901

(Received 20 April 2000; revised 28 June 2000; accepted 23 January 2001)

The sentences in the coordinate response measure (CRM) corpus described in a recent letter to this journal [Bolia *et al.*, J. Acoust. Soc. Am. **107**, 1065–1066 (2000)] have been used to measure speech intelligibility as a function of signal-to-noise ratio with a speech-spectrum-shaped noise masker. The data from this experiment, along with those of an earlier experiment comparing intelligibility with the CRM and the well-known modified rhyme test (MRT), have also been used to estimate performance with the CRM as a function of the articulation index (AI) for a variety of different masking signals. The results provide a normative evaluation of the CRM for potential users of the CRM corpus, and can be used to compare the CRM with other measures of speech intelligibility. [DOI: 10.1121/1.1357812]

PACS numbers: 43.71.Gv, 43.72.Kb [DOS]

I. INTRODUCTION

In a recent letter to this journal, Bolia *et al.* (2000) described a publicly available speech corpus for speech intelligibility experiments based on the coordinate response measure (CRM) originally developed by Moore (1981). This speech corpus consists of sentences of the form ''Ready (call sign) go to (color) (number) now'' spoken by eight talkers with each of eight call signs, four colors (''blue,'' ''red,'' ''green,'' and ''white''), and eight numbers (1–8). Speech intelligibility can be assessed with the corpus from the percentage of correct number and color identifications for the utterances addressed to the listener's assigned call sign. The CRM was originally developed by researchers at the Air Force Research Laboratory (Moore, 1981) to provide an intelligibility test with greater relevance to military communications than the commonly used modified rhyme test (MRT) (House *et al.*, 1965). Moore measured speech intelligibility with both the CRM and the MRT in a variety of jamming conditions and found that performance with the CRM was less sensitive to interfering noise than the MRT, but that overall performance was highly correlated across the two intelligibility tests. However, little additional information about speech intelligibility with the CRM has been reported since this original publication. This short article presents a normative evaluation of speech intelligibility with the CRM speech corpus as a function of signal-to-noise ratio with a speech-spectrum-shaped masking noise. The results of this evaluation, along with the original data provided by Moore, are used to relate performance with the CRM to the well-known articulation index (Kryter, 1962). This provides a basis for comparing the CRM with other measures of speech intelligibility.

[a)]Electronic mail: douglas.brungart@wpafb.af.mil

II. CRM INTELLIGIBILITY EXPERIMENT

A. Procedure

Nine normal-hearing paid volunteers with previous experience in the CRM task (four males, five females) participated in the experiment. The listeners sat at a control computer in a sound-deadened room and heard diotic headphone presentations of the sentences from the CRM corpus. Each utterance was masked by a noise signal that was spectrally shaped (with a 512-point FIR filter) to match the average overall spectrum of the 2048 sentences in the CRM corpus[1] and rectangularly gated to the same length as the speech signal. The overall levels of the masking noises varied randomly from approximately 64 to 70 dB in 1-dB steps,[2] and the levels of the speech signals were scaled to produce signal-to-noise ratios (measured from the rms power of the speech and noise) ranging from −18 to 15 dB in 3-dB steps. Only sentences with the call sign ''Baron'' were used in the experiment, and the sentence presentations were balanced so that each listener heard exactly 120 sentences from each of the eight talkers in the corpus in random order. All other parameters of the sentence presentation, including color, number, and signal-to-noise ratio, were randomly selected (with replacement) on each trial. After each sentence presentation, the listeners identified the spoken color and number combination by moving the mouse pointer to the appropriate colored digit on the screen of the control computer. The trials were divided into 8 blocks of 120 trials, with 1–2 blocks of trials collected during each day of the experiment.

B. Results

The results show that performance in the CRM is related to signal-to-noise ratio by an S-shaped curve typical of most measures of speech intelligibility in noise (Fig. 1). At signal-to-noise ratios (SNRs) of 0 dB or above, performance plateaus at about 98%-correct responses. At SNRs less than 0 dB, performance degrades rapidly, falling by about 10%/dB as the signal-to-masker ratio declines from −6 to −12 dB. At lower SNRs, performance again levels off as color iden-

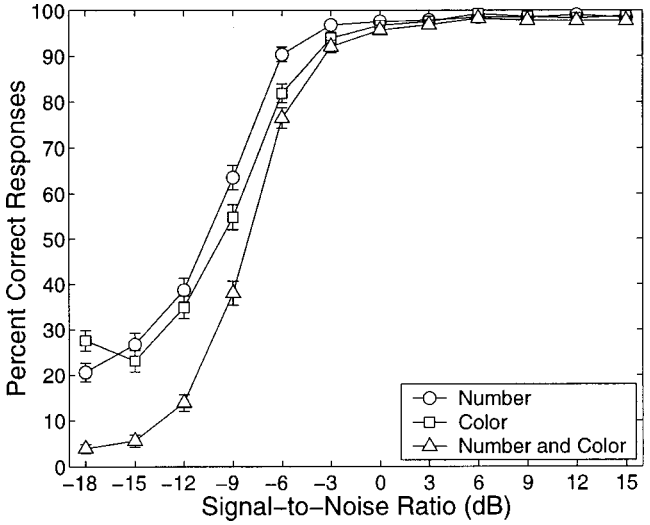


FIG. 1. Percentage of correct identifications of color, number, and both color and number as a function of signal-to-noise ratio with a speech-spectrum-shaped masker. The error bars have been calculated from the arcsine transformation and represent ±1.4 standard errors. Mean values with nonoverlapping error bars would be significantly different at the $p<0.05$ level in a one-tailed t test.

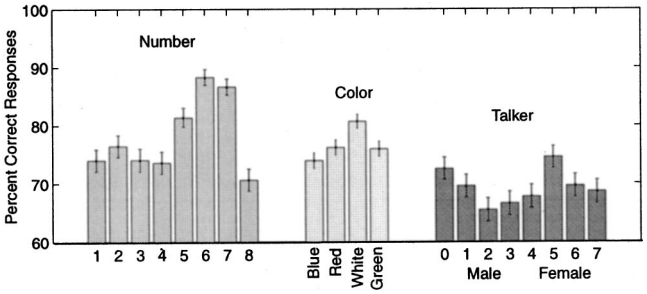


FIG. 2. Effects of number, color, and talker. The first set of data shows the percentage of correct number identifications as a function of the number in the target phrase. The second set shows the percentage of correct color identifications as a function of the color in the phrase. The third set shows the percentage of correct overall responses as a function of the talker (numbered 0–7 as in the CRM corpus). The error bars have been calculated from the arcsine transformation and represent ±1.4 standard errors. Mean values with nonoverlapping error bars would be significantly different at the $p<0.05$ level in a one-tailed t test.

tifications reach chance performance (25%). There are two interesting features of the data. The first is that correct identifications of the color and number in each utterance were essentially independent events. The probability of correct overall (number and color) identifications differs from the product of the probabilities of correct number and color identifications by an average of only 1.1%. This implies that coarticulation played a relatively minor role in the intelligibility of the color–number pair.

The second interesting feature is that the percentage of correct number identifications was consistently higher than the percentage of correct color identifications, in spite of the larger vocabulary in the number category (eight numbers versus four colors). This somewhat counterintuitive result, which indicates a lack of phonetic balance in the four color coordinates used in the response set, can be explained in part by differences in intelligibility across the key words in the CRM vocabulary. The digits ''six'' and ''seven'' were correctly identified more frequently than the other color and number coordinates (Fig. 2). This result is probably related to the high-frequency content of the strident fricative /s/, which stands out against the primarily low-frequency (<5 kHz) speech-shaped masking noise. The second syllable in the word ''seven'' may also have contributed to its enhanced intelligibility. The distinctive diphthong /ai/ may account for the relatively large number of correct identifications of the key words ''five'' and ''white.'' The large differences in intelligibility across the CRM key words suggest that the phonetic features of the color and number words played a larger role in determining overall performance than the number of alternatives in the color and number sets.

Although there were some variations in overall performance across the talkers and listeners used in the experiment, these differences would have relatively little impact on the measurement of speech reception thresholds with the CRM. The threshold SNR values for 50%-correct identification of both the color and number coordinates fell between −8.6 and −7.3 dB for seven of the eight talkers and seven of the nine listeners used in the experiment. The overall percentages of correct responses varied by about 10% across the different talkers used in the corpus (Fig. 2), and by about 10% across the different listeners used in the experiment.

III. RELATIONSHIP BETWEEN THE CRM AND THE ARTICULATION INDEX

The results in Fig. 1 show how the percentages of correct responses in the CRM vary with signal-to-noise ratio for a speech-shaped noise masker. In order to compare the results to other measures of speech intelligibility, it is helpful to relate these results to the articulation index (AI). The AI, which estimates the intelligibility of speech from the spectral properties of the speech and the masking noise, has been shown to accurately predict performance in a variety of phonetically balanced intelligibility tests across a wide range of different listening environments (Kryter, 1969). In Fig. 3, the overall identification results of the CRM test shown in Fig. 1 have been plotted as a function of the AI for each of the 12 SNRs used in the experiment. The articulation indices were calculated using the 20-band method described by Kryter (1962), with the assumption that the overall rms noise level was 70 dB across all of the trials.

The performance data plotted in Fig. 3 provide a guideline for estimating performance with the CRM as a function of the AI, but they should be interpreted with some caution. The CRM has a very restricted vocabulary, and its color and number response words are not phonetically balanced. It is therefore problematic to assume that CRM performance with a speech-shaped noise masker will be the same as performance with all other types of masking signals that produce the same AI value. The original CRM data collected by Moore (1981) provide some insight into the performance of the CRM with different types of masking sounds. Moore collected data using both the CRM and MRT methods with live talkers and a panel of ten listeners wearing full military flight gear, including oxygen masks and flight helmets. He tested five different jammers (an FM tone, an FM drifting

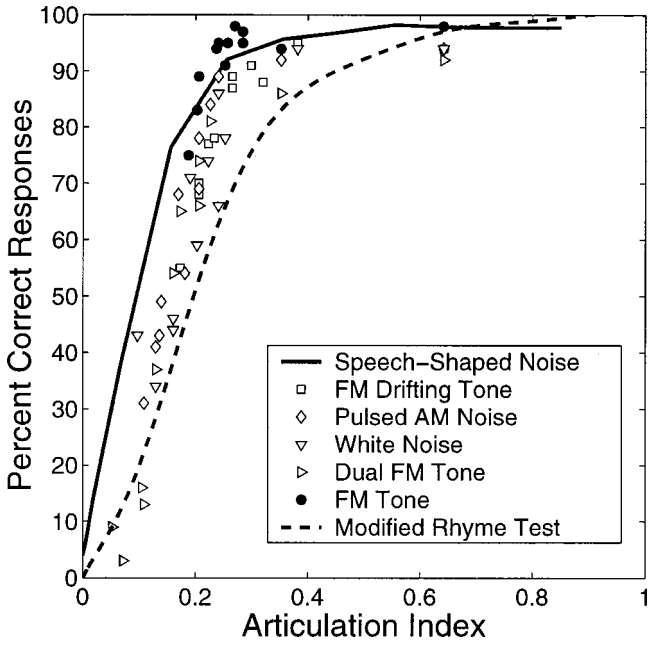


FIG. 3. Performance in the CRM as a function of the articulation index. The solid line represents the percentage of correct color and number identifications in the presence of a speech-shaped noise masker as a function of the articulation index, which was calculated by the 20-band method (Kryter, 1969). Each symbol in the plot represents a single jammer type, jammer-to-noise ratio, and ambient noise level from Moore's 1981 study comparing the CRM to the MRT. The performance curve for the MRT as a function of AI that was used to transform Moore's data is also shown (Kryter, 1969). See the text for details.

tone, a white-noise modulated AM tone, a pulsed AM noise, and dual FM swept tones) at six different jammer-to-signal ratios both in quiet and in the presence of a 105-dB jet-engine-shaped ambient noise. Although Moore did not calculate the AI directly, it is possible to derive a rough estimate of the AI in each of these 60 listening conditions by comparing the MRT performance measured by Moore to the MRT versus AI curve provided by Kryter (dashed line in Fig. 3). These estimates can then be used to plot CRM performance as a function of AI in each listening condition (symbols in Fig. 3). The overall shape of the curve defined by these data points is similar to the response curve measured with the speech-shaped noise masker, but has a steeper slope and is shifted to the right of the speech-shaped noise curve. This shift reflects the additional difficulty introduced into Moore's experiment by requiring the listener to correctly identify one of three call signs (''Ringo,'' ''Laker,'' or ''Baron'') in each stimulus presentation and respond only to those phrases containing the call sign assigned to that listener prior to the experiment. The data from four of Moore's jamming conditions (open symbols) are clustered into a reasonably well-defined response curve, which is remarkable when one considers that these CRM performance estimates are sensitive to errors in both the determination of MRT performance and in the determination of CRM performance in each listening condition. However, the data from the FM jammer (filled symbols) consistently indicate better performance than the other four types of jamming signals (open symbols) under listening conditions with similar AI values. Although Moore does not provide a detailed description of the FM jammer, it is clearly an example of a masking signal that interferes substantially less with the phonetically unbalanced CRM vocabulary than with the phonetically balanced word list used in the MRT. Based on these results, it appears that the relationship between CRM performance and AI shown in Fig. 3 is reasonably stable for many types of masking sounds, but that performance can deviate substantially from this curve for certain specific types of masking signals. It is therefore not appropriate to use the CRM to determine the AI of speech in the presence of an arbitrary masking signal.

It is, nevertheless, useful to use the AI performance curves as a basis for comparing the CRM to other types of speech-perception tests. For example, the AI performance curve for the MRT shown in Fig. 3 allows a direct comparison between the CRM and MRT. For all the different types of masking signals tested, it is clear that the CRM is substantially more sensitive to changes in AI than the MRT at AI values less than or equal to 0.25, and substantially less sensitive than the MRT at AI values greater than 0.25.

IV. APPLICATIONS OF THE CRM

While the limited vocabulary of the CRM prevents its use as a replacement for comprehensive, phonetically balanced intelligibility measures such as the MRT, the CRM does have advantages over other speech-intelligibility tests that justify its use in some special testing situations. One advantage demonstrated by these results is the sensitivity of the CRM to small intelligibility changes in extremely difficult listening environments (characterized by AI values less than 0.25). This feature makes the CRM attractive for testing intelligibility in very noisy environments. It also makes the CRM an excellent threshold test for determining how powerful a jamming signal must be to render a communications channel inoperative. Another advantage of the CRM is its intrinsic portability across different languages. Since all languages contain words for colors and numbers, the CRM can provide a rough functional measure of intelligibility without deriving a phonetically balanced word list for each language tested. A third advantage of the CRM corpus is its applicability to speech-intelligibility testing with multiple simultaneous talkers. The call signs in the corpus allow the experimenter to designate the target phrase without relying on differences in its location, onset time, or talker characteristics to distinguish it from the simultaneous masking phrases. The listeners are simply instructed to respond with the color and number coordinates spoken in the phrase addressed to their assigned call signs. A final advantage of the CRM is the relative ease of setting up and running speech-intelligibility experiments with the publicly available CRM corpus (Bolia *et al.*, 2000). Because the listener is always limited to the same 32 possible responses, every trial of the CRM can be evaluated with exactly the same response list. This makes it much easier to collect and process intelligibility data with the CRM than with sentence-based tests or phonetically balanced tests such as the MRT. Thus, although the CRM should not be viewed as a comprehensive measure of speech intelligibility, it does appear to be representative of other

speech perception tests and it should be considered when a rapid measure of functional intelligibility is desired.

ACKNOWLEDGMENTS

The author would like to thank Mark Ericson, Todd Nelson, Bob Bolia, Brian Simpson, Kim Scott, Tom Moore, and Howard Rothman for their helpful comments in the preparation of this manuscript.

[1]Note that the speech materials in the CRM corpus have been low-pass filtered with an 8-kHz cutoff frequency.
[2]The listeners were allowed to choose a comfortable listening level for the stimuli in the experiment.

Bolia, R., Nelson, W., Ericson, M., and Simpson, B. (**2000**). ''A speech corpus for multitalker communications research,'' J. Acoust. Soc. Am. **107**, 1065–1066.

House, A., Williams, C., Hecker, M., and Kryter, K. (**1965**). ''Articulation testing methods: Consonantal differentiation with a closed response set,'' J. Acoust. Soc. Am. **37**, 158–166.

Kryter, K. (**1962**). ''Methods for calculation and use of the articulation index,'' J. Acoust. Soc. Am. **34**, 1689–1697.

Kryter, K. C. (**1969**). ANSI S3.5-1969, ''American National Standards Methods for Calculation of the Articulation Index'' (American National Standards Institute, New York).

Moore, T. (**1981**). ''Voice communication jamming research,'' in AGARD Conference Proceedings 331: Aural Communication in Aviation, pp. 2:1–2:6, Neuilly-Sur-Seine, France.

PROGRAM OF

The 141st Meeting of the Acoustical Society of America

Palmer House Hilton Hotel • Chicago, Illinois • 4–8 June 2001

NOTE: All Journal articles and Letters to the Editor are peer reviewed before publication. Program abstracts, however, are not reviewed before publication, since we are prohibited by time and schedule.

MONDAY MORNING, 4 JUNE 2001 CRYSTAL ROOM, 8:30 TO 11:30 A.M.

Session 1aAA

Architectural Acoustics: Concert Hall Measurements

J. Christopher Jaffe, Chair
Jaffe Holden Acoustics, 114A Washington Street, Norwalk, Connecticut 06854

Chair's Introduction—8:30

Invited Papers

8:35

1aAA1. Measurement of decay times in coupled spaces. Ning Xiang (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, University, MS 38677), Paul M. Goggans, and Donghua Li (Univ. of Mississippi, University, MS 38677)

To achieve enough variability for different performances, a number of new concert halls with secondary hard chambers coupled to the main floor have recently received positive evaluations. As a result, the acoustics of coupled spaces is drawing growing attention in architectural acoustics. In these coupled spaces, the sound energy decays with different exponential constants under certain conditions. For better understanding and control of the acoustics in coupled spaces an efficient measurement technique of identifying multiple decay times is therefore of critical importance. A Bayesian parameter estimation approach is applied to an algorithmic evaluation of decay times using measured Schroeders decay functions. This work discusses Bayesian analysis of systematic measurements in scaled-down, coupled rooms and real coupled spaces.

8:55

1aAA2. Stage acoustics as affected by an array of apertures incorporated into the orchestra shell of a stagehouse coupled auditorium. Jason E. Summers and Yasushi Shimizu (Sonics in Architecture, Rensselaer Polytechnic Inst., 110 8th St., Troy, NY 12180, summej2@rpi.edu)

A number of performance halls incorporate into their design a volume (or ''chamber'') surrounding the stage platform and coupled to the stage platform through the orchestra shell. Bass Performance Hall utilizes this design concept, while incorporating into the orchestra shell an array of apertures backed by adjustable doors. Measurements have been made to assess the impact of the adjustability of these doors on the soundfield of the stage platform. Consideration was given to both the early and late portions of the soundfield. Variations in impulse response derived measures, in the spectra of early reflections, and in the coupling behavior are presented. These are examined in view of existing models for the reflection and transmission properties of apertures and for the decay of coupled spaces. [Research supported by the Bass Foundation.]

9:15

1aAA3. Concert hall acoustics at Bass Hall. Derrick Knight, Young-Min Kwon, John Kreiger, Brandon Holcomb, Jamin Hemenway, Jason Summers, Yasushi Shimizu, and Christopher Jaffe (Sonics in Architecture, Rensselaer Polytechnic Inst., 110 8th St., Troy, NY 12180, knight427@yahoo.com)

Traditionally, multi-purpose halls are believed to have less than adequate symphonic acoustic listening characteristics compared to one room concert halls such as Boston Symphony Hall. Over the last two years, students and faculty of the Sonics Program at the Rensselaer Polytechnic Institute conducted a series of measurements at the multi-purpose Bass Performance Hall in Forth Worth, Texas. In this investigation, Bass Hall is shown to have excellent concert hall metrics when configured for that purpose. Measurements

of reverberation time, early decay time, clarity (C80), and center time (Ts) were taken according to the ISO 3382 standard. This data was compared to data from the great one-room concert halls. This comparison demonstrates that Bass Hall can function as an equally excellent concert hall, in spite of being a multi-purpose hall. [Research supported by the Bass Foundation.]

Contributed Papers

9:35

1aAA4. Consideration of the propagation loss of the direct sound with change in seat conditions. Yasushi Shimizu (Sonics in Architecture, Rensselaer Polytechnic Inst., 110 8th St., Troy, NY 12180, shimiy@rpi.edu)

Reviewing excess attenuation of the direct sound by seats in concert hall acoustics gives the propagation loss by the performers absorption on the stage platform and diffraction of the front seats as well as seat dip effect in low frequency spectrum. Scale model experiments with 1/10 scale model in anechoic chamber were conducted at YAMAHA Acoustic Research Laboratories, Hamamatsu, focusing on propagation loss by (1) diffraction due to various geometrical conditions of the seating row in front of the listening position and (2) attenuation by the seat rows in front of the listening point. The results show the excess attenuation in middle and high frequency spectra. The optimum height between two adjacent seating rows is more than 1.3 feet and the number of seat rows are less than five seats. Normalized SPL referenced to SPL measured at 6.6 feet over the listening point is presented to be effective for evaluating excess attenuation. Concert Hall at Niigata Performing Arts Center was designed with consideration of these scale model experiments. Measurement data of normalized SPL and power spectra of orchestra sound in this hall are presented.

9:50

1aAA5. Measurement surveys of eight British concert halls. Russell Richardson and Bridget M. Shield (School of Eng. Systems and Design, South Bank Univ., London SE1 0AA, UK, shieldbm@sbu.ac.uk)

Objective measurement surveys and subjective surveys have been carried out in eight leading concert venues in the United Kingdom. In each hall a questionnaire survey of the whole audience at a concert of symphonic music was used to elicit the opinions of nonexpert listeners. Altogether over 3000 responses have been received. An objective survey of each (empty) hall has also been carried out, the parameters measured including reverberation time, early decay time, clarity index and definition. The number of measurement positions in each hall varied from 36 to 92. The questionnaire response data is to be correlated with the results of the objective measurements to investigate the relationship between objective data and the subjective impressions of nonexpert listeners. This paper will present summary data of the measurements in the eight halls and examine the variation in the values of the different parameters across each hall. The differences between the halls will also be considered.

10:05–10:15 Break

10:15

1aAA6. A five-channel microphone system for detecting 3-D acoustic image sources. Chulmin Choi, Lai-Hoon Kim (Seoul Natl. Univ., Seoul 151-742, South Korea), Se-Jin Doo (Dong-Ah Broadcasting College, Ansung, Kyonggi, South Korea), Yang-Ki Oh (Mokpo Natl. Univ., Mokpo, Chonnam 534-729, South Korea), Dae-Up Jeong (Chonbuk Natl. Univ., Jeonju, Chonbuk 561-756, South Korea), and Koeng-Mo Sung (Seoul Natl. Univ., Seoul 151-742, South Korea)

Measuring three-dimensional behavior of early reflections in a sound field has been an important issue in auditorium acoustics, since they are found to be strongly correlated with the subjective experience of spatial impression in rooms. A conventional way of detecting the direction and relative amplitude of reflections is to utilize a four-channel microphone system. This system, however, has the inherent possibility of missing some important reflections while exploring the early part of reflections. A new measurement system, which has five microphones on the apex of a tetrahedron and at the center of gravity, is proposed in the present work. The proposed system enables us to find more early reflections accurately, which previous four-channel microphone systems might miss. A peak detecting pair matching algorithm, which gives dominant peaks of impulse response automatically by typical iteration, was adopted, since we should find exact peak positions of measured impulse response in order to calculate source positions accurately. In the present paper, the theoretical backgrounds and features of the five-channel microphone system are presented. Also, some results from experiments using the system are discussed. Experimental results clearly show the advantage of the five-channel microphone system over the conventional four-channel microphone system. [Work supported by Korean Research Foundation Grant KRF-1999-1-310-004-3.]

10:30

1aAA7. Interaural cross-correlation: Sources of variability in concert halls. Gary S. Madaras (The Talaske Group, Inc., 105 N. Oak Park Ave., Oak Park, IL 60301, gary@talaske.com)

Impulse responses were measured inside multiple concert halls. Data analysis identified the level of the direct sound relative to the level of the architectural reflections as a significant source of IACC80 variability. Excluding the direct sound from the IACC80 integral durations (IACC80 $\times$DIR) significantly decreased the variability. The remaining variability seemed to relate to the arrival direction of the architectural reflections. A testing model capable of various configurations and a scale measurement method were developed. Models that had reflections approaching the receiver from the sides produced much lower IACC80$\times$DIR values than those without reflections approaching the receiver from the sides. The various configurations of architectural elements (i.e., balconies and stage canopy) and surface treatments (i.e., diffusive and specular) affected IACC80$\times$DIR values to a lesser extent than altering the arrival direction of the architectural reflections. It was concluded that the general direction from which the architectural reflections arrive and the finish of the architectural surfaces are both sources of IACC80$\times$DIR variability (the former having a much greater effect). It was also concluded that unless the direct sound is excluded from the integral duration, variability of interaural cross-correlation values does not relate as strongly to the arrival direction of the architectural reflections. [Work supported by the Concert Hall Research Group and the University of Florida.]

10:45

1aAA8. Investigation of early-to-late sound reflection effect in a room with a dome. Mokhtar Harun, Tharek Abdul Rahman, Md. Najib Ibrahim, and Ahmad Khan Said (Universiti Teknologi Malaysia, 81310 Skudai Johor, Malaysia, mokhtar@suria.fke.utm.my)

The material of a dome is normally very sound reflective. This is due to the fact that the material for inner surfaces of the dome is made of hard material for endurance and easy maintenance. Also, the dome is often situated high above the floor of the room with the intention of providing larger volume and thus creating comfortable envelopment of the dome for people inside it. Due to the height from the floor and the hard material of its inner surfaces, the dome naturally manifests strong sound reflection, sound focusing effects and uneven sound pressure levels. These phenomena also give rise to the reverberation time in a room with a dome. It is the intention of this paper to discuss the effects of early-to-late sound reflection with respect to reverberation time and sound reverberant level in a room with a dome.

11:00

1aAA9. Perceived spatial impression from various musical motifs. Lily M. Wang (Arch. Eng. Prog., Univ. of Nebraska–Lincoln, 200B PKI, Omaha, NE 68182-0681, lwang@unl.edu)

Most of the objective measures which are currently used to quantify spatial impression in auditoria, such as lateral energy fraction (LF), interaural cross-correlation coefficient (IACC) and late lateral relative sound level (GLL), are calculated from an impulse response. For subjective testing, these impulse responses are often later convolved with music recorded in an anechoic environment to simulate music performance in the hall. One spatial impression measure, though, interaural level fluctuations (IALFs) which was first proposed by Griesinger [Aud. Eng. Soc., Preprint No. 3292 (1992)] and further developed by L. M. Wang and A. C. Gade [J. Acoust. Soc. Am. **107**, 2891(A) (2000)], is calculated from the musical source signal instead of from the impulse response. Preliminary investigations showed that the IALF, calculated from simulated sound fields generated by ODEON, changed relatively the same for different pieces of music across various halls. If IALF is a valid measure for spatial impression, then this indicates that the musical source is not critical in evaluating the difference in spatial impression between halls. In this paper, results from subjective listening tests are presented to determine if the same trend is perceived by listeners across different musical motifs, which would aid in validating IALF.

11:15

1aAA10. An experiment in the use of resilient materials in orchestra riser design. Ian B. Hoffman, Richard H. Talaske, and Byron W. Harrison (The Talaske Group, Inc., 105 N. Oak Park Ave., Oak Park, IL 60301, ian@talaske.com)

An experiment was assembled under extremely limited time and space constraints to investigate the acoustic effect of resilient materials in orchestra riser design. The riser was a portable, box-type riser for assembly on a flat concert platform, not a lift-type riser. The primary goal of the inquiry was to maximize lateral vibration transfer through the riser (from celli, basses, etc.) without detrimentally affecting airborne sound radiation from the riser. A portion of a full-scale riser was mocked up, including sections of differing elevation. Various design configurations of the riser allowed experimentation with and without resilient materials, and with resilient materials placed at different positions within the riser structure. The experiment was conducted within a reverberant volume. The riser was excited by mechanical shakers and by a cello, independently. Vibration measurements were captured using accelerometers on the riser, and airborne sound pressure levels were captured by a distant microphone. Initial results suggest that the use of resilient materials in certain configurations can considerably increase lateral vibration transfer through the riser (as compared to rigid construction), without significant compromise to airborne radiation from the riser.

MONDAY MORNING, 4 JUNE 2001 ADAMS ROOM, 8:30 TO 11:55 A.M.

Session 1aAO

Acoustical Oceanography: Acoustical Instrumentation for Water Column Measurements I

Kenneth G. Foote, Chair
Department of Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

Chair's Introduction—8:30

Invited Papers

8:35

1aAO1. Multibeam sonars: Applications for fisheries research. Larry A. Mayer, Yanchao Li (Ctr. for Coastal and Ocean Mapping, Univ. of New Hampshire, Durham, NH 03824, lmayer@unh.edu), and Gary Melvin (Dept. of Fisheries and Oceans, St. Andrews, NB, Canada)

Multibeam sonars are rapidly becoming a standard tool for seafloor mapping in support of geological, geophysical, and engineering applications. More recently, the ability of multibeam sonars to provide high-resolution, large areal coverage, and potentially quantitative, coregistered, backscatter has been applied very successfully to problems of defining fisheries habitat. While most multibeam sonars are designed to gate out all midwater returns, newly developed systems now allow access to the full data stream and thus offer the possibility of application to studies of pelagic and demersal fisheries. Traditional acoustic approaches to fisheries issues have used single beam echo sounders that sample a relatively small volume of the water column within a survey area. Multibeam sonars provide a mechanism to greatly enhance both the resolution and the area of coverage. When combined with powerful new 3-D visualization techniques, they can offer immediate feedback on fish behavior as well as the critical question of vessel avoidance. If properly calibrated, multibeam sonars can provide the means for much more robust assessment of stock levels and perhaps even species identification.

9:35

1aAO2. Physical constraints in sonar design. Philip R. Atkins and Tim Collins (School of Electron., Univ. of Birmingham, Edgbaston, Birmingham B15 2TT, UK, p.r.atkins@bham.ac.uk)

The dynamic range and processing gains achievable within a simple passive sonar will be reviewed. It will be demonstrated that the constraints of modern electronic hardware components do not necessarily limit the performance of a passive sonar system when compared with the physical constraints of the acoustic channel. By contrast it will be demonstrated that the pulse design of an active sonar places significant limitations on the capability of the system. For example, range resolution, Doppler resolution, range sidelobes, ambiguities and reverberation performance all depend on the design of the sonar transmission. Techniques available to the designer

for suppressing range sidelobes and pulse designs with both range and Doppler sensitivity will be reviewed. The constraints imposed by the transducer will be considered, particularly in relationship to ultrahigh range resolution capabilities. Finally, a comparison between an early 1960s sonar and a modern equivalent will be made in terms of performance, power consumption and size.

10:05–10:25 Break

Contributed Papers

10:25

1aAO3. Acoustic classification of individual zooplankton using artificial neural network. Andone C. Lavery and Timothy K. Stanton (Woods Hole Oceanogr. Inst., Dept. of Appl. Ocean Phys. and Eng., Woods Hole, MA 02543)

The acoustic scattering characteristics of zooplankton generally fall into one of three categories: gas-bearing, fluidlike, or elastic shell. Acoustic scattering models are available for zooplankton that fall into each of these categories. A multilayered feedforward artificial neural network, based on a backpropagation algorithm, has been used to discriminate between the model-based acoustic scattering responses of individual zooplankton from these different zooplankton classes. Realistic animal size distributions, taken from net tows, have been used as input to the models to create (simulated) acoustic data for use as input into the classification network. Two different approaches have been taken to selecting input features from the model-based zooplankton scattering responses, and the results of the neural network classification for these two features sets are compared. The feasibility of applying this neural-network-based classification technique to both multifrequency and broadband target strength data from field surveys is discussed.

10:40

1aAO4. Simple, nonoptimal, real-time classifiers for euphausiid and copepod numerical density estimation. Philip R. Atkins, Claire Bongiovanni, David T. I. Francis (School of Electron., Univ. of Birmingham, Edgbaston, Birmingham B15 2TT, UK), Kenneth G. Foote (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543), Tor Knutsen (Inst. of Marine Res., N-5817 Bergen, Norway), Peter K. Eriksen, Mette Torp Larsen, and Tom Mortensen (RESON A/S, DK-3550 Slangerup, Denmark)

An active sonar covering the frequency range 1.6–3.2 MHz was used to observe zooplankton at two sites along the Norwegian coast during the period 28 April–9 May 1999. The sonar transmitted LFM signals with an approximate bandwidth-time product of 100. The matched-filter envelope was computed within the receiver at a sampling rate of 10 M samples/s. Freshly caught, swimming specimens were insonified *ex situ* in a tank mounted on the stern of R/V JOHAN HJORT, and the envelope signal was classified by eye as belonging to the euphausiid *Meganyctiphanes norvegica*, or copepod *Calanus finmarchicus*. The separated signals were then averaged to produce templates corresponding to the two classes. All orientation information associated with the specimens was thus removed. A sliding-window-normalized covariance was calculated between the templates and the receiver output for a large number of *ex situ* and *in situ* pings, a technique capable of being implemented at the sampling rate of the receiver. Although simple, this multiple-hypothesis-testing procedure outperformed a conventional split-window (CFAR) normalizer and produced a rapid and apparently reliable method of distinguishing the two classes and estimating their numerical density. [Work supported by EU RTD Contract No. MAS3-CT95-0031.]

10:55

1aAO5. Observation of fish length and behavior of an individual fish using echo trace analysis. Kouichi Sawada, Yoichi Miyanohana, Hideyuki Takahashi, and Yoshimi Takao (Natl. Res. Inst. Fisheries Eng., Fisheries Res. Agency, Hasaki, Kashima, Ibaraki 314-0421, Japan, ksawada@nrife.affrc.go.jp)

A compact sized split-beam echo sounding system was developed and used in Sagami Bay, Japan, to validate echo trace analysis (ETA) that can observe the length and the behavior of an individual fish. Fishing by hook and line was conducted to confirm the fish species and its length distribution. The echo sounding system was tethered from an anchored ship. Electric power was provided through a cable. This system consists of a split-beam transducer (70 kHz, 118 mm), a transmitting and receiving unit, a data-recording unit, and a control unit and the total weight of the echo sounding system is about 30 kg. Recorded data are up linked to the computer on the ship through a LAN cable. The fish length and fish swimming speed were estimated using ETA. From the fishing results, 79l (*Scomber japonicus*) or horse mackerel (*Trachurus japonicus*) had an average fork length (FL) of 27.9 cm. The estimated fish length distribution coincides with that obtained from a fishing well. The estimated swimming speed was 0–1.2 m/s and corresponds to 0–4.3 (FL). These results show the validity of ETA.

11:10

1aAO6. Use of multiple-beam echo-sounders for estimation of target strength *in situ*. Ian Hampton (BENEFIT, P.O. Box 912, Swakopmund, Namibia) and Michael A. Soule (Marine and Coastal Management, Dept. of Environ. Affairs and Tourism, Roggebaai, 8012, Cape Town, South Africa)

In principle, the average target strength of fish in a defined volume can be estimated *in situ* from the ratio between the volume backscattering strength and a count of the number of fish in the volume. The chief problems, which have precluded use of the method in practice, are the difficulty in obtaining reliable counts at common fish densities and in defining beam volume. It is suggested that both problems could be solved to a large extent by using a relatively modest high-frequency, multi-beam echo-sounder as a counter. This presentation outlines the underlying principle of the method, considers the design criteria, and describes a 48-element 420 kHz system, (code named ABACUS) which is being built to test the method. The system, which has been designed to operate to approximately 70 m, for target strengths in the range 40 to 55 dB, generates a single transmit beam and 32 receiving beams orthogonal to it, which intersect to give up to ten 1 deg $\times$ 1 deg resolution cells. Preliminary test-tank results are presented, and plans for further experimental work discussed. It is believed that the method should be capable of delivering *in situ* target strength estimates at densities well beyond the limit of current methods.

11:25–11:55

Panel Discussion

Session 1aPA

Physical Acoustics: Acoustic Detection of Land Mines and Other Buried Objects

James M. Sabatier, Chair
National Center for Physical Acoustics, University of Mississippi, Coliseum Drive, University, Mississippi 38677

Contributed Papers

10:00

1aPA1. Acoustic landmine detection at 40° below zero. James M. Sabatier, Ning Xiang, and George Nock (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, MS 38677)

To study the effects of snow cover and frozen ground, mine detection experiments based on acoustic to seismic (A/S) coupling have been conducted on a roadway at the Defense Research Establishment Suffield (DRES), Alberta, Canada, 5–16 February 2001. The site is a gravel road with grain or gravel sizes as large as several centimeters. The ground was frozen. Only a few measurements of acoustic-to-seismic coupling in frozen grounds have been reported. The coupling ratio is within a factor of 10 of that for unfrozen soils. The fast compressional wave speeds were expected to be significantly larger than those in unfrozen soils, while slow compressional wave speeds were expected to be possibly unchanged, except for temperature effects. Therefore, one might expect significantly larger wavelengths for the fast compressional wave and nonlocally reacting coupling. Compensation for these changes included using higher-frequency sound and normal incidence angles for the sound source. This presentation outlines the results of this testing. [Work supported by the U.S. Army Communcations–Electronics Command Night Vision and Electronic Sensors Directorate.]

10:15

1aPA2. Acoustic scattering by a buried object with a compliant top in a rigid porous material. Doru Velea, Roger Waxler, James M. Sabatier, and Craig J. Hickey (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, University, MS 38677)

Recently, a method for the detection of landmines buried a few inches below the surface of the ground has been developed at the University of Mississippi. This method utilizes acoustic-to-seismic coupling and measures the particle velocity of the insonified ground surface on- and off-target with a Laser Doppler Vibrometer. As part of the on-going theoretical effort to explain the measurements, this paper is concerned with acoustical scattering from an object buried in a rigid porous half-space. The object is assumed to be a rigid right circular cylinder with the exception that its top can move in the vertical direction in a pistonlike fashion with a mechanical resonance at around 100 Hz. A boundary element method has been employed to predict the on-target/off-target particle velocity ratio. Comparisons with observations are made. [Work supported by the U.S. Army Communications-Electronics Command Night Vision and Electronic Sensors Directorate.]

10:30

1aPA3. Measurement of attenuation and speed of sound in soils for the purposes of imaging buried objects. Michael Oelze, Robert Darmody, and William O'Brien (Univ. of Illinois at Urbana–Champaign, 405 N. Mathews, Urbana, IL 61801, oelze@brl.uiuc.edu)

The speed and attenuation of sound vary in different soil types and with different levels of water saturation. Imaging techniques make use of speed and attenuation to range targets and compensate for loss in echo magnitude with depth. Signals from an acoustic source over the range of 2–6 kHz were passed through soil and detected by a coupled hydrophone. Six soil samples were used to represent a range of properties expected to influence acoustic response. Clay ranged from 2%–38%, silt from 1%–82%, sand from 2%–97%, and organic matter from 0.1%–11.7%. The attenuation and propagation speed were determined in the soil as a function of soil moisture and compaction. Attenuation coefficients ranged from 0.12–0.96 dB cm^{-1} kHz^{-1}. Correlation coefficients were 0.35 ($P=0.01$) and 0.31 ($P=0.03$) between attenuation and soil water content and soil bulk density. Propagation speeds ranged from 86 to 260 $m\,s^{-1}$. The correlation coefficient with speed was -0.28 ($P=0.05$) for soil water content and 0.42 ($P=0.002$) for porosity. Given the attenuation, it is theoretically possible to image objects in soils to 40 cm in depth. [Work supported by U.S. Army Construction Engineering Research Laboratory, Champaign, IL.]

10:45

1aPA4. Acoustic landmine detection: Poroelastic model and elastic wave models. Yanqing Zeng and Qinghuo Liu (Dept. of Elec. and Computer Eng., Duke Univ., Box 90291, Durham, NC 27708-0291, qhliu@ee.duke.edu)

Acoustic waves can be a viable tool for the detection and identification of landmines and unexploded ordnance (UXO). Design of acoustic instruments and interpretation and processing of acoustic measurements require accurate numerical models for acoustic wave propagation in heterogeneous soil with buried objects. Two numerical models will be compared in this work: the soil is treated as an elastic material in the first model, and as a poroelastic material in the second model. A multidomain pseudospectral time-domain method is used for the elastic model, while a finite-difference method is used to solve Biot equations in the poroelastic model. Acoustic wave response of a buried landmine is investigated with both methods. When the soil is treated as a fluid-saturated poroelastic medium, additional attenuation is automatically introduced via the relative motion between solid frame and pore fluid. Thus the poroelastic model is presumably a more accurate model; however, the elastic model requires less computation time. This work will study several important scenarios such as landmines buried in dry sand, in wet sand, and in mud. A comparison between results of the PSTD model and results of the poroelastic model is performed to characterize the regime of validity for the elastic model.

11:00

1aPA5. A hybrid numerical model to predict the effect of buried objects on the acoustic-to-seismic coupling on ground surfaces. Shahram Taherzadeh[a] (Natl. Ctr. for Phys. Acoust., Coliseum Dr., Univ. of Mississippi, Oxford, MS 38677, s.taherzadeh@open.ac.uk)

A numerical model based on a hybrid boundary integral equation/wave number integration is presented and is used to predict the influence of rigid objects buried within a layered, porous elastic soil upon excitation of the ground surface induced by an airborne sound source. Recently, acoustic-to-seismic coupling phenomena have been used successfully to detect buried landmines [J. Sabatier and N. Xiang, J. Acoust. Soc. Am. **106**, 2143 (1999), **107**, 2896 (2000)]. If a ground surface is subjected to insonification, the sound wave penetrates the ground and is reflected and refracted from the topsoil layers, resulting in complex acoustic-to-seismic spectra. The presence of an object alters these spectra by scattering and/or enhanc-

ing certain vibrational modes. The model will be applied to predict these changes in the presence of buried landmines. [a] Permanent address: Faculty of Technology, The Open Univ., Milton Keynes, UK.

11:15

1aPA6. 3-D multidomain PSTD for elasticity. Yanqing Zeng and Qinghuo Liu (Dept. of Elec. and Computer Eng., Duke Univ., Box 90291, Durham, NC 27708-0291, qhliu@ee.duke.edu)

A 3-D numerical modeling scheme is developed in general curvilinear coordinates by using a multidomain pseudospectral time-domain (PSTD) method. The computational domain is naturally divided into subdomains of piecewise homogeneous, curvilinear quadrilaterals conforming to the problem geometry. These curvilinear quadrilaterals are then mapped onto a unit square, where the spatial derivatives are approximated by the Chebyshev PSTD method. Characteristic conditions and physical boundary conditions are used to patch the solution between adjacent subdomains. A newly developed unsplit perfectly matched layer (PML) is used to absorb outgoing waves at the computational edge. This multidomain PSTD algorithm allows natural incorporation of curved boundaries into the numerical scheme, thus providing more increased flexibility and accuracy than the conventional finite-difference method. The numerical scheme is validated by analytical solutions and by other numerical methods. Applications of this multidomain PSTD method will be shown for acoustic landmine detection.

MONDAY MORNING, 4 JUNE 2001 STATE BALLROOM, 9:00 A.M. TO 12:00 NOON

Session 1aPP

Psychological and Physiological Acoustics: Perception, Learning, and Cognition

Beverly A. Wright, Chair

Audiology and Hearing Science Program, Northwestern University, 2229 North Campus Drive, Evanston, Illinois 60208

Contributed Papers

9:00

1aPP1. The perceptual interaction between the pitch and timbre of musical sound. Jeremy P. Marozeau, Alain de Cheveigné, Stephen McAdams, and Suzanne Winsberg (IRCAM-CNRS, 1 place Igor-Stravinsky, F-75004 Paris, France)

The perceptual dependency of the timbre of musical sounds on their fundamental frequency ($F0$) was examined through two timbral dissimilarity experiments. In Experiment 1, subjects judged timbral dissimilarity between pairs of stimuli produced by 12 musical instruments equal in $F0$, duration, and loudness. There were three sessions, each at a different $F0$ (B2, Db3, Bb3). In Experiment 2, the same stimuli were presented in pairs with a constant difference in $F0$. Two sessions corresponded to two different intervals (B2-Db3 and B2-Bb3). Subjects were instructed to ignore pitch difference. The results are interpreted in terms of an evolution of timbre space as a function of $F0$. Experiment 1 produced dissimilarity matrices that were similar at different $F0$s, suggesting that the relative positions of instruments within the timbre space varied little with $F0$. Experiment 2 produced dissimilarity matrices that were symmetrical, suggesting further that $F0$-dependent displacements within the timbre space are small. Patterns of dissimilarity were similar with or without an $F0$ difference, showing that subjects could ignore a constant difference in pitch when making timbral dissimilarity judgments. Results suggest that the perception of timbre is relatively independent of pitch and can be abstracted across differences in $F0$ under these conditions.

9:15

1aPP2. Infants' perception of transposed melodies. Daniella A. Kim and Lynne A. Werner (Dept. of Speech and Hearing Sci., Univ. of Washington, 1417 NE 42nd St., Seattle, WA 98105)

It has been suggested that infants are similar to adults in their perception of tone chroma, in the perceptual similarity of notes in octave relationships [Demany and Armand (1984)]. In music, note pitches are defined by height (frequency distance) and chroma (key). The current study further examines this question. Seven- to nine-month-old infants' perceptions of transpositions of six-note tonal melodies were measured using an observer-based procedure. Once given a background melody subjects were trained to respond to an entirely different training melody and not to a simple transposition. Once that discrimination had been learned, transposed versions of the background melody were introduced. If infants perceive tone chroma as adults do, they should learn this discrimination more quickly for an octave transposition. In fact, infants learned to ignore a transposition by a fifth more quickly than transpositions by an octave. The implication is that infants found melodies that were transposed by a smaller distance in frequency as more perceptually similar to the background melody than they did for the octave transposition. A possible conclusion is that octave transpositions of whole melodies have special perceptual status for infants of this age range. [Work supported by NIDCD Grant No. DC00396.]

9:30

1aPP3. Auditory streaming without spectral cues. Nicolas Grimault, Sid P. Bacon (Psychoacoustic Lab., Dept. of Speech and Hearing Sci., Arizona State Univ., Tempe, AZ 85287-1908), and Christophe Micheyl (UPRESA CNRS 5020, Pav. U, Hp. E. Herriot, Lyon 69437, France)

Spectral features are powerful cues in the perceptual organization of sound sequences. The ''peripheral channeling'' explanation of stream segregation predicts that the perception of fusion or fission in sequences, in a repeating ABA format, depends on the degree of spectral overlap of the A and B sounds. Nevertheless, previous studies have shown that other factors can also affect auditory stream segregation. Sequences of fully unresolved complex tones, bandpass filtered in the same frequency region, have been shown to perceptually split apart into two streams when the fundamental frequencies of the complexes are widely separated [Grimault *et al.*, J. Acoust. Soc. Am. **108**, 263–271 (2000)]. In the present experiment, normally hearing adults heard rapid alternations of a pair of flat-spectrum, sinusoidally amplitude-modulated broadband noise bursts. These noise bursts were segregated based on differences in modulation rate, as determined with a constant stimulus procedure as well as with a procedure in which segregation responses were collected continuously while the modulation rate of B was swept up or down during a block of trials. [Work supported by Fondation pour la Recherche medicale and NIDCD.]

9:45

1aPP4. Loudness-independent cues to object striking force. Stephen Lakatos (Dept. of Psych., Washington State Univ., 14204 NE Salmon Creek Ave., Vancouver, WA 98686) and James W. Beauchamp (School of Music, Univ. of Illinois–Urbana-Champaign, Urbana, IL 61801)

To what degree can listeners rely on auditory cues besides loudness to correctly judge the striking forces used to excite an object? Listening experiments used a discrimination task to assess whether individuals could correctly map loudness relations onto striking forces for stimuli derived from tam–tam, tympani, and xylophone-bar recordings. On each trial, listeners heard two sound pairs representing an instrument struck with differing strike forces using a mallet; one pair had its loudness relations intact while the other had its loudness relations inverted. Listeners determined which pair preserved its correct loudness relations. Both striking force and mallet properties served as independent variables. Discrimination varied in general with the extent to which the mallet and striking force excited the characteristic vibrational modes of the instruments: the more acoustic modes available, the higher the discrimination performance. Discrimination was significantly poorer for instruments with fewer acoustic cues generated by such modes, like the xylophone, than it was for instruments with a rich set of cues, like the tam–tam. Mallet density and weight were also important factors in modulating discrimination. The results indicate source-based cues besides loudness can convey information about striking force, albeit with less effectiveness. [Research supported by Air Force Grant F49620-99-1-0293.]

10:00

1aPP5. Perceptual learning and attentional cues in nonsimultaneous masking. Marina M. Rose, Douglas E. H. Hartley, and David R. Moore (Univ. Lab. of Physiol., Oxford Univ., Parks Rd., Oxford OX1 3PT, UK, marina.rose@physiol.ox.ac.uk)

This study investigated training effects in three nonsimultaneous masking conditions, and the possible role of an attentional cue. Three groups of six participants trained on seven consecutive days on either (i) backward masking (BM, 0-ms delay), (ii) combined forward and backward masking with a short inter-masker delay (10-ms, FBS) or (iii) combined forward and backward masking with a long inter-masker delay (110-ms, FBL). The signal was a 10-ms, 1-kHz tone pip, and the masker was a 300-ms, 1.6-kHz wide noise (40 dBSPL/Hz), centered at 1 kHz. Training sessions had 15 runs (30 presentations/run), using a 2AFC, maximum-likelihood procedure. The FBS condition showed the highest amount of masking, followed by the BM and FBL conditions, respectively ($p < 0.05$). A small but significant improvement was found between training days when all three groups were combined ($p < 0.001$). Pre-training data showed, for most subjects, lower FBL than BM thresholds. This could indicate that the first noise burst in the FBL condition did not mask the signal, but rather acted as a temporal and attentional cue, helping to ''listen at the right time.'' [Work supported by the National Health Service, the Wellcome Trust, and the Medical Research Council.]

10:15

1aPP6. Rapid improvements on interaural-time-difference discrimination: Evidence for three types of learning. Jeanette A. Ortiz, Beverly A. Wright, Matthew B. Fitzgerald, and Jagan Pillai (Audiol. and Hearing Sci. Prog., 2299 N. Campus Dr., Northwestern Univ., Evanston, IL 60208-3550, j-ortiz@northwestern.edu)

Listeners often show rapid improvements when first performing auditory discrimination tasks. This learning was explored by investigating how training 300–1500 trials on one of three conditions (all 2IFC) influenced interaural-time-difference (ITD) discrimination on a target ITD condition (lateralization, 0-μs ITD, 500 Hz). Listeners trained on ITD or interaural-level-difference (ILD) discrimination had equivalent ITD-discrimination thresholds that were significantly lower than those of naive listeners. However, listeners trained on duration discrimination had ITD-discrimination thresholds between those of naive and ITD-trained listeners. These data suggest that rapid decreases in ITD-discrimination threshold result partly from learning the general procedure (induced by the duration training) and partly from learning the lateralization task (induced by the ILD training). In contrast, compared to naive listeners, final within-subject standard deviations of the ITD-discrimination thresholds were significantly smaller for ITD-trained, somewhat smaller for ILD-trained, and no smaller for duration-trained listeners. These data suggest that rapid reductions in threshold variabiilty result partly from learning the lateralization task (induced by the ILD training) and partly from learning something provided, here, only by ITD training. Thus, at least three different types of learning can contribute to initial improvements in performance on auditory discrimination tasks. [Work supported by NIDCD.]

10:30–10:45 Break

10:45

1aPP7. The impact of labeling versus detection training on performance in an asynchrony-detection task. Julia A. Mossbridge and Beverly A. Wright (Audiol. and Hearing Sci. Prog., 2299 N. Campus Dr., Northwestern Univ., Evanston, IL 60208-3550, j-mossbridge@northwestern.edu)

It is not clear which training methods yield the fastest improvement on perceptual skills. To address this issue, performance on a detection task was compared following training on a labeling or detection task. The labeling task was to determine, for each presentation, whether the onsets of two tones were synchronous or asynchronous. Tone onsets were separated by one of six durations. The detection task was to determine which of two presentations contained tones with asynchronous onsets. The duration between tone onsets was varied adaptively to establish the asynchrony-detection threshold (79% correct detections). In both tasks, the tonal frequencies were 0.25 and 4 kHz. Labeling-trained listeners ($n = 11$) had detection thresholds lower than those of untrained controls ($n = 6$) following just 600 labeling trials ($p = 0.008$). In contrast, detection-trained listeners ($n = 6$) had detection thresholds lower than controls only following 3660 detection trials ($p = 0.027$), 2160 trials on the asynchrony-detection task plus 1500 trials on related temporal-acuity tasks. Thus, labeling training appears to improve performance on an asynchrony-detection task more quickly than detection training itself. By increasing our understanding of perceptual learning processes, these data may aid the treatment of individuals with disorders in auditory perception. [Work supported by NIDCD.]

11:00

1aPP8. Specificity of learning in an auditory asynchrony-detection task. Beverly A. Wright, Matthew B. Fitzgerald, and Julia A. Mossbridge (Audiol. and Hearing Sci. Prog., 2299 N. Campus Dr., Northwestern Univ., Evanston, IL 60208-3550)

The ability to determine whether the onsets of two tones at different frequencies are asynchronous is a measure of temporal acuity. Learning in this asynchrony-detection task and its generalization to untrained frequencies and temporal-acuity tasks was explored. For six listeners, asynchrony-detection thresholds for tones at 0.25 and 4 kHz improved from 79 to 20 ms with 4320 training trials. A z-score analysis revealed that trained listeners learned significantly more than untrained controls on the trained condition, but only slightly or no more than controls on: (1) the trained task at two sets of untrained frequencies (0.5 and 1.5 kHz, 0.75 and 1.25 kHz), and (2) three untrained temporal-acuity tasks at the trained frequencies (offset asynchrony detection, onset temporal order, and offset temporal order). This generalization pattern parallels that seen when the trained task was to determine the onset order of two tones [B. A. Wright and J. A. Mossbridge, J. Acoust. Soc. Am. **107**, 2882 (2000)]. Thus, learning in both asynchrony-detection and temporal-order tasks appears to be mediated by task-specific, frequency-dependent mechanisms. By increasing our understanding of temporal-acuity mechanisms and their plasticity, these data may guide the treatment of related hearing disorders. [Work supported by NIDCD.]

11:15

1aPP9. Psychophysical studies of auditory sensory traces. Denis McKeown (School of Psych., Univ. of Leeds, Leeds, UK), David Wellsted (Univ. of York, York, UK), and Tim Green (Univ. College London, London, UK)

The evidence is impressive, using an evoked potentials paradigm, that detailed internal models of recent sound stimulation are formed and compared with incoming sounds: Changes are incorporated, it is supposed, in order to update those models or sensory traces. Here psychophysical studies are reported. The first experiment poses the question of how recent tone stimulation affects ongoing performance in a discrimination task. Briefly, it is assumed that such a tone, or cue, may provide a highly frequency-specific sensory trace equivalent to sensory enhancement at that frequency. In a discrimination task two complex sounds occur in successive intervals. A single component of one complex is decreased in intensity: The task is to report whether this is in the first or second interval. Now, prior to each trial a tone cue is introduced: This is at the same frequency as the decreased component, or at a different frequency. The reasoning is that if the cue tone produces sensory enhancement, then it should effectively remove the decrement when cue and decrement are at the same frequency, impairing performance on these trials. This was the result obtained. Other experiments ued a detection task in examining the time source of the presumed sensory trace.

11:30

1aPP10. Unsigned coding of stimulus change. Ervin R. Hafter, Erick Gallun (Dept. of Psych., Univ. of California, Berkeley, CA 94720), Anne-Marie Bonnel (Ctr. Natl. de la Recherche Scientifique, Marseille, France), Brent Edwards, and Neal Viemeister (Univ. of Minnesota, Minneapolis, MI 55455)

Various studies have pointed to a special importance for stimulus transients in such diverse areas as the overshoot in masking and onset dominance in sound localization. Evidence that the internal representation of transience may be unsigned is found in cases where listeners readily detect a change in the ongoing stimulus without successfully identifying the direction of that change. We will review these results and present new data on sensitivity to change without sign in such dimensions as frequency and lateral location. Unusual sensitivity to transience has implications for the perception of natural scenes, where change per se may represent the code for simple alerting signal. [Work supported by grants from NIH, DC00087 and DC00683.]

11:45

1aPP11. Are cognitive processes involved in the perceived degree of urgency of auditory warning signals? Anne Guillaume, Lionel Pellieux, Veronique Chastres (Imassa BP73, 91223 Bretigny sur Orge, France), and Carolyn Drake (Universite de Paris V, Boulogne-Billancourt Cedex, France)

Studies on the perceived degree of urgency of auditory warning signals have demonstrated that the greatest sense of urgency is created by sequences with a high fundamental frequency, many irregular harmonics, and fast sequence rate (Edworthy *et al.*, 1991). However these effects disappeared when subjects were under high workload. In order to further investigate the cognitive processes involved, listeners' perception of two sets of warning sounds were tested: one constructed following Edworthy's indications, and a second composed of real alarms currently used by the French airforce. A first group of listeners evaluated the perceived degree of urgency between pairs of alarms. A second group quantified the perceived difference between the two alarms. A multidimensional analysis demonstrated that the spatial distribution of the *constructed alarms* was similar for both tasks. However, for the *real alarms*, a different pattern was observed: Some alarms were classified as nonurgent although their acoustical characteristics would have indicated a high-urgency classification. The conclusion is that the perceived degree of urgency is determined by both low-level perceptual factors and more complex cognitive processes involving long-term memory and personal experience.

MONDAY MORNING, 4 JUNE 2001 RED LACQUER ROOM, 8:30 A.M. TO 12:00 NOON

Session 1aSC

Speech Communication: Vowels, Sonorants, Consonants: Acoustics and Perception (Poster Session)

Amy T. Neel, Chair

Department of Audiology and Speech Sciences, Purdue University, 1353 Heavilon Hall, West Lafayette, Indiana 47906

Contributed Papers

All posters will be on display from 8:30 a.m. to 12:00 noon. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 8:30 a.m. to 10:15 a.m. and contributors of even-numbered papers will be at their posters from 10:15 a.m. to 12:00 noon.

1aSC1. Effects of instructions to produce hyperarticulated vowels on formant values. Marianne Pouplier (Haskins Labs, 270 Crown St., New Haven, CT 06511 and Yale Univ., marianne.pouplier@yale.edu) and D. H. Whalen (Haskins Labs, New Haven, CT 06511)

Clear speech collected in the laboratory is typically well articulated, but speakers can produce even clearer speech (''hyperarticulated'') when so instructed. Johnson *et al.* [Language **69**, 505–528 (1993)] found that hyperarticulations, averaged across speakers, are more extreme in both the $F1$ and $F2$ dimension. Here, 10 speakers of a Rhode Island dialect of English recorded words with 11 different vowels in (h)Vd environments (e.g., ''heed,'' ''aid''). Individuals show the effect for between 2 and 9 vowels; there was no vowel that had more extreme formant values (for $F1$ or $F2$) for all 10 speakers. (Averaged across talkers, only the low front vowel had more extreme values in both $F1$ and $F2$.) Typically, hyperarticulations that were not more extreme in formant values were equivalent to the normal production, while a few had less extreme values. The lack of consistency in the individual results calls into question the use averaged values, since speakers may be changing articulatory parameters in a more

complex way. The dynamic nature of some of the vowels may have an influence, since the initial measurements were taken from a single point in the syllable. [Work supported by NIH.]

1aSC2. On trading relations in vowel production: An x-ray microbeam study. Angela Slama and Gary Weismer (Univ. of Wisconsin–Madison, 1975 Willow Dr., Madison, WI 53706, slama@waisman.wisc.edu)

The purpose of this study was to investigate the articulatory and acoustic variability in vowel production, across 52 normal speakers of English. Specific to that variability, the notion of trading relations was addressed. In a study by Perkell *et al.* (1993), trading relations describes the complementary nature by which two different articulatory mechanisms are employed by a single speaker, when producing a given speech sound. Perkell and colleagues found weak, negative correlations within three out of four speakers, for the tongue-raising and lip-rounding mechanisms used to produce the English vowel /u/. These findings suggest that within a given speaker, the two primary articulatory mechanisms used to produce /u/ may be traded, or used in a complementary fashion, to constrain the acoustic variability of the vowel. In the current experiment, kinematic and acoustic measurements were used to explore the possibility of the same kinds of trading relations across speakers. The specific question asked was, do speakers who demonstrate large amounts of lip rounding for the vowel /u/, also use small amounts of tongue backing and/or rounding, and vice versa? Interspeaker variability for the acoustic and kinematic measures will be evaluated and discussed relative to this question. [Work supported by NIH No. DC03274.]

1aSC3. Male and female vowels identified by visual inspection of raw complex waveforms. Michael A. Stokes (MAS Enterprises, 3926 Graceland Ave., Indianapolis, IN 46208, masmodel@indy.net)

A perceptual experiment involving visual identification of the raw complex waveforms of American English vowels from two male speakers was presented at the 131st meeting of Acoustical Society of America (Stokes, 1996). In that study, a subject correctly identified five out of nine vowels for both male speakers using only a visual presentation of the raw complex waveforms. The present study replicates those earlier results with the same subject correctly identifying four out of nine vowels and six out of nine vowels for two new male speakers. Furthermore, four out of nine vowels produced by a female speaker were correctly identified. These results demonstrate that the visual cues being used for vowel identification can be applied across genders, in addition to extending the work from 1996 to now show successful visual identification of vowels across four male talkers using only raw complex waveform displays. These results will be discussed, including an analysis of the errors made by the subject. A description of the visual cues and the new model of vowel perception and production that resulted from these cues can be found at http://www.indy.net/~masmodel.

1aSC4. Exploring the ''hyperspace'' effect in vowel perception. D. H. Whalen (Haskins Labs, 270 Crown St., New Haven, CT 06511) and Harriet Magen (Haskins Labs, New Haven, CT 06511)

When using the method of adjustment to choose the best representative of an English vowel in an $F1$ by $F2$ space, listeners have been found to select formant values more extreme than those they actually produce [Johnson *et al.*, Language **69**, 505–528 (1993)]. This preference for ''hyperspace'' targets (those beyond actual productions) is argued to indicate a phonetic target more extreme than the subjects' actual hyperarticulations. The same methodology was used with 10 Rhode Island subjects. They produced vowels under two conditions, citation and then (after the perceptual task) hyperarticulated speech. Perception was assessed with a grid of 298 $F1/F2$ combinations from which listeners chose the best match for a keyword's vowel. Females and males were nearly equivalent in perception, as assumed previously but not shown. However, only one vowel was clearly extreme in both formants, with seven other vowels being more extreme on one formant, two vowels showing hypoarticulation, and the final one showing no effect. This lack of a robust effect may be due to difficulty of using $F2$ in perceiving the synthetic vowels. It may also be that the listeners were selecting perceptual targets from a more neutral dialetic than their own. [Work supported by NIH.]

1aSC5. Variations in coda sonorants at different speaking rates. Jodi P. Bray (Univ. of Florida Prog. in Linguist., P.O. Box 115454, Gainesville, FL 32611-5454)

Despite sonority's role in phonological accounts of syllable structures in the world's languages, phonetic correlates of sonority remain controversial. This study is part of a larger project designed to examine possible phonetic cues to sonority. Specifically, this paper reports on acoustic characteristics (i.e., duration, amplitude, and $F0$) of sonorants [r, l, m, n] occurring word finally or before an obstruent. The goal is to examine (a) whether these acoustic characteristics vary across speaking rate: slow, casual, and rapid; (b) whether these characteristics vary across speakers. Preliminary results based on data from one male native English speaker suggested that sonorant duration did not seem to vary at different speaking rates. Moreover, the ratio of sonorant to rhyme duration is greatest in rapid speech suggesting that sonorants require duration for their perceptibility. Additional speakers will be analyzed to confirm these findings. A perception experiment designed to examine degree of perceptibility of sonorants at different speaking rates and in different degrees of noise will also be conducted. Perceptual data will be examined and discussed in relation to production data.

1aSC6. A documentation of phonetic detail in onset-embedded words. Katherine Crosswhite (Dept. of Linguist., Univ. of Rochester, Rochester, NY 14627, crosswhi@ling.rochester.edu), Mikhail Masharov (Dept. of Brain and Cognit. Sci., Univ. of Rochester, Rochester, NY 14627, mym@bcs.rochester.edu), Joyce M. McDonough (Univ. of Rochester, Rochester, NY 14627, jmmcd@ling.rochester.edu), and Michael Tanenhaus (Univ. of Rochester, Rochester, NY 14627, mtan@bcs.rochester.edu)

As part of a study of online processing of onset-embedded words, we documented phonetic-level properties of speech production, such as duration differences, that may cue for word identification, segmentation and/or lexical access. Forty nouns (20 pairs like cap/captain, doll/dolphin) were placed in sentences in three frames (nuclear, prenuclear, and utterance-final position). Each sentence was read, then pronounced from memory by eight speakers randomly ordered. In keeping with previous research, instrumental analyses revealed strong effects of syllable count on vowel duration [Klatt, J. Phonetics **7**, 279–312 (1979)] and initial consonant VOT [Lisker and Abramson, Lang. Speech **10**, 1–28 (1967)]. A similar effect on initial consonant duration was also found. These effects were stable across speakers, words and contexts, suggesting they may be available as cues to overall word length in online segmentation. To test this, the noun pairs of one speaker were cross spliced to produce congruous and incongruous tokens, and presented to listeners in a ''visual world'' eye-tracking paradigm [Allopenna *et al.*, J. Mem. Lang. **38**, 419–439 (1998)]. We will report on whether beneficial durational cues of congruous tokens and/or potentially misleading incongruous cues had more influence on online word segmentation of our subjects. [Work supported by NIH and NSF.]

1aSC7. What happens to segment durations at the end of a word? Rebecca Schwarzlose and Ann R. Bradlow (Dept. of Linguist., Northwestern Univ., 2016 Sheridan Rd., Evanston, IL 60208)

Coda consonant clusters of monosyllabic words were studied to determine what durational modifications occur at the segment level as a function of the number, the order, and the morphemic status of segments in the coda. We focused on the segment /s/ because it has the fewest constraints in terms of position in coda clusters, allowing reversals such as tax (/tæ ks/) and task (/tæsk/), and because /s/ can also be a separate morpheme, allowing us to compare segment durations for tax and tacks (/tæ ks/). Preliminary results from four talkers reaffirm previous findings that segment durations decrease as the number of segments increases, although all segments do not decrease to the same extent. Our data also reaffirm previous claims that morphemic /s/ has a longer duration relative to the overall word duration than its nonmorphemic counterpart. In addition to these factors, position at the right-edge boundary of a word seems to play a special role in determining the relative durations of segments within a coda cluster. Specifically, we found that the balance between segment durations in /sk/ and /ks/ coda clusters in monomorphemic words tended to tip in favor of the word-final segment and against the word-internal member of the cluster.

1aSC8. Use of formant movement detail in vowel identification. Amy T. Neel, Ananthanarayan Krishnan, and Kent J. Collins (Dept. of Audiol. & Speech Sci., Purdue Univ., West Lafayette, IN 47906)

Many studies have shown that American English listeners identify vowels with dynamic formants better than flat-formant vowels. The amount of formant movement detail used by listeners, however, is not known. This study examined use of formant movement detail in vowel identification and discrimination using five sets of Klatt-synthesized vowel stimuli modeled after natural vowels. One-point stimuli consisted of flat-formant vowels with formant values taken from the center of the natural vowels. The other four sets consisted of tokens with two, three, five, or eleven points having formant values taken from the natural vowels with the remainder of the values being linearly interpolated. Two-point vowels, with sloping formants, were least well identified by listeners, whereas the most detailed stimuli, the five- and eleven-point vowels, were the best identified tokens. Listeners had difficulty discriminating between the five- and eleven-point representations of vowels. Results indicate that listeners may have an intermediate level of formant movement detail representation — more detailed than a simple slope from onset to offset, but less detailed than natural vowel patterns. In addition to behavioral data, frequency-following responses (FFRs), electrophysiologic measures of neural encoding of formant movement patterns, will be presented. [Work supported by 2001 Research Grant in Speech Science, ASHF/ASA.]

1aSC9. Acoustic and articulatory characteristics of Korean affricates. Hyunjoo Chung and Anna Marie Schmidt (School of Speech Pathol. and Audiol., Kent State Univ., Kent, OH 44242, hchung@kent.edu)

The purpose of this study was to examine the three voiceless Korean affricates in terms of differing acoustic and articulatory characteristics which might determine three-way lax, aspirated, and tense distinctions. In previous studies, the three-way stop distinction and the two-way fricative distinction (lax and tense) were investigated. Yet little is known about how the three distinctions are implemented in affricates. The three Korean affricates in CV and VCV syllables produced by two female and one male native Korean speaker were compared. Duration, amplitude, and fundamental frequency in six vowel contexts for each affricate were measured on spectrograms. In addition, temporal and spatial characteristics of lingual gestures during production of affricates such as seal duration and skewness of articulatory movement were measured from electropalatographic data. Endoscopic and air flow data for two different subjects were also examined for differences in measures such as glottal width and air flow. Finally, measurements of affricates will be reported and discussed in comparisons to previous findings on stops and fricatives with respect to the three-way distinctions. [Speech Communication Best Student Paper Award.]

1aSC10. Some acoustic cues for categorizing American English regional dialects. Cynthia G. Clopper and David B. Pisoni (Speech Res. Lab., Psych. Dept., Indiana Univ., Bloomington, IN 47405, cclopper@indiana.edu)

The perception of phonological differences between regional dialects of American English by naive listeners is poorly understood. Using the TIMIT corpus of spoken sentences produced by talkers from a number of distinct dialect regions in the United States, an acoustic analysis conducted in Experiment I confirmed that several phonetic features distinguish between the dialects. In Experiment II recordings of the sentences were played back to naive listeners who were asked to categorize each talker into one of six geographical dialect regions. Results suggested that listeners are able to reliably categorize talkers into three broad dialect clusters, but have more difficulty accurately categorizing talkers into six smaller regions. Correlations between the acoustic measures and both actual dialect affiliation of the talkers and dialect categorization of the talkers by the listeners revealed that the listeners in this study were sensitive to acoustic-phonetic features of the dialects in categorizing the talkers. Taken together, the results of these experiments suggest that naive listeners are aware of phonological differences between dialects and can use these differences to categorize talkers by dialect. Implications for the study of variation and variability in speech will be discussed. [Work supported by NIH.]

1aSC11. Similarities in acoustic modeling for /r/ and /l/. Carol Y. Espy-Wilson (Dept. of Elec. and Computer Eng., Boston Univ., Boston, MA 02115, espy@bu.edu) and Suzanne E. Boyce (Univ. of Cincinnati, Univ. of Cincinnati, Cincinnati, OH 45267)

It is well known that acoustically American English /r/ and /l/ show similar $F1$ and $F2$ values, with /l/ showing a high $F3$ and /r/ showing a low $F3$. Stevens [Acoustic Phonetics, 1998] models tip-down bunched /r/ and light /l/ with similar tongue configurations and lateral acoustic paths around the tongue. Two major articulatory differences in this model are the more forward position of the tongue/palate contact for /l/ vs /r/ and the longer constriction for bunched /r/ vs /l/. Ong and Stone [Phonoscope **1** (1998)] note that the /l/ and /r/ in their data showed very similar pharyngeal volumes, accounting for the similarity in $F1-F2$ patterns for these segments. In this paper, we explore the possibility that these variants of /r/ and /l/ share acoustics mechanisms for generating $F1$ and $F2$, and that the difference in $F3$ for /l/ vs /r/ arises primarily from palatal constriction length and front cavity shape.

1aSC12. Acoustic properties underlying vowel intelligibility in words with mixed consonantal contexts. Sarah Hargus Ferguson, Diane Kewley-Port, and Aaron P. Brown (Dept. of Speech & Hearing Sci., Indiana Univ., 200 S. Jordan Ave., Bloomington, IN 47405, safergus@indiana.edu)

This study explored the relative importance of spectral target, dynamic formant, and duration information to vowel identification in mixed consonantal contexts. Vowel intelligibility was varied naturally by one talker producing test words in sentences using clear and conversational speech. In /bVd/ context only, vowels identified by normal-hearing listeners in noise were significantly more intelligible (15%) in clear speech than in conversational speech. Clear and conversational tokens of the 10 vowels also differed significantly on several spectral and temporal measures. Regression analyses revealed that intelligibility differences for individual vowels were well predicted by various combinations of spectral target, dynamic formant, and duration metrics [S. H. Ferguson and D. Kewley-Port, J. Acoust. Soc. Am. **106**, 2272 (1999)]. A second data set was col-

lected concurrently with the /bVd/ data, with the same listeners identifying vowels in mixed CVC contexts. A significant intelligibility advantage (18%) for clear speech was found for these materials. However, regression analyses were less straightforward for vowels in CVC context than in /bVd/ context. Several acoustic metrics were investigated to determine whether intelligibility differences could be modeled for vowels in mixed consonantal contexts. The results will be discussed in terms of theories of vowel perception. [Work supported by NIHDCD-02229, the American Speech-Language-Hearing Foundation, and NIHDCD-00012.]

1aSC13. Influence of /r/ on burst spectra for stop place. Megan M. Hodge, Janka Hegedus, Megan Kenny, and Carmen Souster (Dept. of Speech Pathol. and Audiol., Rm. 2-70 Corbett Hall, Univ. of Alberta, Edmonton, AB T6G 2G4, Canada)

Bilabial and lingua-alveolar stops in 20 minimally contrastive word pairs in both singleton and stop /r/ (cluster) contexts were recorded from two girls, one with facial paralysis (CFP) and one with normal facial movement (CNM). Auditory identification of these productions by 12 listeners revealed a significant place by context interaction for CFP. Identification scores were high for her lingua-alveolar stops (99.3% singletons; 94.8% clusters). Identification scores for her bilabial stops were lower, with singletons being significantly lower (57.8%) than clusters (77.8%). Acoustic cues for stop place ($F2$ onset frequency, VOT, mean and skewness of burst spectra) were measured for all word productions. For both girls, $F2$ onset and VOT measures were lower for bilabial than lingua alveolar stops in singletons and clusters. CFP's burst spectra for bilabials had a higher mean and more negative skewness than lingua alveolars in singletons and clusters. CNM's burst spectra for bilabials had a lower mean and more positive skewness than lingua alveolars in singletons (expected) but a higher mean and more negative skewness for bilabials than lingua alveolars in clusters (unexpected). This unexpected finding may account for CFP's higher identification rates for bilabial clusters (all acoustic cues ''fit'' those of a child with normal facial movement).

1aSC14. A method for studying variability in fricatives using time-dependent changes in spectral mean. Benjamin Munson (Dept. of Commun. Disord., Univ. of Minnesota, 115 Shevlin Hall, Minneapolis, MN 55455)

A large literature has examined variability in temporal, spectral, and kinematic measures as indices of speech motor control. While much of this literature has examined factors that influence within-speaker variability in temporal parameters, relatively little research has examined variability in acoustic parameters, and, in particular, variability in consonant acoustics. This paper presents a new method for studying within-speaker variability in fricative production by examining dynamic characteristics of the spectral mean. In this method, nonlinear regression is used to predict values of the first spectral moment of a 10-ms window of frication noise from its position in the fricative. The resulting measure of model fit, R^2, is used as an index of within-speaker variability. This method was used on two corpora of archival data. In the first study, three English-speaking adults' productions of /s/ in initial /sp/ and /sw/ clusters was examined. /s/ was found to be spectrally less variable when embedded in /sw/ sequences than when embedded in /sp/ sequences. The second study examined variability in the spectra of /f/, /s/ and /ʃ/ spoken at normal and slow rates of speech by adults and two groups of children. Age-group, rate and phonetic context affected fricative variability.

1aSC15. Evaluating vowel normalization transformations through generalized linear modeling. Patti Adank (Dept. of Linguist. and Dialectology, Univ. of Nijmegen, E9.20, Erasmuslaan 1, 6500 HD, Nijmegen, The Netherlands) and Roel Smits (Max Planck Inst. for Psycholinguist., Nijmegen, The Netherlands)

Nearey [ICSLP (1992)] suggests using generalized linear modeling (GLM) to investigate the role of vowel-intrinsic (e.g., role of $F0$) and vowel-extrinsic (e.g., z-transformation, log-mean transformation) normalization schemes. A generalized linear model consists of two parts: a linear predictor and a nonlinear link function. GLM is used to predict values both for dependent variables with discrete distributions and for variables that are nonlinearly related to the predictors. GLM can therefore be useful for dealing with relations between vowel-intrinsic and vowel-extrinsic information and vowel categories, since intrinsic factors can be tested as predictors in the linear predictor function and extrinsic factors can be tested using the link function. The GLM approach was applied to speech data from 80 speakers from four different regions in The Netherlands. The regional accent spoken in each group was different from the other three groups. The nine monophthong vowels of Dutch were produced in a neutral context. For each vowel, $F0$ and $F1–F3$ were measured. These values were transformed according to several vowel normalization schemes. The coefficients for the normalizations were estimated through maximum likelihood estimation. Overall, the results indicate that GLM can be a useful tool for evaluating normalization transformations.

1aSC16. Temporal trading relationship between singletons and geminates in different vowel contexts. Mohamed Al-Khairy (Prog. in Linguist., P.O. Box 115454, Univ. of Florida, Gainesville, FL 32611-5454, mohamedk@ufl.edu)

Length is phonemically contrastive in both consonants and vowels in Arabic. The goal of this study is to examine whether a trading relationship exists between the duration of singletons and geminates in the environment of short and long vowels in Arabic. Preliminary results based on duration measurements of 48 tokens of word-medial bilabial stop singletons and geminates spoken by four speakers suggested that, as expected, geminates are always longer (2.6 times on the average) than singletons. Long vowels are always longer (2.45 on the average) than short vowels. A trading relationship, however, exists between consonant and vowel durations. First, short vowels became significantly shorter when occurring before geminates. This is true only when the vowel following the geminate is also short. Second, geminates preceding short vowels are significantly longer (3.2 vs 2.3 times on the average) than geminates preceding long vowels. To see if these preliminary results also hold true for other types of stop consonants in Arabic, more data will be collected. Similarities and differences in the trading relationship between these consonants and vowels will be examined. Cross-language comparisons regarding the trading relationship between singletons and geminates, and their neighboring vowels will be discussed.

1aSC17. Identification of Swedish vowel quantity by pre-adults. Dawn M. Behne (Norwegian Univ. of Sci. and Technol., NO-7491 Trondheim, Norway), Peter E. Czigler, and Kirk P. H. Sullivan (Umeaa Univ., SE-901 87 Umeaa, Sweden)

Research on the perception of Swedish vowel quantity by adult native listeners shows that vowel duration is a primary cue for identifying vowel quantity. However, when a vowel has a relatively long duration (e.g., due to its inherent duration, or its context), adult listeners may make use of both duration and spectra to identify vowel quantity. Use of the vowel spectrum in these special cases might be seen as the result of perceptual fine tuning to improve the processing efficiency in vowel identification. If so, we would expect a developmental change in how children use vowel duration and spectra to identify vowel quantity, with the use of vowel spectra as a cue coming relatively late. This investigation examines how young native listeners, from 9 to 15 years old, use vowel duration and $F1$-$F2$ to identify Swedish vowel quantities. For each of three vowel types,

100 /kVt/ words were resynthesized with ten degrees of vowel duration and ten degrees of $F1$ and $F2$ adjustment. Each child responded to a subset of these in a rhyming task. Results show a late developing ability to identify vowel quantity and trace the use of vowel duration and $F1$-$F2$ as perceptual cues to vowel quantity.

1aSC18. Quality of front vowels before /r/. Michael J. Clark and James M. Hillenbrand (Speech Pathol. and Audiol., Western Michigan Univ., Kalamazoo, MI 49008)

The vowels /i/ and /ɪ/ are not contrastive before /r/ in American english, and the phonetics literature is equivocal about which symbol to use for the nucleus in words such as *beer*. Similarly /e/ and /ɛ/ are not contrastive before /r/, and the literature contains varied references to one or the other in words such as *bear*. Trained listeners disagree about representing the vowel in /r/-final words. Ten men and ten women recorded monosyllables containing /i,ɪ,e,ɛ/, with initial /b/ and /h/ and with no final consonant or with final /d/ and /t/. For example, the /b/-initial high front series was *bee*, *bead*, *beet*, *bid*, *bit*. Additionally an /r/-final set (*beer*, *bear*, *hear*, *hair*) was recorded. Acoustic measurements included the first two formants. Formants for /i,ɪ,e,ɛ/ in /d/-final and /t/-final syllables (and for /i,e/ in open syllables) were typical of those measured in an earlier study of subjects from the same region. The high front vowel in /r/-final syllables (*beer* and *hear*) showed formant values intermediate between /i/ and /ɪ/, but closer to /i/. The mid front vowel (in *bear* and *hair*) fell entirely within the range seen for /e/. Listening tests using synthetic vowels and backward versions of the original syllables will be discussed. [Work supported by NIH.]

1aSC19. Duration and rate effects on English vowel identification by native Spanish listeners. Terry L. Gottfried and Adam M. Berman (Dept. of Psych., Lawrence Univ., Appleton, WI 54912)

Several American English vowel contrasts are specified by co-varying formant frequency and duration information. Listeners alter their vowel identification according to speaking rate, apparently making judgments about relative vowel duration. Two series of spectral continua were previously created from natural speech, ''beat-bit'' and ''pat-pet,'' varying $F1$ and $F2$, and syllable duration [Berman *et al.*, J. Acoust. Soc. Am. **105**, 1402 (1999)]. These syllables were inserted into sentence contexts of two rates (normal and fast). Longer syllable duration led to more ''beat'' and ''pat'' responses; faster rate contexts also led to more ''beat'' and ''pat'' responses. The present research tested native speakers of Spanish on their use of duration and rate context in English vowel identification. Despite the absence of Spanish vowel contrasts based on duration, native Spanish listeners were significantly affected by syllable duration and sentence rate in their identification of the ''beat-bit'' contrast. They also showed a significant duration effect for ''pat-pet,'' but the magnitude of this duration effect was significantly smaller for these non-native listeners than for native listeners, and there was no significant effect of sentence rate. With less English experience, non-natives might show even smaller effects of duration and sentence rate.

1aSC20. The acoustics of gestural overlap. Khalil Iskarous (Haskins Labs., 270 Crown St., New Haven, CT 06511)

Acoustic modeling of speech production focuses on the effects of a single tongue constriction and a simultaneous labial constriction on the formant frequencies of an acoustic tube. But research on dynamic speech production reveals that tongue gestures overlap in time, so two or more simultaneous tongue constrictions affect the acoustic output of the tube. Understanding the acoustics of natural speech therefore requires an investigation of the acoustic effects of multiple constrictions. The goal is to generate a generalized formant nomogram that would reveal how the acoustic effects of single constrictions combine and interact in determining the acoustic output of the whole tract. Results will be presented from a finite element simulation of the acoustic effects of two constrictions that vary independently in location and size. The location of each constriction varied in 34 steps and the size varied in 16 steps. Independent variation of the two constrictions yields 295 936 constriction combinations. The function relating the acoustic output of the multiply constricted tract to the acoustic effects of each of the constrictions when made alone is generally nonlinear. Several features of this function will be discussed, and conditions for its linearization will also be presented. [Work supported by NIH.]

1aSC21. Center of gravity effects in the perception of virtual formant transitions. Nandini Iyer, Ewa Jacewicz, Lawrence L. Feth, and Robert A. Fox (Dept. of Speech and Hearing Sci., The Ohio State Univ., 1070 Carmack Rd., Columbus, OH 43210)

When $F2$ and $F3$ of vowels are separated by up to 4.3 Barks (6.3 ERBu) and amplitude modulated to change their spectral centroid over time, listeners perceive a transition similar to that in two-formant diphthonglike vowels [Lublinskaja, 1996]. This study replicates Lublinskaja's finding relating to Russian vowels and extends its application to the consonant-vowel (CV) transitions in /wi/ and /ju/ in English. Stimuli were two-formant or three-formant CVs generated by an upward (/wi/) or downward (/ju/) transition of $F2$ over the frequency range $\Delta f = 100 - 800$ Hz, in steps of 100 Hz. For the three-formant CVs, $F2$ and $F3$ were separated by the same Δf as in the two-formant CVs, and their amplitudes were modulated so as to move the spectral centroid up or down. The differences in amplitude between $F2$ and $F3$ were either 10 dB or 20 dB. The stimuli were 88 ms in duration. Listeners identified the stimuli in a one-interval, forced-choice identification procedure. Results for the English CV combinations are discussed in relation to the findings reported in Lublinskaja's study. [Research supported by a grant from The Ohio State University College of Social & Behavioral Science to L. Feth and an INRS Award from NIH to R. Fox.]

1aSC22. Perception of relative formant amplitude for /i/ and /I/ in English. Ewa Jacewicz (Dept. of Speech and Hearing Sci., The Ohio State Univ., 110 Pressey Hall, 1070 Carmack Rd., Columbus, OH 43210)

This work investigates listeners' sensitivity to variations in the intensity of $F2$ and $F4$ for the English /i/-/I/ distinction using four-formant synthetic vowels. Two nine-step /i/-/I/ continua were synthesized with the amplitudes of $F2$ and $F4$ increasing or decreasing in steps of 2 dB. During 2AFC discrimination tasks, listeners were able to detect the 2-dB difference for both formants. /i/ was perceived as more natural with decreasing intensity of $F2$ and as less natural with increasing intensity of $F2$. The opposite was observed for /I/. Stepwise intensity changes of $F4$ were perceived as a difference in timbre for both vowels but not as a difference in naturalness. These data suggest that, perceptually, the patterns of relative formant amplitude provide information about the naturalness of /i/ and /I/ in English. The selective auditory sensitivity to the differential contribution of the intensities of $F2$ (variation in naturalness) and $F4$ (variation in timbre) is manifested when the 2-dB difference is detected for both frequency peaks. [Work supported by an INRS Award from NIH to R. Fox.]

1aSC23. Spectral tilt versus formant frequency in static and dynamic vowels. Michael Kiefte and Keith R. Kluender (Dept. of Psych., Univ. of Wisconsin–Madison, Madison, WI 53706)

The relative importance of gross (e.g., spectral tilt) versus detailed (e.g., formant peak frequencies) for speech perception is an enduring question of substantial importance in understanding performance of both impaired and unimpaired listeners [cf. Kiefte, J. Acoust. Soc. Am. (submitted)]. In this study, subjects were presented with steady-state stimuli varying perceptually between [i] and [u] for which spectral tilt and formant frequencies were fully crossed. Results show an effect for both tilt and formants indicating some perceptual importance for both in the iden-

tification of statically specified vowels. However, this result does not appear to extend to more natural sounding stimuli with dynamic formants. For example, there was no effect for spectral tilt in a similar experiment using diphthongs ranging from [ai] to [au]. One hypothesis for this difference is that the F_2 of front vowels is masked by F_1 when combined with sharply falling spectral tilt and that this effect is mitigated in dynamic or naturally produced vowels. Results are also presented which compare the perception of these continua with similar stimuli in which F_2 is absent as well as with dichotically presented stimuli in which F_1 and F_2 are heard in different ears. [Work supported by NIDCD DC04072.]

1aSC24. Vowel formant discrimination in quiet and noise for natural speech. Chang Liu and Diane Kewley-Port (Dept. of Speech and Hearing Sci., Indiana Univ., Bloomington, IN 47405)

Everyday communication takes place in background noise. The purpose of this study is to investigate the effects of noise on formant discrimination for naturally spoken vowels. Our analysis of formant discrimination for synthetic vowels found that formant resolution could be accounted for by excitation pattern and specific loudness models [D. Kewley-Port and Y. Zheng, J. Acoust. Soc. Am. **103**, 1654 (1998)]. It is expected that excitation pattern and specific loudness model can be extended to account for the formant discrimination in natural speech in both quiet and noise, even though natural speech has more variability than synthetic speech. The present study measured formant frequency discrimination in quiet and noise for normal-hearing listeners. Thresholds were measured as ΔF in natural speech for $F1$ and $F2$ for a set of vowels /ɪ, ɛ, ʌ, æ/ spoken by a female talker. Noise conditions used a long-term averaged speech spectrum to shape the noise. The signal-to-noise ratio (SNR) was manipulated systematically from -6 to $+6$ dB. Results showed different masking patterns for $F1$ and $F2$ with performance approaching asymptote between 0 and $+3$ dB SNR. Results from an analysis of these data with a specific loudness pattern model will be presented. [Work supported by NIHDCD-02229.]

1aSC25. Dynamic information in the perception and production of Rhode Island vowels. Harriet Magen (Haskins Labs, 270 Crown St., New Haven, CT 06511 and Dept. of Commun., Rhode Island College, Providence, RI 02908, hmagen@ric.edu) and Khalil Iskarous (Haskins Labs, New Haven, CT 06511)

Comparisons were made of the production and perception of 11 vowels for 10 Rhode Island speakers. To assess perception, the method of adjustment (MOA) was used. Subjects were presented with a grid of 298 steady state $F1/F2$ combinations and asked to choose the best match to a visually presented keyword. [Johnson *et al.*, Language **69**, 505–528 (1993)]. The same subjects produced the 11 vowels; acoustic measurements were made 1/3 of the way through. For two pairs of vowels ([eɪ]-[ɪ], [ɔ]-[oʊ]) there was a disparity between production and perception, calling into question the perceptual strategy listeners use when normally present dynamic information is absent. To address the issue of the dynamics of production, measurements were made at additional points. At all three points measured, values of $F1$ of [eɪ] and [ɪ] overlap substantially; however, perceptually they are distinct. While this dialect maintains the [ɑ]-[ɔ] distinction, [ɔ] and [oʊ] show overlap in production early in the vowel, diverging later in the vowel; static formant values chosen as the best instance of [oʊ] overlapped with the formants of producted [ɔ]. The efficacy of the MOA for dynamic information will be discussed. [Work supported by NIH and the RIC Faculty Research Fund.]

MONDAY MORNING, 4 JUNE 2001 PDR 18, 7:45 TO 10:50 A.M.

Session 1aSPa

Signal Processing in Acoustics: Signal Processing for Underwater Acoustics

Brian G. Ferguson, Chair
Maritime Systems Division, Defence Science Technology Organization, P.O. Box 44, Pyrmont 2009, Australia

Chair's Introduction—7:45

Contributed Papers

7:50

1aSPa1. Statistical analysis of acoustic Green's functions in a shallow ocean environment. Sean R. Chapin, George E. Ioup, Juliette W. Ioup (Dept. of Phys., Univ. of New Orleans, New Orleans, LA 70148, schapin@uno.edu), and George B. Smith (Naval Res. Lab., Stennis Space Center, MS 39529)

Knowledge of statistical properties of acoustic Green's functions can assist in the development of effective blind deconvolution algorithms for enhancing the detection and classification performance of active and passive sonar systems. One hundred synthetically generated Green's functions have been produced representing the acoustic propagation of signals in a given shallow ocean environment. Fourier analysis is performed on the Green's functions and various statistical properties are calculated in both the time and frequency domains, including examination of properties such as the mean, variance, and the third- and fourth-order moments in both domains. Statistics are also generated for the entire ensemble of Green's functions as a function of time and frequency. A similar statistical analysis is conducted on the autocorrelation of each of the Green's functions in both domains. Synthetic noise-free signals are produced by convolving the Green's functions with representative source signals. Statistical analysis is performed on the resulting signals in both domains for comparison to the results obtained using the Green's functions alone. Results include large phone-to-phone variations in kurtosis of the Green's functions and ensemble average kurtosis increasing sharply with time and decaying to zero. [Research supported by ONR.]

8:05

1aSPa2. Improved detection using propagation and target responses. Charles F. Gaumond and Ralph N. Baer (Naval Res. Lab., 4555 Overlook Ave. SW, Washington, DC 20375-5320)

The detection results of a monostatic sonar simulation are presented. The bandwidth of the simulation is 200 Hz to 950 Hz and the environment is range-independent, shallow water. A target is modeled as a point with a frequency-dependent response simulated using a finite element model and the propagation is simulated using the RAM PE model. These two broad-

band simulation results are combined using a coherent form of the sonar equation. The improvement of detection is shown for several cases comparing the use of environment-alone, target-alone, and combined-environment-target matched filters with the detection from the use of simple matched filtering. [Research sponsored by ONR.]

8:20

1aSPa3. Wavelet detection and denoising of low-frequency chirp signals. Joseph S. Wheatley, Juliette W. Ioup, and George E. Ioup (Dept. of Phys., Univ. of New Orleans, New Orleans, LA 70148, wheatley@datastar.net)

The detection and classification of underwater acoustic signals embedded in noise is a fundamental problem of interest to the Navy. The use of wavelet transforms is a recent development in digital signal processing that has been applied in many different areas. A particular type of wavelet is the chirplet, which includes frequency variation as well as time shift and scaling. The analysis of low-frequency signals containing multiple chirps using wavelet and chirplet techniques is demonstrated. Examples of low-frequency synthetic chirp signals are generated. Denoising and feature extraction of these signals using various wavelets with wavelet packet techniques are described.

8:35

1aSPa4. Time-reversal operator applied to broadband active sonar. David M. Fromm and Charles F. Gaumond (Naval Res. Lab., 4555 Overlook Ave. SW, Washington, DC 20375-5320)

The decomposition of the time reversal operator (DORT) is a single-frequency technique that isolates scatterers using a system of multiple sources and receivers [Prada *et al.*, J. Acoust. Soc. Am. **99**, 2067–2076 (1996)]. The application of this technique to a broadband active sonar system is shown through simulations in a realistic oceanic environment using RAM PE and a coherent reverberation model. A broadband signal containing reverberation, noise, and a target return is simulated. The broad bandwidth allows point targets or clutter to be isolated in range. The returns of interest are then time windowed, Fourier transformed, and DORT processing is applied. By propagating the eigenvectors of the time reversal operator to create time reversal images, the character of the return can be evaluated. [Research sponsored by ONR.]

8:50

1aSPa5. Iterative deconvolution for source recovery from multipath degraded signals. James P. Larue, George E. Ioup, Juliette W. Ioup (Dept. of Phys., Univ. of New Orleans, New Orleans, LA 70148, hubauto@juno.com), and George B. Smith (Naval Res. Lab., Stennis Space Center, MS 39529)

Known ocean propagation Green's functions can be used to recover an estimate of the source in underwater acoustic passive detection. Good estimates of the source improve the performance of classification algorithms. Since convolution models the propagation, deconvolution using the Green's function as the impulse response produces an estimate of the source. In the presence of moderate amounts of noise, deconvolution methods degrade due to the extreme noise sensitivity of the process. Iterative deconvolution has been shown to be more robust in the presence of noise than some other techniques due to the effect of a gradually changing window that need not be pushed to convergence. Iterative techniques are more difficult to apply for oscillatory data than for non-negative data because conservation of area does not provide a robust normalization. Iterative deconvolution techniques are tested for noise performance in source recovery and compared to other methods. [Research supported by ONR.]

9:05

1aSPa6. Matched field processing classification. Ahmad T. Abawi, Phil Schey (53560 Hull St., San Diego, CA 92152-5001), and W. S. Hodgkiss (Scripps Inst. of Oceanogr., San Diego, CA 92152-6400)

Matched field processing (MFP) has traditionally been performed using vertical line arrays. During the Shallow Water Cell Experiment SWellEx-96, which was conducted off the coast of San Diego in the spring of 1996, it was shown that MFP could also be done using horizontal line arrays (HLA). The MFP correlation using the 250-m long HLA used in this experiment was large enough for detection, but because of the high sidelobes, classification by depth was not possible. However, simulations predicted that the sidelobe levels for an approximately 500-m long HLA would be small enough to make classification by depth possible. To verify this prediction, the SWellEx-99 experiment was conducted in spring of 1999. During this experiment, which was conducted in approximately 200 m of water, a number of large ($\sim$500 m) horizontal line arrays were deployed. A J15-3 source, which transmitted multi-tone signals at various levels, was towed along various tracks. In this paper, data from the SWellEx-99 experiment are used to examine the detection and classification performance of adaptive matched field processing on large horizontal line arrays.

9:20

1aSPa7. Robust adaptive beamforming for hydrodynamic self-noise rejection. Vincent E. Premus and Stephen M. Kogon (MIT Lincoln Lab., Lexington, MA 02420)

White noise gain (WNG) is a metric widely used in the radar community for limiting the degree of adaptivity of an adaptive beamformer (ABF). Constraining adaptation of an ABF algorithm has two key effects: (1) protection against self-nulling associated with steering vector mismatch, and (2) limitation of white noise gain associated with squinting a beampattern to place a null on a mainlobe interferer. Hydrodynamic self-noise, or cable strum, commonly manifests itself as a source of mainlobe interference for passive acoustic towed horizontal line arrays. Cable strum is the result of vibrations excited in the array body by vortex shedding in the presence of flow. Strum is particularly detrimental on beams near endfire and typically exhibits very high dynamic range. As such, it requires an aggressive adaptation strategy for its effective removal. In this work, a white noise gain constraint for the rejection of cable strum is derived based on the scaled orthogonal projection approach of Cox *et al.* [IEEE Trans. ASSP **35** (1987)]. Tradeoffs involved in balancing strum rejection performance against mismatch-induced signal self-nulling are examined. [Sponsored by the Dept. of the Navy, under Air Force Contract No. F19628-00-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the authors and not necessarily endorsed by the U.S. Air Force.]

9:35

1aSPa8. Effective analysis technique of unstable acoustic signature from ship radiated noise. Jong-rak Yoon, Yong-ju Ro (Telematics Eng. Dept., Pukyong Natl. Univ., Pusan 608-737, Korea), and Jae-Jin Chun (Agency for Defense Development, Chinhae 645-600, Korea)

Acoustic signatures of ship radiated noise are important passive sonar parameters in target ship detection and classification. Among these signatures an unstable signature, which is defined as frequency change with respect to the time or frequency modulation, is due to the external loading

variation in specific machinery components. As an example, propeller blade frequency is measured as an unstable signature and is controlled by the speed governor which is dependent on the propeller resistance to sea surface waves or current turbulence. In this study an extended Kalman filtering algorithm is applied to detect the unstable signature. To obtain better performance than with the existing similar technique, state model parameters such as variance and bandwidth are expressed as a function of the sea state wave spectrum. In addition background noise variance in the measurement model is also defined as a sea state function. Its performance is evaluated using measured surface ship radiated noise data and it is also applied to Dopplerized signal analysis which is essential in moving target monitoring. [Work supported by the Underwater Acoustics Research Center.]

9:50

1aSPa9. Wavelet analysis of acoustic Green's functions in a shallow ocean environment. Antonio M. Morgan, Juliette W. Ioup, George E. Ioup (Dept. of Phys., Univ. of New Orleans, New Orleans, LA 70148, ammorgan2@hotmail.com), and George B. Smith (Naval Res. Lab., Stennis Space Center, MS 39529)

Knowledge of the properties of acoustic Green's functions can assist in the development of effective blind deconvolution algorithms for enhancing the classification performance of active and passive sonar systems. Green's functions have been produced representing the acoustic propagation of signals in a given shallow ocean multipath environment. Simulated noise-free received signals are the result of convolving the Green's functions with representative source signals. Wavelet analysis is performed on the Green's functions, the source signals, and the received signals. Forty different wavelets are tested. Comparison of the wavelet transform coefficients illustrates similarities and differences in signal properties. Significant differences can be exploited to separate the source from the Green's functions in the multichannel received signals for blind deconvolution source recovery. [Work supported by ONR.]

10:05

1aSPa10. An experimental investigation on a direct blind equalizer for underwater acoustic multiple FIR channel. Jianmin Liu, Yingmin Wang, Xiangrui Cao, and Zhengqi Zhang (Inst. of Acoust. Eng., Northwestern Polytech. Univ., Xi'an 710072, PROC)

A blind equalizer removes the distortion caused by intersymbol interference (ISI) without the need for training sequences. In order to get the complete phase information of the ISI, high-order statistics have to be used to find the equalizer taps. However, L. Tong *et al.* have shown that most channels can be identified from the second-order cyclostationary statistics through fractionally spaced equalizers. This result has stimulated considerable research activities in fractionally spaced channel identification and equalization. In this paper a single-input multi-output (SIMO) channel blind equalizer for an underwater acoustic communication system is considered. A direct blind equalization scheme was investigated by simulations, which is a novel unsupervised adaptive equalizer and was suggested by J. Labat first for single-input single-output (SISO). An interactive matched processing (MP) is added for the SIMO case. The work verified the prediction that the linear equation form of the direct blind equalizer lends itself naturally to adaptive solutions and optimum combinations of the equalizer outputs and improved performance over existing algorithms in terms of mean-square error and probability of error in the equalized data.

10:20

1aSPa11. Color rendering of range in passive acoustic displays. Frank A. Boyle and Thomas H. Phipps (Appl. Res. Labs., Univ. of Texas, Austin, TX 78713-8029)

Techniques for mapping image color to range in passive time-bearing displays were explored. They included (1) a range-focusing method wherein images focused at multiple ranges were combined; and (2) a parallax technique wherein images were formed with different subarray apertures. In both cases, a Hough transform based image processing method was used to enhance features and to compensate for degradation associated with combination of multiple images. Colormapping techniques were employed to render source ranges in color. The presentation will include a description of the methods and a discussion of their effectiveness under various conditions. [Work funded under ARL:UT IR&D program.]

10:35

1aSPa12. Characterization of the left/right bearing ambiguities for passive line arrays. Frank A. Boyle and David E. Grant (Appl. Res. Labs., Univ. of Texas, Austin, TX 78713-8029)

The symmetry of a straight horizontal line array produces a left/right bearing ambiguity that makes it impossible to determine on which side of the array a source is located. A simple technique for resolving left/right ambiguities is to break the symmetry by curving the array. Under certain common conditions, however, significantly curved arrays still possess left/right ambiguities that depend on the source's range and the beamformer's focal length. These ambiguities can be modeled, given a source and array configuration, to produce a prediction of source localization capability. Comparisons with actual data were generally consistent with modeled results in the cases considered. The presentation will include a description of how the ambiguities are modeled, and a comparison between modeled and experimental results. Considerations on how curved arrays can be configured to minimize ambiguities will also be discussed.

Session 1aSPb

Signal Processing in Acoustics: Signal Processing for Airborne Acoustics

Charles F. Gaumond, Chair
Naval Research Laboratory, Code 7142, 4555 Overlook Avenue, SW, Washington, DC 20375-5320

Chair's Introduction—11:00

Contributed Papers

11:05

1aSPb1. Signal processing for acoustic surveillance of the land environment. Brian Ferguson (Defence Sci. and Technol. Organisation, P.O. Box 44, Pyrmont NSW 2009, Australia, Brian.Ferguson@dsto.defence.gov.au)

Surveillance systems that rely on the propagation of sound in the atmosphere are not constrained by electromagnetic line of sight and offer the advantages of being passive, lightweight, compact and covert. Such systems are unattended in their operation and are often deployed in remote areas. To demonstrate the scientific principles that underpin the operation of such systems, various narrow-band and broadband signal processing algorithms are applied to real acoustic data from both single and multiple sensors. In one example, the acoustical Doppler effect enables the flight parameters and blade-passage frequencies of turboprop and rotary wing aircraft to be estimated from the time-frequency signal analysis of spectrogram data. In another example, the acoustical Lloyd's mirror effect is used to extract the flight parameters of a jet aircraft. Other examples deal with the application of generalized cross-correlation processing to the detection, localization and tracking of aircraft, ground vehicles and weapon fire. Wideband cross-correlation with differential Doppler compensation is required for the acoustic tracking of high-speed sources when the sensors are widely separated and phase tranform prefiltering is required to suppress the ambigouous peaks in cross-correlograms caused by the presence of strong narrow-band lines in the source spectra of air and ground vehicles.

11:20

1aSPb2. Finding the right cross-correlation peak for locating sounds in multipath environments with a fourth-moment function. John L. Spiesberger (Dept. of Earth and Environ. Sci., Univ. of Pennsylvania, 240 S. 33rd St., Philadelphia, PA 19104-6316, johnsr@sas.upenn.edu)

To locate calling animals in reverberant environments from recordings on widely separated receivers, a fourth moment ''augmented-template correlation function'' (ATCF) helps identify which of many peaks in each cross-correlation function is that corresponding to the difference in travel times for the first arrivals (reference-lag). This peak may not be the largest. The ATCF, by providing an approximate correlation between auto- and cross-correlation functions, can be orders of magnitude more efficient in selecting the reference-lag than the alternative of randomly selecting peaks. The ATCF's efficacy increases with the number of paths and their signal-to-noise ratios. [Work supported by ONR.]

11:35

1aSPb3. A reconstruction technique for passive acoustic imaging through a turbulent atmosphere. Jonathan W. Benson, George W. Swenson, Jr. (Wave Propagation Lab., Dept. of Elec. and Computer Eng., Univ. of Illinois at Urbana-Champaign, Champaign, IL 61820), and Yoram Bresler (Coordinated Sci. Lab., Univ. of Illinois at Urbana-Champaign, Champaign, IL 61820)

Interferometric passive acoustic imaging through the near-ground atmosphere is hampered by the effects of turbulence. Traditional techniques aimed at removing the effects of the atmosphere, such as self-calibration and phase-closure, cannot be applied in this case where the distortion is a function of the source coordinates. A method was developed to remove the distortion due to the atmosphere within the context of interferometric imaging. A discrete complex-valued distorting screen was used to model the atmosphere. This type of model is routinely used to model propagation through turbulent media and in this case was found to adequately model the distortion effects seen in field data.Reconstruction is accomplished by joint estimation of the source intensity distribution along with the parameters of the distorting screen. Multiple short integration period images or snapshots are used. The parameters of the screen are allowed to change from one snapshot to another, while the source distribution is assumed constant. The Cramer–Rao bounds of the problem were computed and found to be very favorable. The estimation problem was solved using both maximum likelihood and least-squares methods. Near-perfect reconstruction was obtained for simulated data, and significant image quality gains were obtained with experimental data.

11:50

1aSPb4. Performance bounds of atmospheric acoustic arrays operating in a turbulent medium. Sandra L. Collier and D. Keith Wilson (U.S. Army Res. Lab., AMSRL-CI-EP, 2800 Powder Mill Rd., Adelphi, MD 20783-1197)

The performance bounds of an atmospheric acoustic array operating in a turbulent medium with fluctuations described by von Kármán's spectrum are investigated. This treatment considers a single monochromatic source and no multiple-path effects. The Cramer–Rao lower bounds of the azimuthal and zenith angles of arrival (AOAs) are calculated. The normalized range (to wavelength), turbulence parameters, and signal-to-noise ratio are introduced as additional unknowns. It is found that the Cramer–Rao lower bounds of the AOAs increase significantly in strong turbulence at large values of the normalized range. For weak turbulence and moderate values of the normalized range, the signal-to-noise ratio is the dominating factor. For an incident plane wave, the estimates of the AOAs will decouple from the estimates of the other parameters with the appropriate choice of array geometry; however, the estimates of the azimuth and zenith are coupled to one another due to the constraints of the turbulence model.

Session 1aUW

Underwater Acoustics: Scattering

William M. Carey, Chair

Department of Aerospace and Mechanical Engineering, Boston University, 110 Cummington Street, Boston, Massachusetts 02215

Contributed Papers

8:15

1aUW1. Issues of analytical continuation of the field interior to the target surface and a more general T-matrix for submerged fluid targets. M. F. Werby (NRL Code 7180, Stennis Space Center, MS 39529 and Dept. of Phys., Catholic Univ. of America, Washington, DC 20064)

The T-matrix or extended boundary condition method developed by P. Waterman for wave scattering from classical targets and extended by a host of researchers is a marvelous generalization of normal mode solutions allowed for separable geometries to extend to nonseparable geometries by means of mode coupling. The method is based on Huygens' principle and partial wave expansions of physical quantities. In general the expansions don't obey orthogonality and one must take care in the way expansions are made. One must never take expansions of a quantity and then represent expansions of their differential forms in terms of the partial wave differential forms using the same expansion coefficients lest one violate a mathematical principle (the Rayleigh hypothesis). One may circumvent any issues without resort to analytical continuation of interior solutions by including all solutions subject to interior and exterior Greens functions. This leads to matrix relations (constraining matrices) between the expansion terms which when used lead to new forms of the fluid T-matrix, for example. The new form differs from Waterman's T-matrix for the case when the constraining matrices are not symmetric. Otherwise the original expression by Waterman is valid. These issues are discussed and examples are presented.

8:30

1aUW2. Backscattering from empty and water-filled tilted cylindrical shells due to leaky helical flexural waves: Comparison and ray theory. Florian J. Blonigen and Philip L. Marston (Dept. of Phys., Washington State Univ., Pullman, WA 99164-2814)

For tilt angles smaller than the meridional ray coupling condition previously investigated [S. F. Morse *et al.*, J. Acoust. Soc. Am. **103**, 785–794 (1998)], helical rays on empty and water-filled steel shells can significantly enhance the backscattering. These contributions are compared and modeled in the present work. Such contributions in the water-filled case are found to be weaker in amplitude, which is to be expected since additional energy is lost through radiation leaked to the inner fluid. A modified ray theory taking into account the increased radiation damping was compared with experimental results for the earliest helical wave arrival, yielding satisfactory agreement. However, the energy lost to the interior of the shell may be at least partially recovered, as internal rays can excite additional leaky waves on the shell. Contributions from rays with one internal chord inside the cylinder were superposed with those of the first helical wave, and the resulting backscattering amplitude was nearly that of the air-filled case for the first helical wave alone. This shows that we may regard the ray theory for the empty shell as an upper bound for corresponding amplitudes observed in the water-filled case. [Work supported by ONR.]

8:45

1aUW3. Wavefront scatter and wavefront healing. C. Feuillade (Naval Res. Lab., Stennis Space Center, MS 39529-5004, cf@nrlssc.navy.mil)

While acoustic scattering is essentially a time domain phenomenon, studies of scattering from objects in water, and from boundaries, have often concentrated on time-independent/cw descriptions. There are, however, critical aspects of the scattering problem which become evident only in the time domain. In this paper, time-domain visualizations of impulse scattering from rigid, pressure-release, fluid, and elastic spheres are presented. The results provide important insights into the physical processes which lead to the construction of scattered wavefronts. Using this method, it is possible to time-separate and identify different scattering mechanisms: geometrical scatter; diffraction (which, for rigid spheres, gives rise to the circumferential wave); refraction; and wave conversion in the body of the scattering obstacle. Of particular interest is the characterization of the forward scatter region, where the process of wavefront healing can be directly observed. The presentation will be illustrated by movie animations. [Work supported by ONR.]

9:00

1aUW4. Shape sensitivity calculations for exterior acoustics problems. Gonzalo R. Feijoo and Peter M. Pinsky (Div. of Mech. and Computation, Stanford Univ., Stanford, CA 94305-4040, grfeij@stanford.edu)

We present a method to calculate the derivative of a functional determined by the shape of an object. This functional depends on the solution of a linear acoustic problem posed in an unbounded domain. We rewrote this problem in terms of another one posed in a bounded domain using the Dirichlet-to-Neumann (DtN) map, or the modified DtN map. Using a classical method in shape sensitivity analysis, called the adjoint method, we are able to calculate the derivative of the functional using the solution of an auxiliary problem. This method is particularly efficient because the cost of calculating the derivatives is independent of the number of parameters used to approximate the shape of the domain. The resulting variational problems are discretized using the finite-element method and solved using an efficient Krylov-subspace iterative scheme. Numerical examples that illustrate the efficacy of our approach are presented.

9:15

1aUW5. A two-way coupled mode method, matrix methods, and bandedness of the integration with free scattering solutions and the solution of imbedded objects in a waveguide. M. F. Werby (NRL Code 7180, Stennis Space Center, MS 39529 and Dept. of Phys., Catholic Univ. of America, Washington, DC 20064)

Since the seminal work of Richard Evans in 1983 when he proved to the world the viability of a two-wave general waveguide solution in coupled mode form his method and implementation has been the gold standard of other range-dependent approximations. The method by its very nature is tedious to implement and requires much time in computation.

Often a one-way solution is adequate allowing many to circumvent this more complicated solution. A one-way coupled mode method has been developed at NRL that takes advantage of representation theory in terms of a basis set that leads to analytical forms suitable for use in representing incident fields as input for scattering problems. It is apparent that in many cases the addition of an inclusion renders coupling in both directions in a waveguide and thus the one-way approach is extended to a two-way coupled mode solution and by extension to a generalization of the integrated solution of an object in a waveguide for more general inclusions.

9:30–9:45 Break

9:45

1aUW6. Scattering measurements and interface model validation for a rocky bottom at 2–4 kHz. Raymond J. Soukup (Naval Res. Lab., Code 7144, 4555 Overlook Ave. SW, Washington, DC 20375)

Validation of interface scattering models for the ocean bottom, such as the small slope approximation model of Gragg and Wurmser [J. Acoust. Soc. Am. **109** (2001)] requires acoustic scattering data covering a wide range of grazing angles with minimal sub-bottom penetration. Such data measurements were performed in the frequency band 2–4 kHz over an exposed limestone bottom off the Carolina coast (120 to 310 m water depths) during the Littoral Warfare Advanced Development Focused Technology Experiment 96-2 (LWAD FTE 96-2). Direct-path bottom scattering strengths were obtained at grazing angles ranging from 8 to 75 degrees using data fusion from multiple experimental geometries coupled with careful signal processing. This data processing includes corrections for the surface-reflected and other hybrid paths, and the rapid reverberation decay observed over the length of the pulses at higher grazing angles. The resulting frequency and grazing-angle dependencies exhibit trends consistent with those predicted by the elastic small-slope approximation model of Gragg and Wurmser. [Work supported by ONR.]

10:00

1aUW7. Theory and computation for small-slope scattering from rough elastic ocean floors. Robert F. Gragg and Daniel Wurmser (Naval Res. Lab., Code 7144, Washington, DC 20375-5350)

A bistatic theory is presented for the scattering strength of a random rough interface bounding a uniform medium (either fluid or elastic solid) for cases where the Bragg scale lies within the power-law tail of the roughness spectrum. The physical basis is an inherently reciprocity-preserving, local small-slope theory. A robust numerical implementation is also presented that functions smoothly throughout a wide range of spectral exponents (from 2.4–3.9, at least). The practical result is a significantly improved formulation for the interface component of sea floor scattering. Results are compared to those of perturbation theory, and illustrate the dependence on frequency, environmental parameters, and bistatic geometry. Validation of the model using at-sea data is addressed by Soukup [R. J. Soukup, J. Acoust. Soc. Am. (submitted)]. [Work supported by ONR.]

10:15

1aUW8. A new strategy for applying the parabolic equation to a penetrable rough surface. Daniel Wurmser (Naval Res. Lab., Code 7144, Washington, DC 20375-5350)

The range-dependent parabolic equation (PE) propagates an auxiliary field that roughly corresponds to the square root of the downrange flux. At the ocean bottom, the interface is broken into stair steps, and this auxiliary field is conserved at the vertical interfaces. Because of its success in matching benchmark solutions, this method is widely used to model standard propagation problems in underwater acoustics. However, when the interface is rough at the wavelength scale, the physics of the problem changes, and the standard approach ceases to be practicable. This talk presents a new approach specifically designed to handle this important problem. The Foldy–Wouthuysen transformation generates the PE for this auxiliary field, complete with a full suite of boundary conditions. In the discrete problem, the boundary conditions are used to evaluate the Hamiltonian on the interface, and the result is used to perform the next downrange step. At lowest order, the new theory streamlines the stair step method, and frees it from *ad hoc* impositions such as those spawned by the density jump at the ocean bottom. The approach also systematically provides higher-order corrections. [Work supported by ONR.]

10:30

1aUW9. Effect of bottom ripple on detection of a buried object using shallow-grazing-angle underwater sonar. Raymond Lim (Coastal Systems Station, Code R22, 6703 W. Hwy. 98, Panama City, FL 32407)

Sound penetration into the seafloor at shallow (subcritical) sonar grazing angles has been a topic of much discussion lately because measured levels can be much higher than expected based on simple propagation models assuming a flat seafloor. A popular mechanism for explaining this penetration is based on diffraction of energy into the bottom by seafloor roughness. For objects buried in the bottom, this holds the promise for enhanced long range detection with sonar as long as the backscatter by the roughness itself does not overpower the echo from the object. Here the detectability of a sphere buried under roughness in the form of anisotropic wave-induced ripple is investigated by modifying a scattering solution that combines a stochastic perturbative description of the roughness with a deterministic description of the sphere response. [See Lim *et al.*, J. Acoust. Soc. Am. **107**, 1246–1262 (2000).] Comparison with predictions for an isotropic, power-law seafloor roughness at a given rms amplitude demonstrates greater high frequency detection capability with rippled seafloors over a wide range of illumination angles relative to the ripple direction. Comparisons with experimental data are also discussed. [Work supported by ONR.]

10:45

1aUW10. Scattering from stratified sediments. Anatoliy N. Ivakin (On leave from Andreev Acoust. Inst., Svernika 4, Moscow 117036, Russia, anivakin@yahoo.com)

There is a special class of volume perturbations, which exist within stratified media and can be described in terms of displacement of isosurfaces of these parameters. Scattering from marine sediments with such perturbations is considered using a unified approach to volume and roughness scattering [A. Ivakin, J. Acoust. Soc. Am. **103**, 827 (1998)]. Discrete stratification (layering) is treated as a particular case of a continuous one, where interfaces between layers are replaced by additional (transition) layers with acoustical parameters smoothly dependent on the depth. In the unperturbed medium, the transition layers are flat, which gives a zeroth-order solution. In the first-order approximation, volume perturbations of the aforementioned class are proportional to vertical gradients of corresponding acoustical parameters in the unperturbed medium and to displacement (roughness) of the isosurfaces of these parameters within transition layers. Note that there are only smooth continuous volume inhomogeneities in these layers, that is, there is no real roughness scattering, which is controlled by discontinuity of the acoustical parameters at the interfaces. A new general model is presented, which gives the volume scattering strength as a function of the wave thickness of the transition layers. The model does not require the vertical correlation sclae of inhomogeneities to be small in comparison with the layer thickness. In the case of low frequencies, where the transition layers are thin in comparison with the wavelength, the model provides an exact transition to existing results

for scattering from rough interfaces and, thus, naturally links volume and roughness mechanisms of seabed scattering at low and high frequencies within a unified approach. [Work supported by ONR.]

11:00

1aUW11. The contributors of sound scattering in stratified flows. Victor Prokhorov (Inst. for Problems in Mech. of the RAS, 101(1) prospekt Vernadskogo, Moscow 117526, Russia)

It is generally assumed that marine biomass dominates the backscattering everywhere in the ocean. However, stratified flows form both aggregated and solitary scatterers whose scattering levels are close to those of biologicals. The present work deals with theoretical and laboratory study of scattering phenomena in stratified flows of different types and regimes. Microscale clusters, flat and curved discontinuities, and solitary vortex rings were modeled and examined with both monostatic and bistatic acoustic schemes. Simultaneously they were optically visualized and photographed to provide their direct identification and computation of spatial spectra, and to extract characteristic scales being responsible for scattering. The average scattering levels converted to dimensionless backscattering cross section are of the same order or above those obtained in field measurements. Theoretical study seeks to reveal the mechanisms producing acoustic contrasts and to examine temporal characteristics of scattering. The scattering dynamics essentially depends on whether salinity or temperature stratification dominates the condition under which the scattering inhomogeneities arise. Also, attached internal waves contribution to produce the scattering interfaces was theoretically and experimentally confirmed. Shadowgraphs of scattering structures followed by synchronous scattering profiles, spatial and temporal characteristics of scattering, data of theoretical modeling are presented.

11:15

1aUW12. Acoustic scattering from bubbly cylinders and spheres. Preston S. Wilson, Ryan D. McCormick, Ronald A. Roy, and William M. Carey (Dept. of Aerosp. and Mech. Eng., Boston Univ., Boston, MA 02215)

This presentation deals with acoustic scattering from bubble clouds. In the most general manifestation, the cloud has arbitrary shape and extent and consists of a population of bubbles with arbitrary size distribution; the ensuing scattering problem is considerable. In the asymptotic limit of an acoustically compact cloud consisting of acoustically small bubbles (both conditions are satisfied at sufficiently low frequencies) the situation condenses down to the much simpler problem of scattering from a uniform compressible object, and to lowest order in the Helmholtz number (ka) this object is a sphere. Experimental evidence supporting this approximation was obtained at Lake Seneca [Roy *et al.*, J. Acoust. Soc. Am. **92**, 2993 (1992)], as well as numerous laboratory noise and scattering experiments that followed. We present additional data that supports this view, and extends it to higher order in ka for the case of scattering from a cylindrical bubble cloud. Data obtained at the NUWC underwater test facility details the frequency-dependent scattering from well-characterized bubbly cylinders, and a good match to data is obtained using an effective medium model for the uniform compressible cylinder. Accompanying scattering data from two types of bubbly spheres (voided polyurethane, gel-filled) are also presented. [Work supported by ONR.]

11:30

1aUW13. Measurements of the acoustic properties of bubbly liquids and partially voided polyurethane solids in the resonance regime. Preston S. Wilson, Ronald A. Roy, and William M. Carey (Dept. of Aerosp. and Mech. Eng., Boston Univ., Boston, MA 02215)

Complete understanding of sea-surface scattering, shallow water volume scattering, and bottom interaction requires the understanding of frequency-dependent acoustic propagation in bubbly media. Because of the high attenuation encountered at and near the range of resonance frequencies associated with bubble size distribution, experimental measurements in this regime are difficult, and for many specific cases, have not been obtained. Using an impedance tube method similar to one previously described [J. Acoust. Soc. Am. **106**, 2291(A) (1999)], measurements of the acoustic impedance of bubbly fluids in the vicinity of individual bubble resonance have been obtained. The measurement procedure will be briefly summarized, and contrasted with the previously reported method. Results obtained over a range of frequencies and void fractions with known size distributions will be presented and compared to existing theory. In addition, the acoustic behavior of partially voided polyurethane spheres and plates has been studied using the present impedance tube method. Evidence for collective oscillation of these structures will be shown. [Work supported by ONR.]

11:45

1aUW14. Linear acoustics in bubbly liquids from an effective medium theory. Steven G. Kargl (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105)

Proposed corrections to the lowest-order approximation of an effective wave number obtained from Foldy's exact multiple scattering theory [Foldy, Phys. Rev. **67**, 107 (1945)] has sparked renewed interest in linear wave propagation through bubbly liquids. An alternative approach to wave propagation through a bubbly liquid reduces the governing equations for a two-phase medium to an effective medium. Commander and Prosperetti [J. Acoust. Soc. Am. **85**, 732 (1989)], based on this method, derive an expression for the lowest-order approximation to an effective wave number. At this level of approximation the bubbles interact with the mean acoustic field without higher-order rescattering. That is, the field scattered from a bubble may interact with one or more new bubbles in the distribution, but a portion of that scattered field may not be scattered back to any previous bubble. Simple modifications to the results of Commander and Prosperetti lead to a new expression for the effective wave number, which properly accounts for all higher orders of multiple scattering. [Work supported by ONR, Code 321OA.]

Session 1pAA

Architectural Acoustics: The First 80 Milliseconds in Auditoria

U. Peter Svensson, Cochair
Department of Telecommunications, Norwegian University of Science and Technology, Trondheim N-7491, Norway

Jerald R. Hyde, Cochair
Box 55, Helena, California 94574

Chair's Introduction—1:00

Invited Papers

1:05

1pAA1. The importance of the first 80 ms of room impulse responses: A historical review. John Bradley (Natl. Res. Council, Montreal Rd., Ottawa K1A 0R6, Canada)

The direct sound and reflections arriving within 80 ms after it are particularly important to the perceived quality of sound in rooms such as concert halls and auditoria and have been the focus of the majority of research on concert hall acoustics over the past 50 years. Most of the sound energy is usually concentrated in this time interval and this early sound energy determines many of the perceived characteristics of the hall. Although some remarkable early experiments date back to the 1850s, the major advances concerning how these early reflections are perceived have occurred more recently. Monaural effects including explanations of how our hearing system integrates early reflections, the precedence effect, and the development of measures of clarity and intelligibility have developed from initial work in the 1950s. A little more recently the binaural effects leading to perceived spatial impression have been explained and shown to be critical to achieving good concert halls. Physical measures have been developed that relate to both the monaural and binaural subjective phenomena. This work will review the development of our understanding of the importance of early arriving sound on the acoustical characteristics of concert halls and auditoria.

1:30

1pAA2. Early reflection energy in concert halls: how much, how early, and from where. Anders Chr. Gade (Section of Acoust. Tech., Oersted DTU, Tech. Univ. of Denmark, Bldg. 352, DK 2800 Lyngby, Denmark, gade@dat.dtu.dk)

Today, the importance of distributing early reflection energy to listeners and performers in concert halls is well understood and accepted—also among architects. Still, implementation in the practical design of a large hall is not easy, partly because we still have difficulties quantifying precisely the demands in detail (how much, how early, and from where) due to limitations in our acoustic knowledge—and often also due to little response from the clients regarding their preferences when decisions between alternative demands must be made. Therefore, in general it is up to the acoustician to decide how far to promote the good cause on the basis of his/her experience, taste, and talent in influencing the decision process. The aural presentation will focus on the current limitations in our knowledge regarding the musicians' need for early reflections, which is a special challenge in the design of terraced, ''surround'' concert halls. In the case of the listeners, knowledge has reached a much higher level, and recent results regarding how room shape influence the objective acoustic parameters monitoring the early energy will be presented.

1:55

1pAA3. ''REC Curve'' data measured in halls opened after 1988. Yasuhisa Toyota, Keiji Oguchi, and Minoru Nagata (Nagata Acoustics, Inc., 201 Ocean Ave., Ste. 1205B, Santa Monica, CA 90402, toyota@nagata.co.jp)

As the results of the studies on early reflections in auditoriums, ''Reflection Energy Cumulative Curve excluding direct sound'' (REC Curve) was introduced and defined at the ASA Hawaii meeting in 1988. In the study, early reflections up to around 80 ms after direct sound were found to be substantially important and influenced by the room shape of the auditorium. Some concert halls were designed based on ''REC Curve'' so that REC Curve up to 80 ms showed smoothly increasing curves in the auditoriums. ''REC Curve'' data measured in the halls opened after 1988 are reported in this paper with some subjective comments on the halls.

2:20

1pAA4. The nervous spatial behavior of early-energy measures in auditoria. Diemer de Vries and Edo M. Hulsebos (Lab. of Acoust. Imaging and Sound Control, Delft Univ. of Technol., P.O. Box 5046, 2600 GA Delft, The Netherlands, diemer@akst.tn.tudelft.nl)

Most commonly used room acoustical measures are defined as the ratio between the energies in two time windows of the impulse response. In the Clarity Index and the Lateral Energy Fraction—supposed to be indicators for transparancy and spaciousness, respectively—the time window 0–80 ms plays a predominant role. When impulse responses are recorded along an array of closely spaced microphone positions and the above measures are calculated for each position, it appears that large fluctuations, up to a factor

of 2, occur on a small spatial scale, e.g., in front of one and the same seat. These fluctuations are due to interference of early relections and can be physically well understood. It makes clear that the acoustic quality of a seat or row should not be based on one single measurement but on some spatial average, or that the measures should be redefined such that interference, to which our hearing mechanism apparently is not sensitive, is eliminated.

2:45

1pAA5. Edge diffraction and scattering in the early room impulse response. Rendell R. Torres, Mendel Kleiner (Chalmers Rm. Acoust. Group, Chalmers Univ. of Technol., SE-412 96 Gothenburg, Sweden), and U. Peter Svensson (Dept. of Telecomm., Acoust. Group, Norwegian Inst. of Technol., NO-7491 Trondheim, Norway)

Although reverberation is a common and important measure of quality in architectural acoustics, the early part of a room's impulse response is equally critical, as its spectrum and transient structure predominantly determine the initial filtering (i.e., the natural coloration) that a room endows on an (otherwise anechoic) source signal. Since this initial room filtering could make the difference between an acoustical impression that is, for example, either ''clear and defined'' or ''murky and unfocused,'' various studies have been performed to better understand and model the effects of edge diffraction and surface scattering on the early reflections in room impulse responses. These studies include auralizations with varying ''diffusion'' coefficients, computations and measurements of edge diffraction in early room impulse reponses, and application of ''boss models'' to compute hemispherical scattering in room acoustics auralization. Various results show that these effects can be clearly audible and that their audibility and perceived character depend on the input signal, among other factors. Conclusions are complemented with suggestions for future work. [Work supported by the Axson Johnson Foundation and Teknikbrostiftelsen, Sweden.]

3:10–3:25 Break

3:25

1pAA6. Effects of seating arrays on subjective quality in auditoria. Masayuki Morimoto, Sachio Hirose (Environ. Acoust. Lab., Kobe Univ., Nada, Kobe 657-8501, Japan), and Koji Ishida (Onosokki Co., Ltd., Midori, Yokohama 226-8507, Japan)

Listening tests were performed to clarify the subjective effects of seating arrays. First, a qualitative assessment was made of the subjective effects introduced by the seating array. Test signals were prepared by convolving dry music with impulse responses measured in a scale model consisting of a stage, a floor with audience but no walls or ceiling. The measured impulse responses showed that the seating array caused an increase in amplitude below 80 Hz as well as a large dip at 150–200 Hz. The results of listening tests demonstrated that the seating array affected not only bass perception but also loudness. Second, it was investigated whether reflections from the ceiling and walls can compensate for an excess or deficiency of perceived bass caused by the seating array. In addition, the relation between the frequency characteristics of sound fields and bass perception is discussed. Test signals were prepared by adding a ceiling reflection and reverberation to test signals used in the first test. The results indicated that to a certain extent they can compensate for an excess or deficiency of perceived bass. It is suggested that bass perception should be evaluated according to the frequency characteristics of sound fields within the first 100 ms or so after the direct sound.

3:50

1pAA7. Sound coloration from (very) early reflections. Tor Halmrast (Statsbygg, P.O. Box 8106, Dept. N-0032, Oslo, Norway)

Coloration is defined as changes in Timbre/Klangfarbe. Adding a reflection automatically changes the frequency response of a signal, giving some kind of coloration. This might be looked upon as distortion. However, reflections have been a part of sound distribution since the Greek amphitheatres, indicating that some coloration must be acceptable, even wanted. The question is: Which reflections give disturbing/unwanted coloration? Most room-acoustic criteria assume that everything happening before/after a certain time is supposed to be good/bad (e.g., 50 ms/80 ms for speech/music). We need to take a closer look at the distribution of reflections within these time intervals, to investigate coloration. This paper will give results from measurements in concert halls (Oslo, Munich, Vienna), and opera houses, and compare them with psychoacoustic studies on coloration. It is shown that measurements should be done with the orchestra present on the stage/in the orchestra pit (TOR, through orchestra impulse response). Lonesome/strong reflections with a time delay about 10 ms will give box-klang-coloration. On the other side, colorating reflections might be useful and fun, used in the right way, to support bass instruments, etc. [J. Sound Vib. **232** (2000)].

Contributed Papers

4:15

1pAA8. An investigation of the spectral effect of multiple early reflections. Alexander U. Case (Sonics in Architecture, Rensselaer Polytechnic Inst. and Cavanaugh Tocci Assoc., Consultants in Acoustics, 327F Boston Post Rd., Sudbury, MA 01776, acase@cavtocci.com)

The interaction of a direct sound with a single reflection of short delay is well-understood. Comb filtering is a quantifiable, measurable, inevitable result. In actual auditoria, one rarely hears a single early reflection. Typically a volley of several reflections follows the direct sound. This paper presents an analytical evaluation of the change in frequency response due to, not one, but several reflections. The combination of a direct sound with several early reflections does not result in a more pronounced, more audible comb filter quality. Rather, the interaction of the direct sound with each of the several reflections in combination with the interaction among the many reflections themselves tends to average out or randomize the expected, comb filtered, spectral distortion.

4:30

1pAA9. When all sound is mere reflections. Richard H. Talaske (The Talaske Group, Inc., 105 N. Oak Park Ave., Oak Park, IL 60301)

For many decades in the recent past, impulse response measurements have been performed in music and drama spaces to assess the clarity of sound and ease of communication offered on a stage or within the audience chamber. Much of the focus has been on the strength and timing of early sound reflections arriving within 80 or so ms following the direct sound. Generally, the thought and design process revolve around an un-

occluded, straight-line path between source and receiver. This paper, however, presents the results of numerous studies and completed buildings in which excellent communication has been achieved in drama and music conditions where the direct sound is attenuated due to source directivity factors and/or obstructions normally encountered within rooms. Results of computer model studies, photos of completed spaces and acoustic data will be presented along with general conclusions.

4:45

1pAA10. The acoustics of the church San Francisco of Santiago of Chile. Jaime Delannoy (Dept. of Acoust., Tech. Univ. Vicente Perez Rosales, Brown North 290 Nunoa, Santiago, Chile, jaime.delannoy@uvipro.cl)

The church San Francisco, declared a national monument, is the most important representative of the colonial architecture that remains in Chile. The definitive construction data of 1618, and after supporting to earthquakes and fires, conserves its original central shape to the present time. The interest of diverse musicians to interpret its works in this enclosure was the reason for the accomplishment of this study. Measurements of the following objective parameters were made: EDT, C80, C50, D50, U50, ti, and NC. The results indicate that a high background noise exists (NC=46) and a poor intelligibility (D50 lower 0.34). Nevertheless its inner distribution provides an initial favorably short time delay gap (ti=22 ms) that provides an acoustic privacy that favors the execution of small orchestras. [Work supported by Department of Acoustics–Technical University Vicente Perez Rosales.]

5:00

1pAA11. The acoustic design of the Concertgebouw, Amsterdam, and resolution of the hall's early acoustical problems. Pamela Clements (Kirkegaard Assoc., 801 W. Adams St., Chicago, IL 60607)

From the outset, acoustical considerations were a driving force in the design of the Concertgebouw. Several halls were used as acoustical models, including the Tonhalle in Düsseldorf, the Leipzig Gewandhaus, and the concert room of the Felx Meritus Society in Amsterdam. When the Concertgebouw opened in 1888, however, it had acoustical problems: the newly formed orchestra struggled with the hall's long reverberation time, and balance between brass and strings was difficult. Various remedies were tried, including potted plants to ''soak up'' the sound. The installation of the organ in 1891 was thought by some to make the acoustics worse. Between 1895 and 1900 the problems were gradually resolved. Orchestral balance was improved in 1899 when the orchestra platform was rebuilt to reduce the steepness of the risers. Other factors included the approach of the new conductor, Willem Mengelberg, improved performance standards, and the arrival of the late Romantics, particularly Tchaikovsky, in the repertoire. Happily this coincided with improved transportation that enabled newly enthusiastic audiences to fill the hall. The paper uses plans, photographs and other historical evidence to explore the intent of the acoustic design of the Concertgebouw, and its transformation from near failure to musical and acoustical success.

MONDAY AFTERNOON, 4 JUNE 2001 ADAMS ROOM, 1:30 TO 4:55 P.M.

Session 1pAO

Acoustical Oceanography: Acoustical Instrumentation for Water Column Measurements II

Kenneth G. Foote, Chair

Department of Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

Chair's Introduction—1:30

Invited Papers

1:35

1pAO1. Wideband and correlation techniques and their application to fisheries acoustics: Existing prototypes and future trends. Manell E. Zakharia (Ecole Navale, French Naval Acad., Batiment Bougainville, 29240, Brest Naval, France, zakharia@ecole-navale.fr)

The aim of this work is to describe several ''research sonar'' prototypes (one-of-a-kind), to show the performance improvement by using both wideband techniques and associated echo processing procedures such as correlation, synthetic aperture, or spectral signature analysis. We will describe the improvement of fisheries sounders provided by widening the frequency range: enhanced axial resolution, smoothed spatial and angular variations. The advantages will first be illustrated by experiments on individual fish. Then, the performance of such systems will be investigated at sea using wideband prototypes. A wide range of processing algorithms will then be detailed: pulse compression, dynamic focusing, beam control on a wide frequency range, synthetic aperture, echo to echo information, etc. Some of them have been commonly used in military sonar systems, some others are still under exploration in other areas (echo–echo correlation, synthetic aperture processing of sidescan data). A multibeam approach based on wideband correlation will also be described. Finally, the performance of wideband classification of fish species at sea will be shown. The technical limitations (transducers, acoustic compatibility, real-time processing, etc.) that could limit the development of such systems will be developed and several solutions will be proposed to overcome the development complexity.

1pAO2. A synopsis of state-of-the-art high-frequency sonars for water column measurements. Mark V. Trevorrow (Defence Res. Establishment Atlantic, P.O. Box 1012, Dartmouth, NS B2Y 3Z7, Canada, Mark.Trevorrow@drea.dnd.ca)

In recent years there has been a dramatic surge in the use of high-frequency (10 kHz–2 MHz) backscatter sonar systems for oceanographic and bioacoustic measurements in oceans, lakes, and rivers. These applications have generally focused on measuring the distribution of scatterers, such as bubbles, particulates, microstructure, fish, and zooplankton. Additionally, oceanographic flows and wave motions can be illuminated using these scatterers as tracers, using both the intensity and Doppler signals. Advances in electronics, data storage, signal processing, and transducer materials technology have made, and will continue to make, HF sonars smaller yet more capable in terms of bandwidth, resolution, and endurance. Some novel applications of simple echo sounders and sidescan sonars will be reviewed, showing examples of near-surface bubble distributions, internal hydraulic flows, and fish and zooplankton detection. Newer developments in multifrequency, volumetric multibeam and parametric sonars will be discussed.

3:05–3:25 Break

Contributed Papers

3:25

1pAO3. Expanding capabilities of acoustic instruments: Benefits of calibrated output signals, bandwidth, midwater signals, and dynamic range. Kenneth G. Foote (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

Echo sounders, multibeam sonars, acoustic Dopper current profilers (ADCPs), and other acoustic instruments are widely used, if often for quite specific applications. Sometimes, these applications have been pursued at the design stage to the detriment of general system performance and versatility. Availability of a calibrated output signal in a fishery echo sounder, for example, may enable this to be used in quantitative applications. The availability of bandwidth may enable scientific echo sounders to classify and quantify multispecies assemblages of zooplankton. The provision of midwater signals may enable multibeam bathymetric sonars to be used for imaging scatterers in the water column. Extension of the dynamic range of ADCPs may enable a single unit to be used as four synchronous echo sounders with noncollinear beams. In general, the capacities of most underwater acoustic systems can be expanded significantly by attention to issues such as those mentioned here. Resulting opportunities for new applications are described, with justification in the physics of scattering. [Work supported by the Alfred P. Sloan Foundation.]

3:40

1pAO4. Multibeam SONAR from the bottom and up. Jeff Condiotty (SIMRAD, Inc., 19210 33rd Ave. W., Lynnwood, WA 98036) and Oywind (Irv) Bjorkheim (Kongsberg Simrad, Inc., Lynnwood, WA 98036)

Acoustic multibeam technology has traditionally been used for hydrographic surveying but is now being introduced as a useful tool in data collection within the water column. A new PC-based system operating at frequencies of 90 or 200 kHz allows visualizing targets within the water column for fisheries behavioral and enumeration studies as well as bottom characterization and profiling, at the same time. The system takes advantage of a wide sampling swath, up to 150 with up to 128 individual beams. The system can be operated in a wide beam mode or used in a Mill's Cross mode to control the beamwidth in two dimensions. Data are displayed as a SONAR image in real time while raw data (pre-beamformed or beamformed) can be stored for further processing utilizing commercial software. The transducer can be either hull mounted, installed on a submersed vehicle, or attached to a pole mount on a small boat. Applications and data collection methodology to visual and measure target reflectors will be discussed along with calibration of individual beams to produce quantitative results.

3:55

1pAO5. Target strength analysis in high fish densities using a probing split beam transducer. Egil Ona and Ingvald Svellingen (Inst. of Marine Res., P.O. Box 1870, 5024 Bergen, Norway)

Unbiased measurements of the mean target strength are needed if the absolute fish abundance is to be measured. For most pelagic fish, this also requires direct target strength techniques to work in high-density situations. A probing, pressure stabilized split beam transducer has been lowered from the research vessel directly into dense herring layers in order to resolve a single target at short range. Specialized hardware has been used for single fish target tracking, from which data on target strength, target directivity, swimming speed and tilt angle are derived. The paper will focus particularly on calibration procedures for transducer depth performance, target-tracking methodology and on the monitoring of essential transducer movement parameters.

4:10

1pAO6. Instantaneous target strength measurement of live squid with synchronized video image. Tohru Mukai and Kohji Iida (Grad. School of Fisheries Sci., Hokkaido Univ., 3-1-1 Minato-cho, Hakodate, Hokkaido 041-8611, Japan, mukai@fish.hokudai.ac.jp)

The TS is a complex function of the fish's size, shape, orientation, behavior, and internal structure, as well as of the acoustic wavelength employed. Therefore, for practical purposes, it is necessary to measure the TS experimentally. There is a considerable amount of knowledge about the TS of fish. Information about the TS of squid, however, is much less. Through previous studies over the past 6 years at our experimental field, acoustic equipment has been modified and gradually stabilized. It was also possible to use live squid in our specially designed experimental barge equipped with two underwater video cameras which look at the side view and upward. Preliminary TS and orientation measurements were done on single tethered live squid *Todarodes pacificus*. The live squid was suspended with monofilament nylon string near the acoustic beam axis and positioned 3.5 m below the transducers. The echo signals of the squid at six frequencies were converted to amplitude modulated signals in audible frequency, then they were recorded on audio channels of video tape with a video image simultaneously. Afterward, the echo envelopes were demodulated to analyze the relationship between TS and the swimming aspect of live squid by synchronized video image. Preliminary results from this measurement will be presented.

4:25–4:55

Panel Discussion

Session 1pEA

Engineering Acoustics: Hydrodynamics in Ducts, Jets, and Blowers

Daniel M. Warren, Chair
Knowles Electronics, 1151 Maplewood Drive, Itasca, Illinois 60143

Chair's Introduction—2:00

Contributed Papers

2:05

1pEA1. Determination of fan inlet duct mode structure using in-duct microphone array measurement. Carl H. Gerhold (MS 461, NASA Langley Res. Ctr., Hampton, VA 23668, c.h.gerhold@larc.nasa.gov)

A methodology is developed in which the amplitudes of propagating modes are extracted from the sound measured in the duct. The method is derived from a beamforming technique in which the weighting function is the wave description of a chosen mode. The modal distribution that arises from the analysis is used to synthesize the sound propagating in the duct which is projected to the far field to estimate the sound radiated from the duct inlet. The methodology is tested using an array of dynamic pressure transducers in the inlet duct wall of a turbofan engine. The data show that the modal structure of the blade passage frequency tone at lower fan speeds is predominately from the interaction of the rotor with downstream obstructions. At higher speeds, when the fan blade tip speed becomes supersonic, many modes are excited. Good agreement is found between measured and estimated far field sound levels in the forward quadrant of the engine at the lower engine speeds. Significant divergence between measured and calculated far field sound occurs when the blade tip speed becomes supersonic. The methodology is a useful identification of the noise generation mechanisms present in the fan.

2:20

1pEA2. Proof of existence of monopole-type sound generation mechanism in a centrifugal blower. Ashnapreet Nagi and Sean Wu (Dept. of Mech. Eng., Wayne State Univ., 5050 Anthony Wayne Dr., Detroit, MI 48202)

Previous studies have shown that sound radiation from a centrifugal blower is primarily of the dipole and quadrupole characteristics, which describe the effects of fluid-structure interaction and turbulence generated by rotating blades. However, in many engineering applications the sound powers from centrifugal blowers are found to be proportional to the speed to the power of 4–6, which implies the existence of a monopole-type sound according to the Lighthills acoustic analogy [Lighthill, Proc. R. Soc. London, Ser. A **222**, 564–587 (1952)]. This paper demonstrates that such a monopole sound generation mechanism indeed exists in a centrifugal blower used in an HVAC unit of a passenger vehicle. LDV measurements show a velocity fluctuation at the trailing edge tangential to the blade surface of an impeller. Laser tachometer measurements further indicate a fluctuation in shaft speed for all models of blowers tested. The resulting sound powers from these blowers show a predominant dependence of the speed to the power of 4, indicating the characteristic of a monopole. When these blowers are installed on a nonfluctuating shaft, however, the flow fluctuations are significantly reduced and the resulting sound powers are found to be proportional to the speed to the power of 6, which implies that the monopole sound generation mechanism is effectively eliminated.

2:35

1pEA3. Radial temperature variations in cylindrical waveguides and implications for flow measurement. Morten Willatzen (Mads Clausen Inst., Univ. of Southern Denmark, Grundtvigs Alle 150, DK-6400 Sonderborg, Denmark)

A quantitative treatment of radial temperature variations in a cylindrical duct on ultrasonic flow meter performance is discussed in the laminar flow regime. First, based on the continuity equation, the Navier–Stokes equations, and an energy equation including loss mechanisms due to heat conduction and viscous effects, the steady-state temperature and flow spatial distributions are determined in two cases of practical interest: (a) cylinder wall temperature is maintained at a constant value, and (b) cylinder wall temperature decreases linearly with cylinder axial distance. It is shown that while radial temperature variations are insignificant in case (a), radial temperature gradients as large as 100–200 K/m are possible in case (b) for a fixed axial temperature gradient decrease of 0.1 K/m. Such strong temperature gradients have strong and unfortunate consequences for flow measurement. Large flow meter errors—up to several percentages—are possible using typical parameter values for water as medium. Finally, it is shown that effective ways exist such as to diminish the influence of temperature gradients on flow meter performance. Besides the obvious choice of insulating the flow meter tube, flow measurement errors due to radial temperature variations can be effectively suppressed by reducing the cylinder radius and/or ultrasound frequency.

2:50

1pEA4. Computational investigation of acoustic streaming in a two-dimensional duct. W. R. Erskine and A. Selamet (Dept. of Mech. Eng., Ohio State Univ., Columbus, OH 43210)

The present work investigates the nonlinear phenomenon of Rayleigh streaming characterized by acoustically driven flows attenuated by wall friction, resulting in nonzero average streaming velocities and circulation zones within the duct. A two-dimensional, long and narrow duct is considered in the present study with one end closed and the other subject to a sinusoidal acoustic velocity of varying frequency and amplitude, leading to a standing wave within the duct. A computational fluid dynamics code is used to solve the unsteady, compressible Navier–Stokes equations in two dimensions with turbulence modeling. Since the formation of streaming depends on the velocity profile within the Stokes layer, a high resolution grid is employed near the walls. The effect of inlet conditions is examined in terms of the circulation patterns, including the number of zones per wavelength and penetration depth into the fluid, as well as the mean velocity profiles.

3:05–3:20 Break

3:20

1pEA5. On the sound sources of the screech tone radiated from a choked circular jet. Yoshikuni Umeda and Ryuji Ishii (Yoshida Honmachi, Sakyo-ku, Kyoto, Japan, umeda@kuaero.kyoto-u.ac.jp)

In this experimental study, the generation of the screech tone in the dominant helical oscillation mode *C* was investigated by using a series of instantaneous Schlieren photographs and it was observed that the moving sound source for mode *C* appears to be the prominent point on the helical-shaped coherent vortical structure. This sound source is generated by the interaction between the helical-shaped vortical structure and the oblique shock at the rear edge of the third shock cell. The sound source rotates about the jet axis in the plane perpendicular to the jet axis at a constant supersonic speed and forms a Mach cone. The envelope of the Mach cone widens upstream and downstream as the sound source moves along the circular orbit and forms two helical-shaped wave-fronts of the screech tone propagating in upstream and downstream directions.

3:35

1pEA6. The multiple source feedback of supersonic jet screech. IV. Space–time diagrams and the fast/slow waves of transitory wave theory. Alan Powell (Dept. of Mech. Eng., Univ. of Houston, Houston, TX 77204-4792)

In Part II [J. Acoust. Soc. Am. **108**, 2588 (2000)] the space–time wave diagrams of multiple (point) source screech were elucidated, showing the condition for perfect, or nearly so, reinforcement of the feedback sound at the nozzle. In the transitory wave theory [J. Acoust. Soc. Am. **106**, 2194 (1999)] the velocity modulation of a transitory instability wave as it traverses the periodic cell structure results in a *continuous* phased array which was resolved into fast (approx. sonic) upstream and slow downstream source waves (of constant sinuous amplitude, apart from the envelope factor), only the former radiating strongly, in the upstream direction. The two points of view are reconciled by constructing fast and slow waves in the space–time diagrams by: (a) introducing intermediate negative sources: (b) taking the sources at half strength and adding virtual sources at the zero positions on the fast wave traces, of the same sign as adjacent sources: (c) taking the other half-strength sources and adding virtual sources of opposite sign, the sources then being of the same sign along the slow wave traces. The implications of the need to introduce negative sources in the point source array are discussed.

3:50

1pEA7. Vortex shedding characteristics of two cylinders with differing diameters. Muhittin Sylemez (Dept. of Ocean Eng., Istanbul Tech. Univ., Ayazaga Kampusu, 80626 Istanbul, Turkey) and Mehmet Atlar (Univ. of Newcastle upon Tyne, Newcastle upon Tyne NE1 7RU, UK)

This paper presents an experimental investigation of the vortex shedding characteristics of two circular cylinders with differing diameters in representing a simple riser configuration. The smaller of the two cylinders is located at varying positions around the circumference of the larger cylinder. Measurement of forces on the cylinders and the associated shedding frequencies are recorded and analyzed for five different flow velocities and orientations of the riser in the Emerson Cavitation Tunnel of Newcastle upon Tyne University. The Reynolds number corresponding to the test conditions varies from 9.64×10^4 to 4.82×10^5 for the large diameter of the cylinder and from 3.5×10^4 to 1.75×10^5 for the small diameter cylinder. The experiments were split into two distinct phases: the first was flow visualization using the hydrogen bubble technique and the second phase was force and shedding frequency measurements.

MONDAY AFTERNOON, 4 JUNE 2001 PDR 18, 1:00 TO 2:15 P.M.

Session 1pPA

Physical Acoustics: Nondestructive Testing

David H. Chambers, Chair
Electrical Engineering, Lawrence Livermore National Laboratory, L-154, P.O. Box 808, Livermore, California 94551-5508

Contributed Papers

1:00

1pPA1. Propagation of Lamb waves in a bonded plate. Xiao-Min Wang and Ming-Xuan Li (Inst. of Acoust., Chinese Acad. of Sci., 17 Zhongguancun St., Beijing 100080, PROC, xiaomwang@2911.net)

A solution to the propagation of Lamb waves in a two-layered elastic plate for a variety of interfacial conditions is obtained. The bond rigidity between the adjacent layers is accounted for by a spring model proposed by many researchers. Instead of being in the conventional form of a determinant, the resulting dispersion equation is given in a form of polynomials with the normal and tangential interfacial stiffness constants as the sum-coefficients. This novel expression of dispersion equation has an explicit physical interpretation and application for inversion of the stiffness constants. The influences of the variations of the stiffness constants on the propagation of Lamb waves are numerically studied. The results for different adhesive states show that the degradation of the bond rigidity may cause the decreasing velocity and the frequency-shifts of a given Lamb wave mode, thus having a possibility for inspecting adhesively bonded joints by using the ultrasonic wave velocity measurements. Experimental results to verify the theoretical analysis are given and the empircal relationship between the bond rigidity and the Lamb wave velocity of several lower modes is presented. [Work supported by the National Natural Science Foundation of China.]

1:15

1pPA2. Ultrasound propagation in random, two-phase media with application to damage detection in concrete. Joseph A. Turner, Sean P. M. Dugan, and Liyong Yang (Dept. of Eng. Mech., W317.4 NH, Univ. of Nebraska–Lincoln, Lincoln, NE 68588, jaturner@unl.edu)

The scattering of ultrasonic waves in a random two-phase medium, such as concrete, is discussed. Explicit expressions are derived for the attenuation of longitudinal and shear waves in terms of the statistics of the density and Lam parameter fluctuations. The derivation is based upon diagrammatic methods. The problem is posed in terms of the Dyson equation, which is solved for the mean field response. The attenuation results are given here in a straightforward manner. The general description is further simplified in accord with assumptions representative of a random, two-phase material. The medium is assumed statistically homogeneous and statistically isotropic, with a single correlation length. With these assumptions, the attenuations reduce to simple integrals on the unit circle. Expressions for mean free paths and elastic diffusivity follow from the general scattering theory. Applications to diffuse ultrasound experiments in concrete are then discussed. In particular, the appropriate form of the two-point spatial correlation function is examined and compared with experimental results from digitized concrete images. These results show that the assumed form of the correlation function describes the spatial statistics well. Finally, damage detection prospects relevant for concrete inspection are discussed. [Research supported by NSF Grant No. CMS-99-0078707.]

1:30

1pPA3. A crack model for nonlinear wave modulation spectroscopy. Boris Gommerstadt (Northeastern Univ., Boston, MA, 02115 Gommers@coe.neu.edu) and Alexander Sutin (Stevens Inst. of Technol., Hoboken, NJ 07030)

Nonlinear wave modulation spectroscopy is a novel technique for damage and crack evaluation based on the effect of modulation of ultrasonic waves by vibration. This technique was developed as an experimental approach and our goal is to discuss some mathematical models simulating these phenomena. The first model describes variation of resonance frequencies of the ultrasonic wave by breathing crack oscillation under low frequency vibration. This model considers an opened crack with the length changing under the action of vibration. The second model is concerned with plastic deformation in the vicinity of the crack tips that develops a hysteretic stress–strain relationship leading to the increase of ultrasonic wave damping. The Dugdale model is employed to describe ultrasonic wave damping and its variation under the vibration. These models relate modulation parameters to crack size and position. The suggested models explain the basic features of the observed modulation phenomena and could be used for characterization of the dimension and localization of cracks and crack-like damage.

1:45

1pPA4. Crack location using nonlinear means applying modulation of ultrasonic pulses by cw vibration. Vyacheslav V. Kazakov (Inst. of Appl. Phys. RAN, Nizny Novgorod, Russia), Alexander M. Sutin (Stevens Inst. of Technol., Hoboken, NJ), and Paul A. Johnson (Los Alamos Natl. Lab., Los Alamos, NM)

Nonlinear wave modulation spectroscopy (NWMS) was recently introduced as a new tool in NDE. This technique employs the nonlinear interaction of ultrasound and vibration in the presence of defects. Vibration changes the contact area within a defect effectively modulating an ultrasonic wave sensing the defect. NWMS has high sensitivity to the presence of cracks and can be a very good tool for a ''pass–fail'' test but cannot provide the crack location. A new method for locating defects or cracks in a material, presented in this paper, is based on the modulation of ultrasonic pulses by vibration. The cw vibration induces modulation of the signal reflected from cracks. Measurements of the spatial distribution of the modulation level are the basis of nonlinear acoustic imaging. Experimental verification of the method has been conducted in steel plates containing a circular hole and a fatigue crack. The high-frequency (carrier) impulse was about 3 MHz, and the modulating vibration frequency was 5–15 Hz. This technique images a crack very well while a hole was not detected. As the location of damage is the next important issue in NWMS, this work represents an important step. [This work was partially supported by International Science Technical Center, Grant No. 1369.]

2:00

1pPA5. Laser-induced weak shock in fused silica for destructive testing of thin films. Junlan Wang and Richard L. Weaver (Theoretical and Appl. Mech., Univ. of Illinois, Urbana, IL 61801)

Dynamic testing of the adhesion of thin films to solid substrates is a challenging problem. Work of the past decade testing such films by the impact of a compressive stress wave pulse generated by a laser pulse has determined that a critical parameter is the fall time of the stress pulse. Here we report the development of a variation that allows the necessary fall times to be readily achieved. A YAG laser of moderate power (100 mJ) and pulse duration (10 ns) illuminates a 0.4-μm aluminum layer on a fused silica substrate. A transparent 300-μm water–glass layer provides an inertial backing. The resulting compressive stress pulse that enters the substrate has a strain amplitude of the order of 2% and a duration of 10 ns. The negative elastic nonlinearity of the fused silica causes this pulse to evolve, over a distance of 1 mm, into a linear ramp followed by a weak shock that has a theoretical fall time of 10 ps. Interferometric measurements of the pulse, and observations of the failure of a micron-thick aluminum layer on the opposite side of the substrate, confirm this understanding. [Work supported by NSF.]

Session 1pPP

Psychological and Physiological Acoustics: Exploring the Dynamics of Auditory Perception Through Neuroimaging

Ann Clock Eddins, Chair
Center for Hearing and Deafness, University of Buffalo, 215 Parker Hall, Buffalo, New York 14214

Invited Papers

1:30

1pPP1. ''What'' and ''where'' in the auditory cortex. Josef P. Rauschecker (Georgetown Univ. Medical Ctr., WP15 NRB, 3970 Reservoir Rd., N.W., Washington, DC 20007)

Several cortical fields have been defined in the superior temporal gyrus (STG) of the rhesus monkey. These ''belt'' areas are further characterized by their responses to spatially localized, species-specific vocalizations. Neurons in the caudal belt are more specific for spatial location, whereas neurons in the rostral belt are more selective for communication calls. Furthermore, the caudal and rostral belt project to distinct regions of the prefrontal cortex previously identified as subserving roles in visual spatial analysis and object recognition, respectively. Functional neuroimaging in humans also supports the existence of dual streams for the processing of sound content and location. Complex sounds generally activate a larger area in the belt region of the STG than tones. Consonant–vowel combinations give rise to activity distributions anterior to Heschls gyrus, which are organized into ''phonetic maps.'' If visual input is added in a McGurk-type fusion experiment, these maps are found to shift in individual subjects. Virtual auditory space stimuli administered via headphones activate two regions in the posterior parietal cortex. When this activation is compared with that by a visual spatial task, no overlap is found in the inferior parietal cortex, whereas convergent input is present in a superior parietal region.

2:00

1pPP2. Functional imaging of temporal auditory processing. Timothy D. Griffiths (Functional Imaging Lab., Queen Square, London WC1N 3BG, UK, t.d.griffiths@fil.ion.ucl.ac.uk)

Functional imaging of the auditory system can be carried out using positron emission tomography (PET) and functional MRI (fMRI). Both techniques depend on the blood flow response to stimuli as a measure of brain activity, a process with a delay of 10 s in the auditory cortex. This limits the ability of the techniques to investigate temporal processing in the auditory system. Experiments to investigate the basis for temporal processing are critically dependent on the model that is used to interpret the brain blood flow data. The blood flow response in such experiments depends on local mean synaptic activity; this can be increased either by the firing rate in a subpopulation of cells or by increased synchronization of neural activity. At the millisecond level, an increase in the fMRI BOLD response with the temporal regularity of a signal at the millisecond level has been demonstrated as early as the cochlear nucleus. This may be due to an increased firing rate of onset chopper cells tuned to a particular autocorrelation delay or to increased synchronized activity in the whole population. The temporal processing of patterns over longer periods (seconds) is likely to involve areas of the cortex distinct from the primary auditory cortex.

2:30

1pPP3. Cognitive and sensory influence on the perception of complex auditory signals. Ann Eddins, David Eddins (Dept. of Communicative Disord. and Sci., Univ. at Buffalo, Buffalo, NY 14214), Mary Lou Coad, Alan Lockwood (School of Medicine, Univ. at Buffalo and WNY VA Medical Ctr.), and Charles Watson (Indiana Univ.)

Using positron emission tomography (PET), we have been studying the relationship between psychophysical performance and cortical activation during auditory discrimination tasks. In one experiment, we studied normal hearing listeners who had markedly different perceptual performance on a battery of auditory discrimination tasks. Changes in regional cerebral blood flow (rCBF) were measured while subjects discriminated changes in a tonal sequence or a temporally analogous syllable sequence. The results showed that behavioral performance was correlated with cortical activation in auditory sensory areas as well as cognitive areas related to attention and memory, depending on the demands of the task. To evaluate further whether attention to specific features of complex auditory signals influences the recruitment of cognitive processing areas of the cortex, we studied changes in rCBF when subjects

attended to different features of a 1-octave band signal that varied in spectral envelope frequency (i.e., ripple frequency) or carrier frequency. The results showed that when subjects attended to ripple frequency relative to carrier frequency, activation in the thalamus increased. Moreover, when they discriminated ripple frequency in a low-carrier versus a high-carrier frequency, there was increased activation in the anterior cingulate gyrus, which is known to be involved in attention function.

3:00–3:15 Break

3:15

1pPP4. Some neural correlates of speech and nonspeech perception. J. R. Binder, E. Liebenthal, J. N. Kaufman (Dept. of Neurology, The Medical College of Wisconsin, 9200 W. Wisconsin Ave., Milwaukee, WI 53226), R. L. Piorkowski, and R. E. Remez (Barnard College, New York, NY 10027)

In a series of functional magnetic resonance imaging experiments aimed at examining the perceptual boundary between speech and nonspeech stimuli, brain activation signals were measured while subjects listened passively to a continuum of sounds ranging from noise to speech, detected speech signals in varying levels of noise, and perceived the same sinewave speech replicas as either nonspeech or speech while performing an auditory discrimination task. Activation at the anterolateral aspect of left Heschl's gyrus increased with signal-to-noise ratio in the passive listening and noise masking experiments, and was higher when subjects heard sinewave speech as nonspeech than when they heard the same stimuli as speech while performing an auditory task that conflicted with the speech percept. Activation in this area was correlated with behavioral measures of phoneme perception and was unrelated to attentional demands. The data suggest a key role for this region in the representation of spectrotemporal information underlying phoneme recognition.

3:45

1pPP5. The functional anatomy of tinnitus. Alan H. Lockwood (VAWNYHS and Univ. at Buffalo, Buffalo, NY 14215), Richard J. Salvi, Robert F. Burkard, and Samuel A. Reyes (Univ. at Buffalo, Buffalo, NY 14215)

We have used positron emission tomography to map the neural systems that mediate subjective tinnitus. In two studies, subjects altered their tinnitus by voluntary jaw muscle contraction or sustained lateral gaze. In the third, tinnitus was altered by intravenous lidocaine. These experiments all linked the loudness of tinnitus with unilateral spontaneous neural activity in auditory cortex. This unilaterality contrasts with the bilaterality of cortical responses during acoustic stimulation, evidence for a noncochlear central origin for tinnitus. Plasticity in the central auditory system, thought to be associated with deafferentation, was demonstrated in hearing-impaired or unilaterally deaf patients by the greater extent of auditory cortex activation by tonal stimuli when compared to normals. Other sensory and motor systems may interact with these aberrant pathways, further upsetting the balance of auditory system activity. This is manifested by patients' reports of transient voluntary loudness and/or pitch control. In controls, lateral gaze suppressed auditory cortical activity: this cross-modal inhibition was absent in the patients. Subjective tinnitus appears to be due to spontaneous neural activity in aberrant central auditory pathways that have abnormal interactions with other sensory and motor systems. This complexity may explain the unresponsiveness of tinnitus to pharmacological interventions.

Contributed Paper

4:15

1pPP6. Functional imaging of brain activity during speech production. John Sidtis (Dept. of Psych., New York Univ., 6 Washington Pl., 8th Fl., New York, NY 10012)

The functional anatomy of the cerebral cortex, subcortical nuclei, and cerebellum during speech production is understood primarily through the study of defects following brain damage. Functional imaging methods may provide another view of this system in normal as well as in impaired speakers. In the present study, 13 right-handed, native-English-speaking subjects performed a 60-s syllable repetition task during imaging of cerebral blood flow using positron emission tomography (PET). Each subject was scanned eight times (alternating speech and resting control conditions) and their repetitions of the syllables \pa, ta, ka\, produced as quickly as possible, were recorded for acoustic analysis. Blood flow values were measured from 11 regions of interest on each side of the brain. Speech samples were analyzed with respect to several characteristics including rate. Multiple linear regression analyses were used to predict each acoustic measure. Speech rate was predicted by a linear combination of increased flow in the left inferior frontal region (Broca's area) and decreased flow in the right basal ganglia. This functional anatomy is consistent with clinical observations, but was only apparent when performance measures were considered in the analyses. Functional imaging is most useful when both psychological and physiological measures are integrated.

Session 1pSC

Speech Communication: Speech Perception Potpourri (Poster Session)

Joan E. Sussman, Chair
Department of CDS, University of Buffalo, 122 Cary Hall, Buffalo, New York 14214

Contributed Papers

All posters will be on display from 1:30 p.m. to 5:00 p.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 1:30 p.m. to 3:15 p.m. and contributors of even-numbered papers will be at their posters from 3:15 p.m. to 5:00 p.m.

1pSC1. A computational model for detecting and perceiving dynamic properities of speech. Partha Niyogi (Dept. of Computer Sci., The Univ. of Chicago, 1110 E. 58 St., Ryerson 157, Chicago, IL 60637)

We investigate many aspects of a computational framework for detecting phonetically salient dynamic properties of the speech signal from continuous speech input. As a prototypical example of this problem, we consider frameworks for the detection of the transition from closure to burst for stop consonants in continuous speech. We formulate the problem as an optimal non-inear filtering problem. This involves the mapping of the dynamic acoustic data into a low dimension Riemannian manifold in an infinite dimensional Hilbert space. This leads to many kinds of metrics that can now be imposed on the dynamic speech signal and we explore these metrics in the context of the perceptual magnet effect. An extensive discussion of the robustness of the strategy, the nature of the errors made, the relationship to acoustic phonetis and distinctive features and to categorical perception will be conducted. Related problems that have a similar structure include nasal detection using transitions from vowels into adjacent nasals, transitional models of vowels and so on. Results of automatic speech recognition experiments using these dynamic features will be presented.

1pSC2. Auditory context effects for nonequivalent sources. Lori L. Holt (Dept. of Psych., Carnegie Mellon Univ., 5000 Forbes Ave., Pittsburgh, PA 15213) and Andrew J. Lotto (Washington State Univ., Pullman, WA 99164)

One of the most compelling examples of the symmetry in speech perception and production is the ability of the perceptual system to maintain phonetic constancy despite the context sensitivity in acoustic realization of the phonetic segment arising from coarticulation. It may be suggested that the perceptual system is able to accomplish this compensation for coarticulation by recovering the dynamic information in the surrounding acoustic context of the segment. We will present previous data and novel demonstrations that suggest that the contextual information need not come from the same source nor even resemble speech. For example, combinations of pure tones, which are not heard as speech, can shift the identification of neighboring consonants. These nonspeech effects can be described as cases of general auditory contrast. Whereas it is typical in dynamic theories of speech to refer to information specific to vocal tracts in accounting for speech perception phenomena, we feel that the present data are not completely inconsistent with a dynamic approach. The presence of contrast effects suggest to us that the perceptual system respects the dynamic nature of the environment by enhancing sudden canges in kinematics, which are presumed to be caused by application of novel forces.

1pSC3. Perception of English consonants from dynamic acoustic properties. Robert E. Remez, Rebecca L. Piorkowski, Stephanie Wissig, and Abigail Batchelder (Dept. of Psych., Barnard College, New York, NY 10027-6598)

Phonetic perception is often said to depend on short-term attributes of vocally produced sound. Counterevidence to this premise is found in the example of sinewave replicas of speech. The dynamic properties of a sinewave replica evoke phonetic impressions despite the absence of typical or natural sounding short-term elements. To gauge the robustness of phonetic perception from dynamic acoustic properties, we compared identification performance by normal listeners across identification tests composed of signals that preserve dynamic properties of speech despite the absence of short-term speechlike properties: sinewave replicas modeled on the natural samples of 18 English consonants in CV syllables, noise-band vocoded signals derived from the natural samples, and chimerical signals exhibiting the coarse-grain spectrum envelope of natural speech and the fine structure of nonspeech signals. Identification from dynamic acoustic properties was well predicted from control performance using sampled natural speech in noise, showing that short-term acoustic properties and their corresponding auditory qualities are inessential in perceiving consonants. This outcome is consistent with a perceptual account that favors attention to the characteristic modulations of an acoustic carrier, and opposes an account appealing to auditory essences of each phonetic element.

1pSC4. The perception of rate induced resyllabification in English. Kenneth de Jong, J. Byung-jin Lim, and Kyoko Nagao (Dept. of Linguist., Indiana Univ., Bloomington, IN 47405)

Stetson (1951) noted that, when repeated, singleton coda consonants (VC) appear to modulate into onset consonants (CV) as the rate of repetition increases. The current study examines whether nave listeners perceive such resyllabifications, and whether such perceptions are affected by the voicing of the resyllabified consonants. Stimuli were extracted from production experiments in which talkers repeatedly produced singleton voiced and voiceless stops in the CV or VC position. Speakers entrained to a metronome which increased tempo from 450 to 150 ms/syllable. Open-set identification revealed that (1) while slow VCs are identified as such, fast VCs are identified as CVs a majority of the time and (2) CVs are rarely identified as VCs; however (3) both CVs and VCs at fast rates are often identified as CVCs, especially when the consonant is voiceless. A forced-choice identification task indicates that fast VCs and CVs, while clearly differentiated at slow rates, are both identified as CVs 80% of the time at fast rates. These results support Stetson's observations, but like previously reported production results, indicate that fast rate tokens are partially ambiguous between CV and VC forms, an ambiguity which can get resolved by splitting the consonant in question. [Work supported by NIDCD and NSF.]

1pSC5. Dynamic information signals voicing for syllable-final stops. Susan Nittrouer, Sandy Taylor, Lesley Larive, Carol Manning, and Tammy Mertes (Boys Town Natl. Res. Hospital, 555 N. 30th St., Omaha, NE 68131)

Several acoustic properties can influence listeners' judgments of voicing for syllable-final stops, including vocalic length. Developmental and cross-linguistic experiments have suggested that listeners with little experience about how vocalic length is related to final-stop voicing weight this property less than more-experienced listeners. For five (English, CVC) minimal pairs, we examined the acoustic consequences of final-stop voicing, as well as listeners' weighting of two properties that signal voicing: vocalic length and syllable-offset transitions. First, an acoustic analysis of words showed that differences in vocalic length associated with final-stop voicing are attenuated when words occur in continuous speech, raising questions about how much experience any listener really has hearing vocalic-length differences correlated with final-stop voicing. Next, adults and children (3, 5, and 7 years old) labeled natural stimuli made from words that originally had voiced or voiceless stops, in which we modified vocalic length. Partial correlations revealed that listeners of all ages weighted offset transitions more than vocalic length. In fact, listeners were reluctant to label any stimulus that did not have transitions clearly indicating closure as ''voiced,'' and so the traditional ''vocalic-length effect'' was apparent only for originally voiced stimuli. We conclude that dynamic information takes precedence in this perceptual decision.

1pSC6. Change deafness: The inability to detect changes in a talker's voice. Michael Vitevitch (Dept. of Psych., Indiana Univ., 1101 E. 10th St., Bloomington, IN 47405-7007, mvitevit@indiana.edu)

Change blindness is a failure to detect a change in a visual scene. A shadowing task was used to demonstrate an auditory analog to change blindness–change deafness. Participants repeated words varying in lexical difficulty. After a rest-break they heard more words from either the same or a different talker. Answers to explicit questions about the change in talker and implicit measures of behavior (i.e., response latencies) demonstrate that processing is affected by the change, even if participants do not explicitly report a change in talker. Specifically, listeners who did not detect the change in the talker had a greater difference between conditions of lexical difficulty than listeners who noticed the change, or listeners who heard the same talker throughout. These results suggest that failures to detect changes are not limited to the visual domain and that processing at some level may be affected by changes in the environment. Furthermore, these results have implications for models of spoken word recognition.

1pSC7. Formant continuity and auditory scene analysis: The effect of vowel formant manipulations on the perception of synthetic nasal consonants. Susan M. Harding and Georg F. Meyer (Mackay Inst. of Commun. and Neurosci., Keele Univ., Keele, Staffs ST5 5BG, UK, s.m.harding@cns.keele.ac.uk)

The perception of a synthesized nasal /m/ changes to /n/ as the frequency of the second formant of a preceding vowel is increased when there are no transitions between the vowel and nasal; but /m/ is heard consistently when 20-ms transitions are introduced. A possible explanation is provided by auditory scene analysis: Formants of the vowel and nasal that are contiguous and close in frequency may be grouped using principles of similarity and good continuity into a single perceptual stream. Further experiments found a robust change in percept for prototype /n/ as well as /m/. Transitions from the vowel that conflict with the formant structure of the consonant also cause a similar change in percept for both nasal prototypes. The proximity of the second formant of the vowel to a formant in the nasal prototype is therefore unnecessary for the change in percept to occur; the presence of the vowel formant near to one of two target frequencies at the boundary with the nasal seems to be sufficient to determine the nasal percept and takes precedence over the structure of the nasal prototype. Thus, these results do not show strong evidence for auditory scene analysis applied to formants.

1pSC8. Perceptual experiment on Korean number production. Byunggon Yang (Dongeui Univ., 24 Kayadong, Pusanjingu, Pusan 614-714, Republic of Korea, bgyang@hyomin.dongeui.ac.kr)

Acoustic parameters of the nine Korean numbers were analyzed by PRAAT, a speech analysis software, and synthesized by SENSYNPPC, a Klatt formant synthesizer. The overall intensity, pitch, and formant values of the numbers were modified dynamically by a step of 1 dB, 1 Hz, and 2.5%, respectively. This study explored the sensitivity of listeners to changes in the three acoustic parameters. Twelve male and female subjects listened to 390 pairs of synthesized numbers and judged whether the given pair sounded the same or different. Results showed that subjects perceived the same sound quality within the range of 6.6 dB of intensity variation, 10.5 Hz of pitch variation, and 5.9% of the first three formant variations. The male and female groups showed almost the same perceptual ranges. Also, an asymmetrical structure of high and low boundary was observed. The ranges may be applicable to the development of a speaker identification system while the method of synthesis modification may apply to its evaluation data. [This work was supported by a grant from the Interdisciplinary Research Program (Contract No. 1999-2-302-106-5) of the KOSEF.]

1pSC9. Discrimination and labeling differences in audiovisual spech perception. Nicole L. Marrone and Arlene E. Carney (Dept. of Commun. Disord., Univ. of Minnesota, 164 Pillsbury Dr. SE, Minneapolis, MN 55455)

In an earlier report, Clement *et al.* [J. Acoust. Soc. Am. **107**, 2887 (2000)] showed that listeners' confidence ratings of the identity of audiovisual syllables produced with conflicting auditory and visual cues were typically lower than those for audiovisually congruent syllables. This occurred even when both types of syllables were labeled consistently as the same phoneme. Confidence ratings varied across listeners, suggesting a graded nature of auditory–visual integration even across listeners who exhibit visual bias. In the current experiment, listeners were asked to identify audiovisual tokens from two talkers who elicit a high proportion of visually biased responses [Carney *et al.*, J. Acoust. Soc. Am. **106**, 2270 (1999)] and to rate their confidence in their labeling judgment. Only listeners who exhibited strong visual bias were tested in a discrimination task in which a true audiovisual /di/ was a standard. Comparison stimuli were visual /gi/-auditory /bi/, visual /bi/-auditory /gi/, and true audiovisual /bi/, /di/, and /gi/. Listeners were able to discriminate the visually biased and true /di/ stimuli while labeling them both as alveolar tokens. Results support the notion of a graded perception of bimodal stimuli. [Research supported by NIDCD.]

1pSC10. Sinewave speech/ nonspeech perception: An fMRI study. Einat Liebenthal, Jeffrey R. Binder (Neurology Dept., Medical College of Wisconsin, 8701 Watertown Plank Rd., Milwaukee, WI 53226, einatl@mcw.edu), Rebecca L. Piorkowski, and Robert E. Remez (Barnard College, New York, NY 10027)

Sinewave speech replicas follow the center frequency and amplitude of vocal tract resonances of natural speech but lack the attributes of vocal chord vibrations. They can be perceived as speech or nonspeech depending on expectation (Remez *et al.*, 1981). Brain areas activated by sinewave replicas when perceived as nonspeech or as speech were compared with functional magnetic resonance imaging. In an auditory task, 30 nave subjects determined whether isolated second-tone formants were included in a following three-tones sinewave complex. The second-tone formant was either aligned (phonetic stimulus) or temporally reversed relative to the other tones in the complex (auditory stimulus). Halfway through the auditory task, subjects were informed about the phonetic derivation of the stimuli and trained to recognize the original speech. In a subsequent phonetic task, subjects identified sinewave stimuli containing the phoneme /p/. Following training, an increase in behavioral response time, concurrent with a decrease in left Heschl's gyrus activation was observed specifically with the phonetic stimuli. This suggests automatic

recognition of speech structure in the auditory cortex and interference with auditory processing of phonetic stimuli, acquired by training. In the phonetic task, the left inferior frontal gyrus was specifically activated implicating this area in phonetic processing.

1pSC11. Perceptual metathesis of obstruent clusters. Matthew J. Makashay (Dept. of Linguist., Ohio State Univ., 222 Oxley Hall, 1712 Neil Ave., Columbus, OH 43210, makashay@ling.ohio-state.edu)

This study examined acoustic and perceptual cues of obstruent clusters in order to test the hypothesis that metathesis can be a process that maintains identification of the consonants involved. In an auditory lexical decision task, there were effects of both optimality of cues, and lexical frequency of clusters. Nonword targets were created by metathesizing medial obstruent clusters in English words (e.g. [nækpɪ n] for *napkin*). For the clear listening level group, there was a slow rejection of nonword targets with optimal clusters that occur with higher frequency in the lexicon. For the speech reception threshold group, nonword targets with optimal clusters were more likely to be perceptually metathesized and realized as words than nonword targets with nonoptimal clusters were, presumably because subjects are more likely to hear both consonants in optimal clusters. Clusters with fricatives and stops were less likely to be perceptually metathesized than clusters containing only stops, since the continuity of manner features in a stop cluster hinders perception of consonant order. Whether some listeners are more adept than others in the ordering of temporal acoustic events will be investigated as well. [This material is based upon work supported under a National Science Foundation Graduate Fellowship.]

1pSC12. Flexibility of acoustic cue weighting in children's speech perception. Catherine Mayo, Alice Turk (Dept. of Theoretical & Appl. Linguist., Univ. of Edinburgh, Edinburgh EH8 9LL, UK), and Jocelynne Watson (Univ. of Edinburgh, Edinburgh EH9 1UW, UK)

Nittrouer and colleagues [Nittrouer, J. Phonetics **20**, 1–32 (1992); Nittrouer and Miller, J Acoust. Soc. Am. **101**, 2253–2265 (1997); Nittrouer *et al.*, Percept. Psychophys. **62** (2000)] have found that in identifying certain syllable contrasts, young children make more use of syllable–internal formant transitions (relative to other available acoustic cues) than do older children and adults. The evidence for this change in the degree to which listeners weight, or use, certain cues comes predominantly from studies of fricative contrasts (e.g., /sV/–/ʃV/, /sV/–/stV/, /Vs/–/Vʃ/). The current study tests the flexibility of children's weighting of acoustic cues by examining cue weighting across a wider range of phonetic contexts. In particular, this study attempts to determine whether children's focus of perceptual attention can be led away from transitions in contexts where such cues are relatively less salient. Additionally, the study tests children's ability to identify phonemes in an extreme situation, in the complete absence of transitional information. [Work supported by Wellcome Trust.]

1pSC13. Backward masking in speech perception by children and adults. Joan Sussman and Elizabeth Laczi (Dept. of CDS, Univ. at Buffalo, 105 Cary Hall, 3435 Main St., Buffalo, NY 14214, jsussman@acsu.buffalo.edu)

The current investigation measured discrimination and identification of tone-noise and [b]-[a] stimuli by adults, children with normally developing language, and children with language impairment. The children were aged 4–6 years. These tasks were chosen instead of the more traditional two alternative forced choice (2AFC) technique because it was believed possible that task difficulty influenced previous results (e.g., Wright *et al.*, 1997). The stimuli in the current set of experiments had longer tones (40 ms) than prior studies but used higher masker levels (52 dB spectrum level). However, results showed that all the participants discriminated the stimuli containing the tone or [b] signal significantly above chance, contrary to previous findings. More difficulty was found in the identification task, similar to prior backward masking results. [Supported by an Individual Development Award, NYS/UUP.]

1pSC14. Overt visual attention in spoken sentence perception. Charissa R. Lansing (Dept. of Speech and Hearing Sci., Univ. of Illinois at Urban-Champaign, 901 S. Sixth St., Champaign, IL 61820, crl@uiuc.edu) and George W. McConkie (Beckman Inst., Univ. of Illinois at Urbana-Champaign, Urbana, IL 61801)

People's eye movements and performance accuracy were recorded as they attempted to understand sentences spoken by two talkers under two conditions: vision only and vision plus low-intensity sound. Percent word-correct scores were higher for the vision-plus-sound than for the vision-only presentation and for the male compared to the female talker. Eye movement records showed a tendency to gaze at the talker's eyes when the talker was not speaking, but to shift the gaze to the mouth and make long eye fixations when the talker was speaking, particularly under vision-only conditions and for the female talker. In a task requiring verbatim word identification, people with average speech-reading proficiency direct their gaze to the talker's mouth most of the time during the talker's speech production, contrary to the finding of Vatikiotis-Bateson, Eigsti, Yano, and Munhall (1998), and they produce very long eye fixations. For these people, the gaze is drawn to the mouth, not by facial motion alone, but also on some other basis that is assumed to be prior knowledge of the location of critical visual cues, with an accompanying suppression of saccadic activity. [Work supported by NIH-NIDCD Grant DC02250.]

1pSC15. Limits of sentence identification in gated and continuous noise. Peggy B. Nelson, Su-Hyun Jin, and Arlene Earley Carney (Dept. of Commun. Disord., Univ. of Minnesota, 164 Pillsbury Dr. SE, Minneapolis, MN 55455)

Listeners with normal hearing sensitivity are able to take advantage of temporal dips in fluctuating noise. Their word identification in gated noise is better than identification in continuous noise. The limits of this ability are not well understood. Young adult listeners with normal hearing sensitivity were tested for their understanding of speech in gated and continuous noise at a variety of signal-to-noise ratios (SNRs) and gate frequencies from 2–32 Hz. Stimuli were digitized IEEE and CID sentences spoken by 10 talkers (5 male and 5 female). Pseudorandom noise was generated with the same long-term spectrum as the speech. Sentences were presented at an overall level of 65 dB SPL, with noise at +16-, +8-, 0-, −8- and −16-dB SNR. Listeners' responses were scored for the number of keywords correct. Results showed that at −16-dB SNR of continuous noise, listeners were unable to repeat any keywords. Gate frequencies of 4 and 8 Hz resulted in the greatest keyword identification. Performance decreased at the slowest gate frequency for IEEE sentences and at fastest gate frequencies for all sentences. Results will be discussed in terms of listeners with hearing loss and cochlear implants. [Work supported by NIDCD.]

1pSC16. Spatial distribution of early reflections and speech intelligibility. Yang-Ki Oh (Mokpo Natl. Univ., Mokpo, Chonnam 534-729, South Korea), Dae-Up Jeong (Chonbuk Natl. Univ., Jeonju, Chonbuk 561-756, South Korea), Se-Jin Doo (Dong-Ah Broadcasting College, Ansung, Kyonggi, South Korea), Hee-Won Lee (Seoul Natl. Univ. of Technol., Seoul, South Korea), Chul-Min Choi, Lai-Hoon Kim (Seoul Natl. Univ., Seoul, South Korea), and Il-Doo Ko (Seoul Natl. Univ. of Technol., Seoul, South Korea)

The strong early reflections and short delay times have been known to improve the intelligibility of speech heard in rooms. D50 and C80, the most frequently used physical parameters, were developed taking this fact into consideration. However, these monaural parameters have limited applications for the practical design of rooms because of their lack of spatial

information. The present work investigates how temporal changes in three-dimensional distribution of early reflections influence speech intelligibility in rooms. A new measurement method, using a five microphone array and an omnidirectional source setup, is employed, and a series of post-processing procedures are involved, for getting different early reflections in their spatial distributions. The changes were made for the impulse responses obtained through a five microphone array in the arrival times of early reflections from all, and the horizontal and vertical directions, respectively. Anechoic samples of the Korean language were convolved binaurally with the reproduced impulses by applying a head-related transfer function. A series of speech intelligibility tests, conducted for 22 university students, found that the percentage of correct responses significantly deteriorated by increasing delay times of early reflections from the vertical direction. The result suggests that vertical components of early reflections play a significant role in improving speech intelligibility. [Work supported by Korean Research Foundation Grant KRF-1999-1-310-004-3.]

1pSC17. Consonants and vowels discriminated differently even when acoustically matched. A. Min Kang (Haskins Labs., 270 Crown St., New Haven, CT 06511 and Yale Univ., New Haven, CT 06520, min.kang@yale.edu)

Vowels are reportedly discriminated differently from consonants, but there have typically been large between-class acoustic differences. Discrimination still differed when acoustic differences were reduced by removing the mostly vocalic center portion of CVCs [silent center (SC)] [A. M. Kang and D. H. Whalen, J. Acoust. Soc. Am. **107**, 2855–2856 (2000)]. The present study compared consonant and vowel identification and discrimination of synthetic CVCs varying in equal-sized $F2$ steps along /b-d/ and /ɛ-ʌ/ continua (full syllables), and in truncated syllables corresponding to the initial 60 ms of the previously examined SC syllables. To lower listener uncertainty, only consonant, or only vowel, information was varied within a test block. Consonant discrimination for full syllables was much higher than in the earlier SC experiment; it was slightly higher for the truncated stimuli than for the full. Vowel discrimination was much higher than consonant, near ceiling for both full and truncated stimuli. Thus, even when acoustic steps are equalized and the speech presented (in the truncated stimuli) is limited to the syllable portion that contains most of the constant information, vowels remain better discriminated than consonants. This indicates a true difference processing of the two phonetic classes, even when the acoustics are well matched. [Work supported by NIH.]

1pSC18. Perception of Cantonese Parkinsonian speech. Patrick C. M. Wong, Randy L. Diehl (Dept. of Psych., Univ. of Texas, Austin, TX 78712), Shu Leong Ho, Leonard S. W. Li (Univ. of Hong Kong, Hong Kong, PROC), and Kin Lun Tsang (Queen Mary Hospital, Hong Kong, PROC)

The current study is a continuation of our previous case study investigating the effect of reduced pitch range in Parkinsonian speech on a tone language [P. C. M. Wong and R. L. Diehl, J. Acoust. Soc. Am. **105**, 1246(A) (1999)]. In the first experiment, listeners were asked to identify the last word of semantically neutral sentences produced by Cantonese-speaking Parkinson's disease (PD) patients, normal speakers, and a resynthesized version of PD speech with expanded pitch range. Identification of normal and PD speech did not differ, perhaps due to the insignificant difference in pitch range between the two types of speech. However, listeners were better at identifying the resynthesized PD speech which contained a larger pitch range than the original PD speech. This latter result supports the theory of context-target pitch distance proposed by Wong and Diehl which states that lexical tone perception relies on a sufficiently large pitch distance between the context and target of an utterance [J. Acoust. Soc. Am. **104**, 1834(A) (1998)]. In the second experiment, subjects were asked to identify the intended intonation (angry, happy, neutral, and question) of sentences produced by normal and PD speakers. Performance was better for normal speech. [Work supported by NIDCD.]

1pSC19. Do listeners to speech perceive gestures? Evidence from choice and simple response time tasks. Carol A. Fowler, Julie M. Brown, Laura Sabadini-Grant (Haskins Labs. and Dept. of Psych., Univ. of Connecticut, 406 Babbidge Rd., Unit 1020, Storrs, CT 06269, fowler@tom.haskins.yale.edu), and Jeffrey Weihing (Haskins Labs., New Haven, CT 06511-6695)

According to the motor and direct realist theories, listeners perceive speech gestures. The following experiments test this claim. Experiments 1 and 2 replicate the findings of Porter and Castellanos [J. Acoust. Soc. Am. **67**, 1349–1356 (1980)]. Participants shadowed vowel-consonant-vowels (VCVs) produced by a model. Responses were timed. The difference between response times (RT) in simple and choice speech shadowing tasks (26 ms) is shorter than the canonical choice/simple RT difference [100–150 ms, Luce (Oxford, New York, 1986)]. This is interpreted as supporting Porter and Castellanos, in that when the task is to shadow speech, the element of choice is considerably reduced as the listener receives instructions for her response from the speech sounds she perceives. In experiment 3, the timing of gestures of the models' speech was manipulated by extending the voice onset time (VOT) of the models' production of voiceless stops in half of the speakers VCVs. VOTs of participants shadowed responses were measured. Our findings suggest that listeners' productions of phonemes can be influenced by their perception of the timing of the models' gestures in speech shadowing tasks. This provides additional support for the interpretation that participants' shadowing responses are guided by their preception of the models' gestures.

1pSC20. Bandpass filtered faces and audiovisual speech perception. Kevin Munhall (Dept. of Psych. and Otolarngol., Queen's Univ., Kingston, Canada), Christian Kroos, and Eric Vatikiotis-Bateson (ATR Intl.—Information Sci. Div., Kyoto, Japan)

The visual system processes images in terms of spatial frequency-tuned channels. However, it is not clear how complex object and motion processing are influenced by this early visual processing. In two studies this question was explored in audiovisual speech perception. Subjects were presented with spatial frequency filtered images of the moving face during a speech in noise task. A wavelet procedure was used to create five bandpass filtered stimulus sets. The CID Everyday sentences were presented with a multivoice babble noise signal and key word identification accuracy was scored. Performance varied across the filter bands with peak accuracy being observed for the images containing spatial frequencies spanning 7–14 cycles/face. Accuracy for higher and lower spatial frequency bands was found to be lower. When viewing distance was manipulated no change in the overall shape or peak in the key word accuracy function was observed. However, at the longest viewing distance the performance in the highest spatial frequency band decreased markedly. The results will be discussed in terms of visual information processing constraints on audiovisual integration.

1pSC21. Neighborhood effects in Japanese word recognition. Kiyoko Yoneyama and Keith Johnson (Dept. of Linguist., Ohio State Univ., 222 Oxley Hall, 1712 Neil Ave., Columbus, OH 43210, yoneyama@ling.ohio-state.edu)

This paper reports on the results of a naming experiment that investigated lexical neighborhood effects in Japanese word recognition. A naming experiment was conducted with 28 Japanese adult listeners. Each participant responded to 700 words that had varying neighborhood density (in terms of Greenberg–Jenkins' phoneme substitution, deletion, and insertion rules). The lexicon used for this calculation consisted of only nouns from the NTT Japanese psycholinguistic database [Amano and Kondo (1999)]. A preliminary regression analysis showed that such neighborhood density was negatively correlated with naming reacting time. The words with higher neighborhood density were responded to faster than those with lower neighborhood density. We plan to report further analyses that (1)

include prosodic information as another dimension of the neighborhood calculation in order to reflect the finding that prosodic information has a vital role in Japanese word recognition, and that (2) calculate neighborhood density based on the auditory properties of the words in the lexicon. Neighborhood density is measured by comparing similarity of audio files in the Amano and Kondo database. Altogether, we plan to discuss which level of lexical representation (a lower-level acoustic-auditory representation or a higher-level abstract phonemic representation) is used to calculate phonological similarity within the lexicon.

1pSC22. Experimental paradigm for examining the formation of complex auditory categories. Andrew J. Lotto (Dept. of Psych., Washington State Univ., Pullman, WA 99164) and Lori L. Holt (Dept. of Psych., Carnegie Mellon Univ., Pittsburgh, PA 15213)

Phonetic perception is at its essence a categorization task involving multiple imperfectly valid cues to category membership. To examine the general cognitive processes underlying the formation and utilization of complex auditory categories such as speech, we have created a set of stimuli crafted from 300-m bursts of white noise. These stimuli vary in onset duration (or attack) and the center frequencies of two spectral gaps. Categories of varying complexity can be created from this stimulus set and input distributions can be strictly controlled. In addition, category formation can be observed ''online'' in this microgenetic design as subjects make responses (identification and discrimination) while learning the categories. In one such task, subjects were trained with categories defined by boundary values across three attributes. After 10 h of training, the subjects demonstrated extensive learning despite the fact that the categories suffered from ''lack of invariance;'' that is, no cue could be relied on exclusively to determine category membership. This demonstrates the viability of this paradigm for examining complex categorization. Implications for speech categorization and phonetic acquisition will be discussed. [Work supported by NSF.]

1pSC23. The role of spectral contrast in the perception of stop consonants following vowels and their spectral complements. Jeffry A. Coady and Keith R. Kluender (Dept. of Psych., Univ. of Wisconsin, 1202 W. Johnson St., Madison, WI 53706)

The ability of listeners to recover speech information, despite dramatic articulatory and acoustic assimilation between adjacent speech sounds is remarkable and central to understanding perception of connected speech. In recent years, studies have revealed that, to some extent, auditory processes of spectral contrast compensate for assimilative effects of coarticulation. In the present studies, series of CV syllables varying acoustically in $F2$-onset frequency and perceptually from [ba] to [da] were identified either following front (e.g., [i], [e]) and back (e.g., [u], [o]) vowels or following complementary nonspeech spectra. These nonspeech stimuli were harmonic spectra in which the amplitudes of individual harmonics were the inverse of their amplitudes for those same harmonics in the vowels. A prediction of a spectral contrast account is that these complementary spectra should affect perception of following sounds in a manner complementary to that for the vowels from which they were modeled. This predicted effect was obtained. Nevertheless, one may expect that not all perceptual accommodation of coarticulation is explained by spectral contrast. In addition to these initial results, findings from experiments in which stimuli and task variables are manipulated to assess the influence of other processes will be reported. [Work supported by NIDCD DC04072.]

1pSC24. Effects of modulation and laterality on the perception of time-varying sinusoidal sentences in noise. Melanie M. Richter and Thomas D. Carrell (Univ. of Nebraska, 318 Barkley Ctr., Lincoln, NE 68583)

Researchers have shown amplitude modulation significantly increases the intelligibility of time-varying sinusoidal (TVS) sentences [T. Carrell and J. Opie, Percept. Psychophys. **52**, 437–445 (1992); J. Barker and M. Cooke, Speech Commun. **27**, 159–174 (1999)] and ameliorates the effect of multispeaker babble on intelligibility at moderate signal-to-noise ratios [T. Carrell, J. Acoust. Soc. Am. **93**, 2327 (1993)]. These effects have been attributed to the ability of amplitude comodulation to create auditory objects. However, other characteristics of acoustic signals cause components to group together, the classic example being sound location in the cocktail party effect [C. Cherry, J. Acoust. Soc. Am. **25**, 437–445 (1953)]. In natural environments, acoustic cues work together to assemble and segregate auditory objects for auditory scene analysis [A. Bregman, *Auditory Scene Analysis* (1990)]. The present study examined the interaction of amplitude modulation and laterality on the intelligibility of TVS sentences in noise. Forty females listened to TVS stimuli in a $2\times2\times2$ factorial design experiment: modulation (100 Hz versus unmodulated) by laterality (diotic versus dichotic) by noise (S/N$=+5$ dB vs -15 dB). Modulation and noise demonstrated large main effects in the expected direction, but effects of laterality were minimal.

1pSC25. Infant-directed speech helps adults separate different streams of speech. Rochelle S. Newman and Tammy Weppelman (Dept. of Psych., Univ. of Iowa, Iowa City, IA 52242)

Infants, like adults, are frequently faced with multiple people speaking at one time. What cues help them separate different voices? One possible cue is infant-directed speech (IDS). Speech directed toward infants differs from adult-directed speech (ADS) in a number of ways. IDS typically involves higher pitch, wider pitch excursions, longer durations, and may involve greater amplitude variability [Aslin (1993); Fernald and Simon (1984); Fernald *et al.* (1989)]. Many of these same features have been shown to be important for adult musical streaming. Furthermore, IDS may serve to make the caretakers voice more distinct from background speech (which is likely to be adult directed). This suggests that it might be easier to separate IDS from ADS than it is to separate two ADS samples. To test this, adult listeners were asked to shadow a target voice while a distractor voice spoke in the background. Performance when one of the voices was infant directed was compared with that when both voices were adult directed. We found a significant effect of speech register on stream segregation, such that listeners found it much easier to follow the target voice when it (but not the distractor) spoke in an infant-directed manner. [Work supported by NSF.]

1pSC26. The basic units of rate normalization. Rochelle S. Newman (Dept. of Psych., E11 Seashore Hall, Univ. of Iowa, Iowa City, IA 52242)

Individuals vary their speaking rate, and listeners use the duration of adjacent segments to adjust for these changes [Miller and Liberman (1979); Newman and Sawusch (1996)]. However, it has not been clear what these ''segments'' actually are. We examined whether two-phoneme sequences would produce two separate rate normalization effects (one for each phoneme), or only a single effect. We created series such as ''shkib-chkib'' and ''shwib-chwib'' and examined effects of the /k/ or /w/ and /I/ durations on perception of the initial contrast. Altering the /k/ duration resulted in a rate normalization effect different from that caused by varying the duration of the sequence ''kib'' as a whole. This suggests that /k/ and the vowel are treated as separate units of speech. However, we did not

find this pattern for the /w/. For sequences such as ''kih,'' which have obvious acoustic boundaries, each phoneme has a separate rate normalization effect. For sequences such as ''wih,'' without such obvious cues, the sequence is treated as a single (larger) segment by the rate normalization process. This suggests that rate normalization is not based on phonemes, but on segments with obvious acoustic boundaries, emphasizing the role of basic auditory processing in speech recognition.

1pSC27. Identification of multiple consonants in binaural presentation. Bom Jun Kwon (House Ear Inst., 2100 W. 3rd St., Los Angeles, CA 90057)

How well do we process multiple speech sounds presented concurrently? Consonant identification was tested with normal-hearing listeners using 13 voiced consonants, which were recorded by 10 speakers and had similar durations of murmur or voice bar, in syllable-initial position. Listeners were presented with multiple (up to four) stimuli with a time delay between the stimuli via headphone either diotically or dichotically (to simulate spatial separation between sound sources). Results showed that while the identification scores decreased as the delay decreased, performance was still quite robust even with the zero delay. The scores for diotic presentation were lower than those for dichotic presentation, but the difference was not overwhelming compared to the decrease in the score due to the delay. The data suggest that (i) listeners are capable of processing and identifying multiple consonants in concurrent presentation, and (ii) the effect of onset differences is more substantial on identification of multiple sounds than binaural separation. In other words, the ''cocktail party'' effect can occur even when there is no spatial separation between sound sources. A possible mechanism of identification of multiple consonants, when the cues to segregate each stimulus is highly limited, will also be discussed. [Work supported by NIDCD and Quota Scholarship.]

1pSC28. Intelligibility of frequency-shifted speech. Jack M. Scott, Peter F. Assmann (School of Human Development, Univ. of Texas–Dallas, P.O. Box 830688, Richardson, TX 75083), and Terrance M. Nearey (Univ. of Alberta, Edmonton, AB T6G 2E7, Canada)

A significant fact about speech perception is that intelligibility can be preserved when the formant pattern is shifted up or down along the frequency scale. To study the relationship between fundamental frequency ($F0$) and spectrum envelope shifts, we used a high-quality vocoder (''Straight'') to process a set of 280 English sentences from the Hearing in Noise Test. Upward shifts in spectrum envelope (shift factors of 1, 2, 3, or 4) were combined with upward shifts in $F0$ (shift factors of 1, 2, or 4). Word recognition accuracy dropped from near perfect for a spectrum envelope shift factor of 1 (unshifted speech) to 77%, 24%, and 7% for shift factors of 2, 3, and 4, respectively. $F0$ shifts had a small but significant effect on identification. For a spectrum envelope scale factor of 1, increasing $F0$ by two octaves (shift factor of 4) resulted in a 16% decline. However, for spectrum envelope shift factors of 2 or greater, increasing $F0$ led to an improvement in performance. The results are consistent with the hypothesis that learned relationships between $F0$ and spectrum envelope affect the intelligibility of frequency-shifted speech.

1pSC29. English-speaking listeners' sensitivity to allophone appropriateness in real and nonword speech perception. Amanda K. Jones (Dept. of Linguist., UCLA, 405 Hilgard Ave., Los Angeles, CA 90025, amandajo@ucla.edu)

The question of whether adult listeners are capable of evaluating speech for the appropriateness of its allophonic content has received relatively little attention. Whalen *et al.* (1997) showed that English-speaking listeners preferred context-appropriate allophones when evaluating real words. However, for nonwords they found that listeners tended to prefer one particular allophone, regardless of its contextual appropriateness. In the current study, listeners provided naturalness ratings for pronunciations of both words and nonwords containing either an appropriate or an inappropriate allophone of the phoneme /t/ in initial /st/ clusters. The words and nonwords were divided into four categories determined by pretesting: highly familiar real words (e.g., ''start''), relatively unfamiliar real words (''stave''), highly wordlike nonwords (''storch''), and less wordlike nonwords (''stimf''). Results showed that listeners consistently preferred pronunciations that contained an appropriate allophone of /t/, regardless of stimuli category, e.g., real word vs nonword. This result contradicts Whalen *et al.*'s suggestion that ''allophonic contrast . . . may require contact with the lexicon in order to affect production and perception.'' Instead, the results indicate that English-speaking listeners have abstract concepts of allophonic appropriateness that they can apply to both existing lexical items and nonwords when evaluating these items for the naturalness of their pronunciations.

1pSC30. Automatic gender identification. Alireza Afshordi Dibazar, Jim-Shih Liaw, and Theodore W. Berger (Dept. of Biomed. Eng., Univ. of Southern California, OHE-500, Los Angeles, CA 90089-1451)

In this paper, a new automatic gender identification method is proposed as a part of an Automatic Speaker Recognition (ASR) system. The short time raw speech signal (180–450 ms) was filtered by the nine-order Butterworth low pass filter and decomposed to different frequency bands by the wavelet filter bank analyzer. The energy of the 120 sub-bands was used as a feature vector and applied to a standard classifier. This classifier was trained by the gradient descent method with 1594 utterances spoken by various males and females. The system was tested with different 2542 utterances giving 99.2% correct classification rate. The high performance and simplicity of implementation are the characteristics of this system in comparison to the other methods. [Work supported by DARPA.]

1pSC31. Efficient automatic recognition of spoken digit strings. Douglas O'Shaughnessy and Hesham Tolba (INRS-Telecommunications, 900 de la Gauchetiere west, P.O. Box 644, Montreal, PQ H5A 1C6, Canada)

Automatic recognition of spoken digit sequences (such as credit card numbers) is now feasible even in speaker-independent applications over the telephone. However, all recognition tasks have lower performance in noisy conditions. If significant limitations are also imposed on the computational resources used for recognition, then robust speech recognition is still a significant challenge, even for a simple digit vocabulary. Since recognition of continuously spoken digits over telephone links is a very practical application, such recognition was investigated here under different conditions. Traditional hidden Markov model approaches with cepstral analysis were not used, because they are computationally intensive and have not always worked well under adverse acoustic conditions. Simpler spectral analysis was used, combined with a segmental approach. The analysis focuses on locations of spectral peaks, similar to formant tracking, but without the need to estimate peaks for all time frames. The limited nature of the vocabulary (i.e., ten digits) allows this simpler approach. High recognition accuracy is maintained despite being very efficient in both memory and computation. Recent progress will be reported.

Session 1pSP

Signal Processing in Acoustics: Nondestructive Testing: Signal Processing

David H. Chambers, Chair
Electrical Engineering, Lawrence Livermore National Laboratory, L-154, P.O. Box 808, Livermore, California 94551-5508

Contributed Papers

2:30

1pSP1. Ultrasonic pulse compression system with new approaches of signal processing. Khan M. Mahmud and Ryoji Ohba (Div. of Appl. Phys., Grad. School of Eng., Hokkaido Univ., Sapporo 060-8628, Japan, mah@eng.hokudai.ac.jp)

A nonscanning method for ultrasonic flaw detection technique is described. The system employs an *M*-sequence modulated ultrasonic wave as the excitation signal. Irrespective of the number of faults to be detected, the data acquisition system predeterminedly consists of a transmitter and a few receivers to be fixed at convenient locations. The cross-correlation function (CCF) between the original *M* sequence and the demodulated received sequence indicates the presence of faults in the form of sharp peaks. In order to clearly detect the expected peaks corresponding to the faults in the CCF, new signal processing techniques are proposed. The novel design approach of inverse filter and synchronous moving average are found to be very effective in eliminating the false peaks from the CCF while increasing the signal-to-noise ratio significantly. A way of using the CLEAN algorithm from a diverse field is proposed, which shows tremendous potential in ultrasonic NDT for the distinction of close peaks removing all sidelobes. Experimental results (on metal plate) demonstrate the capability of the system and confirm the feasibility of the approach under heavy ambient noise condition. An algorithm, only on the basis of travel time of the signal, to determine the exact location of a fault is described. [The authors would like to express their gratitude to the Ministry of Educaion, Japan, for the funds needed for this research.]

2:45

1pSP2. Space resolution improvement in pulse crack location. Vyacheslav V. Kazakov, Vitaly A. Zverev (Inst. of Appl. Phys., RAN, Nizhny Novgorod, Russia), and Alexander M. Sutin (Stevens Inst. of Technol., Hoboken, NJ 07030)

The new method of space resolution increasing in pulse location method was suggested by Professor Zverev is applied for space resolution improvement of ultrasonic pulse location technique. According to this method, the spectrum of the received signal is divided to the spectrum of the reference radiated signal and the complex inverse Fourier transformation is applied to the divided spectrum. This approach extends the frequency band used for analysis and highly increases the space resolution of crack location. The feasibility test was conducted with pulse ultrasonic device working with carried frequency about 3 MHz. The suggested processing was applied to the rectified signal so the frequency band of processing was restricted by 700 kHz. Even in this not optimal processing we observed the shortening of the signal reflected from a crack more than 20 times. The suggested technique was also used for spatial resolution improvement of nonlinear modulation crack imaging based on modulation of ultrasonic pulses by a crack vibration. [Work was partially supported by International Science Technical Center Grant No. 1369.]

3:00

1pSP3. Processing of ultrasonic measurements for nondestructive evaluation of critical optical components. Alan Meyer and James Candy (Lawrence Livermore Natl. Lab., P.O. Box 808, L-154, Livermore, CA 94551)

Real-time nondestructive evaluation of critical optics in high-energy, pulsed laser experiments is crucial from a safety as well as maintenance viewpoint. Fluence levels in short pulse, high-energy lasers can produce pits and cracks in the surfaces of the lasers' optical components. These flaws in the optical glass can adversely effect the production of the laser light or even result in a catastrophic failure of the optical component itself. Consequently, the identification and characterization of these flaws is critical. The production requirements of our target laser system mandate the automation of any inspection system. Consequently, it is necessary to develop an autonomous solution to process the reflected acoustic signals and, using this information, detect, localize, and characterize the optical flaws. In this paper we develop sophisticated signal and image processing techniques to automatically analyze laser-induced damage in optical glass. After preprocessing the raw channel measurement data from two orthogonal, narrow beamwidth, transducer arrays, a two-dimensional power image is created. A physics-based 2D matched filter is developed and used to detect and localize the flaws. An iterative solution is developed to sequentially search the resulting 2D image to extract and characterize the flaws.

3:15

1pSP4. Inversion methods for Lamb and Rayleigh wave surface holography. Kenbu Teramoto (Dept. of Mech. Eng., Saga Univ., Saga-shi 840-8502, Japan) and Hajime Yuasa (Akishima Lab., Mitsui Zosen, Inc., Tokyo, Japan)

Lamb and Rayleigh wave surface holographies have advantages for nondestructive testing of concrete and steel structures that is based on the use of generated surface acoustic waves that propagate over the concrete and steel and are reflected by external cracks. When the surface wave field is observed by a circular array of surface wave transducers, the proposed method that is called ''eigenfunction series expansion (EFSE)'' can reconstruct exact images of the scatterers within the array. This holographic imaging system has the following features: (1) An inversion via expansion in eigenfunction series produces much higher resolution than one which a back propagation produces. (2) An iso-azimuthal signal processing enables to design a first image reconstruction procedure. In this study, the computational process in EFSE and wave fields near the cracks are discussed and their physical meanings are investigated through FEM simulations. The surface wave transfer function by EFSE permits an effective regularization of the source field with cracks: this is possible because the transfer function includes several wave modes at the cracks and surface wave fields. In particular, it is clearly shown that the reconstructed image can be improved by adopting EFSE method that suppresses the spurious lobes.

3:30

1pSP5. Application of subband adaptive filtering techniques to ultrasonic detection in multilayers. Jie Mao and MingXuan Li (Inst. of Acoust., Chinese Acad. of Sci., 17 Zhongguancun St., Beijing 100080, PROC, jiem@sohu.com)

The subband adaptive filtering method used in the evaluation of multilayered steel–rubber debonding is discussed in the paper. Because the acoustic impedance of the steel differs far from that of the couplant water and the rubber, and the signal decay in the rubber is heavy, the energy of the signal reflected from the debonded rubber layers is very weak. The echoes caused by the steel/rubber interface almost mask the flaw echoes, which makes the extraction of the debonding echoes very difficult. In recent years, adaptive deconvolution [X. M. Jian, Chin. J. Acoust. **18**, 280–288 (1999)], wavelet inverse filtering, homomorphic filtering and echo cancellation techniques are adopted to extract the flaw echoes, but the signal reflected from the water/steel interface is ignored because of its large amplitude. A subband adaptive filtering method is discussed in the paper, where the subband decomposition is performed using a mutual wavelet packets decomposition on the criterion of maximizing the cross-correlation between the signals. The simulations on both synthetic and real signals are presented. [Work supported by National Natural Science Foundation of China.]

3:45

1pSP6. Acoustic scattering from laser damage spots. David H. Chambers, L. Peter Martin, and Graham Thomas (Lawrence Livermore Natl. Lab., P.O. Box 808 L-154, Livermore, CA 94551)

High energy laser systems sustain damage to critical optical components over their lifetime. This damage is manifested as localized pitting and cracking in the glass components. The damage pits have complicated morphology consisting of a hemispherical pit lined with crushed material, and surrounded by an outer zone with cracks. Monitoring of damage is necessary in some applications to prevent catastrophic failure of critical components. An acoustic system for monitoring optical damage spots was investigated to determine its applicability for the National Ignition Facility (NIF) at LLNL. Experiments and simulations of acoustic scattering from damage spots were performed to assess the ability to detect and size optical damage before failure becomes imminent. Pulses with a center frequency of 5 MHz were scattered from damage spots with nominal sizes of 0.5 mm to 7 mm. The amplitude of the measured return showed evidence of resonant scattering from the complicated morphology of the pits. This was confirmed with numerical simulations of scattering from simplified models of damage spots. This talk will emphasize the physics of scattering from damage spots and the comparison between experiments and numerical simulations.

4:00

1pSP7. A new acoustical method for inspection of industry piping. Ishtiaq Rasool Khan and Ryoji Ohba (Div. of Appl. Phys., Grad. School of Eng., Hokkaido Univ., Sapporo 060-8628, Japan)

A new acoustical method for inspection of the industry piping against development of corrosion and early stage cracks is presented. Impulse response of a defective pipe is obtained by launching an M-sequence at an end of the pipe, and cross-correlating it with the sequence received at the same end. The received sequence contains several replica of the M-sequence, reflected from various objects including the defects in the inner wall of the pipe. Therefore, the impulse response contains a peak corresponding to each reflection, the location of which determines the location of the reflecting object. Signal processing techniques are used to detect these very small peaks from heavy noise, and to distinguish the wanted peaks (due to defects) from unwanted peaks (due to other objects). A new technique is presented, which removes all kinds of unwanted peaks and other noise and significantly enhances the wanted peaks. This technique is proved much better than the classical noise removal techniques of moving average, subtraction and inverse filtering. A very simple experimental setup is presented and the results of successful detection of defects of different sizes in pipes of different dimensions are presented. [The presented research was financially supported by Japan Society for Promotion of Science.]

MONDAY AFTERNOON, 4 JUNE 2001 MONROE ROOM, 2:00 TO 4:45 P.M.

Session 1pUW

Underwater Acoustics: Matched Field Processing

Gregory J. Orris, Chair

Naval Research Laboratory, Code 7145, 4555 Overlook Avenue, SW, Washington, DC 20375-5000

Contributed Papers

2:00

1pUW1. Locating a sound source of unstable frequency by use of phase tracking. Travis L. Poole (MIT/WHOI, Joint Prog. in Oceanogr. and Oceanogr. Eng., Woods Hole Oceanogr. Inst., Woods Hole, MA 02543) and George V. Frisk (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

When the frequency of a cw source is known precisely, it is possible to relate the time rate-of-change of the measured phase to the time rate-of-change of the distance between source and receiver. However, when the source frequency is not known precisely, or when the source frequency is unstable, an unknown time dependence will remain after demodulation, causing the phase to appear to vary rapidly. Thus, the phase cannot be used to find the rate of separation of the source and receiver. By examining the difference between the phase measured by two independently moving receivers, it is possible to eliminate this unknown time dependence from the phase. In this talk, a source localization method based on this idea is presented and applied to data from a frequency-unstable source (near 20 Hz) in a shallow-water, low-frequency acoustics experiment. [Work supported by ONR and the WHOI Education Office.]

2:15

1pUW2. Matched-phase matched field processing: Searching for the phases. Gregory J. Orris and Michael Nicholas (U.S. Naval Res. Lab., Washington, DC 20375)

There have been several methods explored in the literature by which standard single-frequency matched-field techniques can be extended to multifrequency coherent algorithms. The foremost difficulty in implementing any coherent algorithm is the lack of reliable estimates of the relative phases of the processed frequency components. Thus many coherent techniques have side-stepped this issue by assuming an inter-frequency phase relationship that is either known *a priori*, or trivially zero. To address this

deficiency we have developed the matched-phase matched-field processor [Orris *et al.*, J. Acoust. Soc. Am. **107**, 2563–2575 (2000)]. The approach is to enlarge the search space from only the spatial dimensions to include the interfrequency phases. This processor has been applied to data from a source of known characteristics and was found to increase the spatial resolution while simultaneously increasing array performance significantly (received levels below −15 dB). We also present an application to a source of opportunity (i.e., a passing surface vessel) with unknown acoustic signature.

2:30

1pUW3. Information theory for matched field processing: preliminary results. John R. Buck (ECE Dept. and SMaST, Univ. of Massachusetts–Dartmouth, 285 Old Westport Rd., North Dartmouth, MA 02747)

Matched field processing (MFP) has traditionally been considered as an estimation problem in which the goal is to use the observed pressure field and an acoustic propagation model to estimate an unknown source location with the smallest possible variance. An alternative perspective is to divide the search region into cells whose size is dictated by logistical constraints, and attempt to assign the source to one of these cells. In this perspective, the goal of MFP is to minimize the probability of error in assigning the source to a grid cell, rather than minimizing the variance of the source location estimate. This perspective leads one to consider MFP as a communication problem in which the source transmits its location to the receiver array. Information theory provides techniques which can bound the amount of information available to the receiver in this problem. The amount of information observable has implications for the performance limits achievable by any MFP algorithm. Additionally, this perspective leads to methods of designing arrays for MFP which maximize the average information observed over the search space. [Work supported by ONR Young Investigator Program.]

2:45

1pUW4. Single snapshot array processing: The CLEAN algorithm applied to the pressure field. J. de Rosny, H. C. Song, and W. A. Kuperman (Scripps Inst. of Oceanogr., Univ. of California, San Diego, La Jolla, CA 92093-0701)

The CLEAN algorithm, originally developed in radio astronomy [Astron. Astrophys. Suppl. **15**, 417–426 (1974)] for intensity images, has been modified to process complex pressure signals. The iterative method allows one to find the position and strengths of a distribution of sources which are assumed to be grouped in small regions of the sky. Recently, we have applied the CLEAN algorithm successfully to matched field processing where the ambiguity surface corresponds to an image. The original CLEAN algorithm, however, is applied only to the field intensity (magnitude). In this paper, we apply the CLEAN algorithm to the complex pressure field received by a vertical array to enhance the matched field processing. This imaging method appears to be useful when just one or a few snapshots are available so that a reliable cross-spectral density matrix cannot be obtained. Numerical simulations will be presented for the modified CLEAN algorithm that extracts a weak source masked by a stronger source.

3:00

1pUW5. Improvement in matched field processing by the CLEAN algorithm. H. C. Song, J. de Rosny, and W. A. Kuperman (Scripps Inst. of Oceanogr., Univ. of California, San Diego, La Jolla, CA 92093-0238)

Matched field processing can be improved by the CLEAN algorithm [Astron. Astrophys. Suppl. **15**, 417–426 (1974)], introduced into radio astronomy for image restoration. The CLEAN algorithm is based on the *a priori* assumption that most images of radio sources are essentially blank sky with point-source components distributed in small regions. A simple iterative approach is employed to find the positions and strengths of these point sources. On convolving the point-source components with a ''clean'' beam, a final image known as the ''clean'' map is derived. In this paper, we apply this approach to match field processing where the ambiguity surface corresponds to the astronomical map. The initial ambiguity surface called a ''dirty'' map is evolved to a clean map through the iteration process. Numerical simulations demonstrate that matched field processing combined with the CLEAN algorithm provides better performance especially when a weak source is hardly visible due to sidelobes generated by a much stronger source.

3:15

1pUW6. Source localization through knowledge of the scattered field near two air-filled spheres: Numerical modeling results. Paul A. Lepper, Gerald L. D'Spain, and William S. Hodgkiss (Marine Physical Lab., Scripps Inst. of Oceanogr., La Jolla, CA 92093-0704)

Knowledge of the complex nature of the underwater sound field measured near two scatterers that are in the far field of an acoustic source may provide valuable information on the location of the source. Traditional techniques of source localization such as time-of-arrival (phase) and amplitude differences measured by multiple receivers are complicated when placed within the near field of an acoustic scatterer due to effects such as diffraction and scattering. The use of the interference pattern field due to the presence of the scatterers is proposed to enhance traditional source localization techniques. Numerical results are presented using a 2-D finite-difference time-domain (FDTD) scheme for scattering from two fluid-loaded, thin-walled, air-filled elastic spheres for multiple source aspect angles and frequencies. Matched field processing techniques are used to quantify the source localization capabilities. [Work supported by ONR.]

3:30–3:45 Break

3:45

1pUW7. Array element localization for towed marine seismic arrays. Stan E. Dosso and Michael Riedel (School of Earth and Ocean Sci., Univ. of Victoria, Victoria, BC V8W 3P6, Canada, sdosso@uvic.ca)

This work presents a new approach to array element localization (AEL) for the sensors of a towed marine seismic array based on regularized inversion of direct and bottom-reflected acoustic ray travel times picked from recorded seismic sections. Depth-sensor measurements at a number of points along the array are included as *a priori* estimates (with uncertainties) in the inversion. The smoothest array shape consistent with the acoustic data and prior estimates is determined by minimizing the array curvature or roughness. A smooth array shape is physically reasonable; in addition, minimizing array curvature provides an *a priori* model for the correlation between hydrophone positions that allows the estimation of both the offset and depth of hydrophones that record only one (or even no) acoustic arrival due to the shadowing effects of water-column refraction or reflection from arbitrary bathymetry. The AEL inversion is applied to a 102-sensor, 1.2-km towed array to correct receiver positions in the seismic velocity analysis of a seabed gas hydrate survey.

4:00

1pUW8. Matched field processing using a large aperture planar towed array. Jennifer Munro (MIT Lincoln Lab., 244 Wood St., Lexington, MA 02420), Arthur Baggeroer (MIT, Cambridge, MA 02139), James Ward, and Brian Tracey (MIT Lincoln Lab., Lexington, MA 02420)

Passive detection and localization in shallow water is complicated by high transmission loss and coherent multipath generated by bottom interaction. To compensate for this loss, one must achieve as much array gain as possible. The seismic community has developed large aperture arrays designed for low-frequency operation, which are on the order of 3–10 km by 0.5 km in size. This presentation explores the applicability of these arrays to passive detection and localization of underwater targets using matched field processing (MFP). This approach offers the advantages of a high number of sensors, large gain focusing, and three-dimensional localization through MFP when the propagation environment is well known. This talk presents simulated results of conventional MFP using planar

arrays in the presence of surface interferers. This presentation continues showing improvement through adaptive MFP. It then examines the problem of target motion. For these large arrays, robustness to target motion becomes a significant challenge. [This work was sponsored by the Defense Advanced Research Projects Agency under Air Force Contract No. F19628-00-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the authors and not necessarily endorsed by the United States Air Force.]

4:15

1pUW9. Matched-field processing in shallow water in the presence of internal waves. Catherine Stamoulis (Dept. of Ocean Eng., MIT, Cambridge, MA 02139 and Naval. Res. Lab., Washington, DC 20375)

The performance of matched-field processing algorithms in shallow water, in the presence of soliton packets, has been investigated through analysis of data from the SWARM95 experiment. In particular, acoustic data collected with a vertical array and a moving up-sweep source have been analyzed, in the frequency range 275–350 Hz. The source was placed below the surface mixed layer and its distance from the array varied between 2.5 and 27 km. Source localization using different matched-field algorithms has been performed, in order to assess the robustness of these processors in the presence of environmental mismatch and their degradation as a function of range. The primary result of the analysis is that these processors failed to localize the source accurately, even at the shortest range (2.5 km). Although broadband processing and inclusion of corrections for array tilt improved the results, the errors in source range (of the order of 1 km at short ranges and 5 to 10 km at long ranges) and depth remain significant. Alternative processing via mode filtering and subsequent matched-mode processing yielded more accurate results, particularly in cases were mode coupling was not an issue of concern.

4:30

1pUW10. Multitarget tracking using simulated annealing. David P. Knobles (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029), Tracianne Neilsen, and Robert A. Koch (Appl. Res. Labs., Univ. of Texas, Austin, TX 78713)

It is common for passive detection systems to have multiple detections over a given time interval. The focus of this research is the simultaneous inversion for environmental parameters and the localization of multiple moving broadband sources in shallow water environments. Individual sources are isolated using track beam methods on array subapertures. The subaperture track beam spectra are then input as element data into an inversion approach that uses simulated annealing to estimate both environmental and source parameters. Environmental parameters include those associated with the seabed and source parameters include source depth, initial range, initial bearing, speed, and course. A key concept is that of space-time processing manifested in the cost function by coherent sums of product spectra over frequency, receiver pairs, and time sequence. Insight into the problem is obtained with simulated data by examining the coupling between individual parameters, including the coupling between source and environmental parameters. [Research supported by ONR.]

NOTE: Attendance at this session requires payment of an additional registration fee.

MONDAY EVENING, 4 JUNE 2001 STATE BALLROOM, 7:00 TO 10:30 P.M.

Session 1eID

Interdisciplinary: Tutorial Lecture on Demonstration Experiments: Videos and Audios for Teaching Acoustics

Thomas D. Rossing, Cochair
Physics Department, Northern Illinois University, DeKalb, Illinois 60115

Uwe J. Hansen, Cochair
Department of Physics, Indiana University, Terre Haute, Indiana 47809

Invited Paper

7:00

1eID1. Demonstration experiments, videos and audios for teaching acoustics. Thomas D. Rossing (Northern Illinois Univ., Dekalb, IL 60115), Uwe J. Hansen (Indiana Univ., Terre Haute, IN 47809), and a team of Chicago-area physics teachers

Since the time of Pythagorus, if not before, experiments have been used by physics teachers to illustrate acoustic phenomena and important acoustic principles. We owe much to Michael Faraday who established a rich tradition of lecture demonstrations at the Royal Institution in London. Lord Rayleigh, in addition to his well-known papers on the theory of sound, published a number of short notes which he entitled ''Acoustical Observations.'' A potpourri of acoustics experiments will be demonstrated by a team of college and high school teachers with instructions for doing these experiments with simple equipment. Reproduction of these demonstrations may be useful in a formal educational setting for teaching acoustics, and should be useful also in any number of informal venues—including ones not always thought of as ''teaching.'' Videos and audios demonstrating acoustics also will be previewed.

Session 2aAAa

Architectural Acoustics: Halls for Music Performance . . . Another Two Decades of Experience 1982–2002 (Poster Session)

Christopher A. Storch, Cochair
Artec Consultants, Inc., 114 West 26th Street, New York, New York 10001

Timothy J. Foulkes, Cochair
Cavananaugh Tocci Associates, Inc., 327F Boston Post Road, Sudbury, Massachusetts 01776

Ian B. Hoffman, Cochair
The Talaske Group, Inc., 105 North Oak Park Avenue, Oak Park, Illinois 60301

Contributed Papers

All posters will be on display from 9:00 a.m. to 5:00 p.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 9:00 a.m. to 10:30 a.m. and contributors of even-numbered papers will be at their posters from 10:30 a.m. to 12:00 noon. To allow for extended viewing time, posters will remain on display until 5:00 p.m.

2aAAa1. Severance Hall, Cleveland, OH. Mark Holden (Jaffe Holden Acoustics, Inc., 114A Washington St., Norwalk, CT 06854, mholden@jhacoustics.com)

2aAAa2. Squires Recital Salon, Blacksburg, VA. Noral D. Stewart (Stewart Acoust. Consultants, P.O. Box 30461, Raleigh, NC 27622, noral@ix.netcom.com)

2aAAa3. Joan and Irving Harris Concert Hall. Elizabeth A. Cohen (Cohen Acoust., Inc., 132 S. Lucerne Blvd., Los Angeles, CA 90004, akustik@mediaone.net) and David Schwind (Charles M. Salter Assoc., Inc., San Francisco, CA 94914)

2aAAa4. Sibelius Hall, Lahti, Finland. Christopher Storch (Artec Consultants, Inc., 114 W. 26th St., 9th Fl., New York, NY 10001-6812)

2aAAa5. LG Arts Center, SangNam Hall, Seoul, Republic of Korea. Christopher Storch (Artec Consultants, Inc., 114 W. 26th St., 9th Fl., New York, NY 10001-6812)

2aAAa6. Kultur- und Kongresszentrum Luzern, Switzerland. Christopher Storch (Artec Consultants, Inc., 114 W. 26th St., 9th Fl., New York, NY 10001-6812)

2aAAa7. New Jersey Performing Arts Center, Newark, NJ. Christopher Storch (Artec Consultants, Inc., 114 W. 26th St., 9th Fl., New York, NY 10001-6812)

2aAAa8. The Morton H. Meyerson Symphony Center, Dallas, TX. Christopher Storch (Artec Consultants, Inc., 114 W. 26th St., 9th Fl., New York, NY 10001-6812)

2aAAa9. Symphony Hall, Birmingham, United Kingdom. Christopher Storch (Artec Consultants, Inc., 114 W. 26th St., 9th Fl., New York, NY 10001-6812)

2aAAa10. The Winspear Centre for Music. Christopher Storch (Artec Consultants, Inc., 114 W. 26th St., 9th Fl., New York, NY 10001-6812)

2aAAa11. The Chan Centre for the Performing Arts, Vancouver, British Columbia. Christopher Storch (Artec Consultants, Inc., 114 W. 26th St., 9th Fl., New York, NY 10001-6812)

2aAAa12. Auditorium de Dijon, Dijon, France. Christopher Storch (Artec Consultants, Inc., 114 W. 26th St., 9th Fl., New York, NY 10001-6812)

2aAAa13. Sala Sao Paulo, Sao Paulo, Brazil. Christopher Storch (Artec Consultants, Inc., 114 W. 26th St., 9th Fl., New York, NY 10001-6812)

2aAAa14. Hobby Center for the Performing Arts, Houston, TX. Mark Holden (Jaffe Holden Acoustics, Inc., 114A Washington St., Norwalk, CT 06854, mholden@jhacoustics.com)

2aAAa15. The Church of Jesus Christ of Latter-Day-Saints Assembly Hall, Salt Lake City, UT. Mark Holden (Jaffe Holden Acoustics, Inc., 114A Washington St., Norwalk, CT 06854, mholden@jhacoustics.com)

2aAAa16. Bass Performance Hall, Ft. Worth, TX. Mark Holden (Jaffe Holden Acoustics, Inc., 114A Washington St., Norwalk, CT 06854, mholden@jhacoustics.com)

2aAAa17. Oklahoma City Civic Center Hall, Oklahoma City, OK. Mark Holden (Jaffe Holden Acoustics, Inc., 114A Washington St., Norwalk, CT 06854, mholden@hacoustics.com)

2aAAa18. Tokyo International Forum, Tokyo, Japan. Mark Holden (Jaffe Holden Acoustics, Inc., 114A Washington St., Norwalk, CT 06854, mholden@jhacoustics.com)

2aAAa19. John F. Kennedy Center Concert Hall, Washington, DC. Mark Holden (Jaffe Holden Acoustics, Inc., 114A Washington St., Norwalk, CT 06854, mholden@jhacoustics.com)

2aAAa20. Zankel Recital Hall, at Carnegie Hall, New York City, NY. Mark Holden (Jaffe Holden Acoustics, Inc., 114A Washington St., Norwalk, CT 06854, mholden@jhacoustics.com)

2aAAa21. River Center for the Performing Arts, Columbus, GA. Mark Holden (Jaffe Holden Acoustics, Inc., 114A Washington St., Norwalk, CT 06854, mholden@jhacoustics.com)

2aAAa22. Murchison Performing Arts Center, University of North Texas, Denton, TX. Mark Holden (Jaffe Holden Acoustics, Inc., 114A Washington St., Norwalk, CT 06854, mholden@jhacoustics.com)

2aAAa23. Paul F. Sharp Concert Hall, University of Oklahoma, Norman, OK. (Acentech, Inc., Cambridge, MA)

2aAAa24. Rogers Performing Arts Center, Merrimack College, North Andover, MA. (Acentech, Inc., Cambridge, MA)

2aAAa25. Weis Auditorium, Bucknell University, Lewisburg, PA. (Acentech, Inc., Cambridge, MA)

2aAAa26. Center for Performing Arts, Univ. of Florida, Gainesville, FL. (Acentech, Inc., Cambridge, MA)

2aAAa27. Spivey Hall, Clayton College and State University, Morrow, GA. (Acentech, Inc., Cambridge, MA 02138)

2aAAa28. College of New Jersey Recital Hall, Ewing, NJ. (Acentech, Inc., Cambridge, MA 02138)

2aAAa29. Auckland Town Hall. Christopher W. Day, Ewen R. Kitto, and Keith O. Ballagh (Marshall Day Acoust., P.O. Box 5811, Wellesley St., Auckland, New Zealand, auckland@marshallday.co.nz)

2aAAa30. Bruce Mason Theatre—Takapuna. Christopher W. Day, Joanne O. Valentine, and Keith O. Ballagh (Marshall Day Acoust., P.O. Box 5811, Auckland, New Zealand)

2aAAa31. Saginaw Valley State University, Recital Hall, Saginaw, MI. Richard H. Talaske (The Talaske Group, Inc., 105 N. Oak Park Ave., Oak Park, IL 60301)

2aAAa32. Coronado Theatre, Rockford, IL. Richard H. Talaske and Jonathan P. Laney (The Talaske Group, Inc., 105 N. Oak Park Ave., Oak Park, IL 60301)

2aAAa33. Goshen College, Concert Hall and Recital Hall, Goshen, IN. Gary S. Madaras and Richard H. Talaske (The Talaske Group, Inc., 105 N. Oak Park Ave., Oak Park, IL 60301)

2aAAa34. Central Michigan University, Concert Hall, Mt. Pleasant, MI. Richard H. Talaske (The Talaske Group, Inc., 105 N. Oak Park Ave., Oak Park, IL 60301)

2aAAa35. California Center for the Arts, Concert Hall, Escondido, CA. Richard H. Talaske (The Talaske Group, Inc., 105 N. Oak Park Ave., Oak Park, IL 60301)

2aAAa36. Meymandi Concert Hall, Raleigh, NC. Edward Dugger (Kirkegaard Assoc., 801 W. Adams St., 8th Fl., Chicago, IL 60607)

2aAAa37. Verbrugghen Hall. Edward McCue (Kirkegaard Assoc., 954 Pearl St., Boulder, CO 80302)

2aAAa38. Concert Hall at the Clarice Smith Performing Arts Center. Edward McCue (Kirkegaard Assoc., 954 Pearl St., Boulder, CO 80302)

2aAAa39. Blumenthal Performing Arts Center, Charlotte, NC (1992). Joseph W. A. Myers (Kirkegaard Assoc., 801 W. Adams St., 8th Fl., Chicago, IL 60607, jmyers@kirkegaard.com)

2aAAa40. Auer Concert Hall. Edward McCue (Kirkegaard Assoc., 954 Pearl St., Boulder, CO 80302)

2aAAa41. Great Hall of the Washington Pavilion. Edward McCue (Kirkegaard Assoc., 954 Pearl St., Boulder, CO 80302)

2aAAa42. Jacoby Symphony Hall, Jacksonville, FL. Edward Dugger (Kirkegaard Assoc., 801 W. Adams St., 8th Fl., Chicago, IL 60607)

2aAAa43. Barshinger Center for Musical Arts, Lancaster, PA. Edward Dugger (Kirkegaard Assoc., 801 W. Adams St., 8th Fl., Chicago, IL 60607)

2aAAa44. Merrill Auditorium, Portland, ME. Edward Dugger (Kirkegaard Assoc., 801 W. Adams St., 8th Fl., Chicago, IL 60607)

2aAAa45. Broward Center for the Performing Arts, Ft. Lauderdale, FL (1991). Joseph W. A. Myers (Kirkegaard Assoc., 801 W. Adams St., 8th Fl., Chicago, IL 60607, jmyers@kirkegaard.com)

2aAAa46. Cerritos Center for the Performing Arts, Cerritos, CA (1993). Joseph W. A. Myers (Kirkegaard Assoc., 801 W. Adams St., 8th Fl., Chicago, IL 60607, jmyers@kirkegaard.com)

2aAAa47. Concert Hall at the Frank L. Moody Music Building. R. Lawrence Kirkegaard (Kirkegaard Assoc., 801 W. Adams St., 8th Fl., Chicago, IL 60607)

2aAAa48. Ordway Music Theatre, St. Paul, MN (1985). Joseph W. A. Myers (Kirkegaard Assoc., 801 W. Adams St., 8th Fl., Chicago, IL 60607, jmyers@kirkegaard.com)

2aAAa49. Corbett Auditorium, University of Cincinnati College-Conservatory of Music, Cincinnati, OH (1996). Joseph W. A. Myers (Kirkegaard Assoc., 801 W. Adams St., 8th Fl., Chicago, IL 60607, jmyers@kirkegaard.com)

2aAAa50. Orchestra Hall in Chicago, IL. Dawn Schuette (Kirkegaard Assoc., 801 W. Adams St., 8th Fl., Chicago, IL 60607)

2aAAa51. Aronoff Center for the Performing Arts. Dawn Schuette (Kirkegaard Assoc., 801 W. Adams St., 8th Fl., Chicago, IL 60607)

2aAAa52. Benedict Music Tent at the Aspen Music Festival. Dawn Schuette (Kirkegaard Assoc., 801 W. Adams St., 8th Fl., Chicago, IL 60607)

2aAAa53. North Shore Center for the Performing Arts. Dawn Schuette (Kirkegaard Assoc., 801 W. Adams St., 8th Fl., Chicago, IL 60607)

2aAAa54. Dewan filharmonik petronas (Petronas Concert Hall), Kuala Lumpur, Malaysia (1998). Joseph W. A. Myers (Kirkegaard Assoc., 801 W. Adams St., 8th Fl., Chicago, IL 60607, jmyers@kirkegaard.com)

2aAAa55. Maui Community Arts and Cultural Center, Kahului, HI (1994). Joseph W. A. Myers (Kirkegaard Assoc., 801 W. Adams St., 8th Fl., Chicago, IL 60607, jmyers@kirkegaard.com)

2aAAa56. Lagerquist Concert Hall, Pacific Lutheran University, Tacoma, WA (1995). Joseph W. A. Myers (Kirkegaard Assoc., 801 W. Adams St., 8th Fl., Chicago, IL 60607, jmyers@kirkegaard.com)

2aAAa57. Davies Symphony Hall, San Francisco, CA. Martha Larson (Kirkegaard Assoc., 801 W. Adams St., 8th Fl., Chicago, IL 60607)

2aAAa58. Peace Center for the Performing Arts, Greenville, SC (1990). Joseph W. A. Myers (Kirkegaard Assoc., 801 W. Adams St., 8th Fl., Chicago, IL 60607, jmyers@kirkegaard.com)

2aAAa59. Shepherd School of Music, Rice University, Houston, TX (1991). Joseph W. A. Myers (Kirkegaard Assoc., 801 W. Adams St., 8th Fl., Chicago, IL 60607, jmyers@kirkegaard.com)

2aAAa60. Heinz Hall, Pittsburgh, PA. Carl Giegold (Kirkegaard Assoc., 801 W. Adams St., 8th Fl., Chicago, IL 60607)

2aAAa61. Joseph M. Meyerhoff Symphony Hall, Baltimore. Carl Giegold (Kirkegaard Assoc., 801 W. Adams St., 8th Fl., Chicago, IL 60607)

2aAAa62. St. Luke Centre, London. Carl Giegold (Kirkegaard Assoc., 801 W. Adams St., 8th Fl., Chicago, IL 60607)

2aAAa63. Barbican Concert Hall, London. Carl Giegold (Kirkegaard Assoc., 801 W. Adams St., 8th Fl., Chicago, IL 60607)

2aAAa64. Jordan Hall at the New England Conservatory. Scott Pfeiffer, Clete Davis, and Larry Kirkegaard (Kirkegaard Assoc., 801 W. Adams St., 8th Fl., Chicago, IL 60607)

2aAAa65. Jemison Concert Hall, University of Alabama, Birmingham, AL (1996). Joseph W. A. Myers (Kirkegaard Assoc., 801 W. Adams St., 8th Fl., Chicago, IL 60607, jmyers@kirkegaard.com)

2aAAa66. Casals Hall, Tokyo, Japan. Toshiko Fukuchi, Keiji Oguchi, Katsuji Naniwa, and Minoru Nagata (Nagata Acoustics, Inc., Hongo Segawa Bldg. 3F, 2-35-10, Hongo, Bunkyo-Ku, Tokyo 113-0033, Japan, fukuchi@nagata.co.jp)

2aAAa67. Sumida Triphony Hall, Tokyo, Japan. Toshiko Fukuchi, Hideo Nakamura, and Minoru Nagata (Nagata Acoustics, Inc., Hongo Segawa Bldg. 3F, 2-35-10, Hongo, Bunkyo-Ku, Tokyo 113-0033, Japan, fukuchi@nagata.co.jp)

2aAAa68. Nara Centennial Hall, Nara, Japan. Satoru Ikeda, Chiaki Ishiwata, Motoo Komoda, and Minoru Nagata (Nagata Acoustics, Inc., Hongo Segawa Bldg. 3F, 2-35-10, Hongo, Bunkyo-ku, Tokyo)

2aAAa69. The Sogakudo Concert Hall of the Tokyo National University of Fine Arts and Music, Tokyo, Japan. Tosiko Fukuchi, Katsuji Naniwa, and Minoru Nagata (Nagata Acoustics, Inc., Hongo Segawa Bldg. 3F, 2-35-10 Hongo, Bunkyo-Ku, Tokyo 113-0033, Japan, fukuchi@nagata.co.jp)

2aAAa70. Tokyo Metropolitan Art Space, Tokyo, Japan. Satoru Ikeda, Keiji Oguchi, Makoto Ino, and Minoru Nagata (Nagata Acoustics, Inc., Hongo Segawa Bldg. 3F, 2-35-10, Hongo, Bunkyo-Ku, Tokyo, Japan)

2aAAa71. The Toppan Hall, Tokyo, Japan. Tosiko Fukuchi, Katsuji Naniwa, and Minoru Nagata (Nagata Acoustics, Inc., Hongo Segawa Bldg. 3F, 2-35-10 Hongo, Bunkyo-Ku, Tokyo 113-0033, Japan, fukuchi@nagata.co.jp)

2aAAa72. Tsuda Hall, Tokyo, Japan. Satoru Ikeda, Katuji Naniwa, and Minoru Nagata (Nagata Acoustics, Inc., Hongo Segawa Bldg. 3F, 2-35-10, Hongo, Bunkyo-Ku, Tokyo)

2aAAa73. Philia Hall, Tokyo, Japan. Akira Ono, Katsuji Naniwa, Takesi Yamamoto, and Minoru Nagata (Nagata Acoustics, Inc., Hongo Segawa Bldg. 3F, 2-35-10 Hongo, Bunkyo-Ku, Tokyo 113-0033, Japan, ono@nagata.co.jp)

2aAAa74. Katsushika Symphony Hills, Tokyo, Japan. Satoru Ikeda, Akira Ono, Katsuji Naniwa, and Minoru Nagata (Nagata Acoustics, Inc., Hongo Segawa Bldg. 3F, 2-35-10, Hongo Bunkyo-ku, Tokyo 113-0033, Japan, ikeda@nagata.co.jp)

2aAAa75. Asahikawa City Taisetsu Crystal Hall, Hokkai-do, Japan. Akira Ono, Makoto Ino, and Minoru Nagata (Nagata Acoustics, Inc., Hongo Segawa Bldg. 3F, 2-35-10, Hongo Bunkyo-ku, Tokyo 113-0033, Japan, ono@nagata.co.jp)

2aAAa76. Kioi Hall, Tokyo, Japan. Akira Ono, Katsuji Naniwa, and Minoru Nagata (Nagata Acoustics, Inc., Hongo Segawa Bldg. 3F, 2-35-10, Hongo Bunkyo-ku, Tokyo 113-0033, Japan, ono@nagata.co.jp)

2aAAa77. Kumamoto Prefectural Theatre, Kumamoto, Japan. Satoru Ikeda, Hideo Nakamura, and Minoru Nagata (Nagata Acoustics, Inc., Hongo Segawa, Bldg. 3F, 2-35-10 Hongo, Bunkyo-Ku, Tokyo 113-0033, Japan, ikeda@nagata.co.jp)

2aAAa78. Concert hall ATM in Art Tower Mito, Mito, Japan. Keiji Oguchi, Hideo Nakamura, and Minoru Nagata (Nagata Acoustics, Inc., Hongo Segawa Bldg. 3F, 2-35-10 Hongo, Bunkyo-Ku, Tokyo 113-0033, Japan, oguchi@nagata.co.jp)

2aAAa79. Okayama Symphony Hall, Okayama, Japan. Satoru Ikeda, Keiji Oguchi, Hideo Nakamura, and Minoru Nagata (Nagata Acoustics, Inc., Hongo Segawa, Bldg. 3F, 2-35-10 Hongo, Bunkyo-Ku, Tokyo 113-0033, Japan, ikeda@nagata.co.jp)

2aAAa80. The Harmony Hall, Matsumoto, Japan. Toshiko Fukuchi, Keiji Oguchi, Katsuji Naniwa, and Minoru Nagata (Nagata Acoustics, Inc., Hongo Segawa, Bldg. 3F, 2-35-10, Hongo, Bunkyo-Ku, Tokyo 113-0033, Japan, fukuchi@nagata.co.jp)

2aAAa81. Fukushima Concert Hall, Fukushima, Japan. Yasuhisa Toyota, Keiji Oguchi, Katsuji Naniwa, and Minoru Nagata (Nagata Acoustics, Inc., Hongo Segawa Bldg. 3F, 2-35-10, Hongo, Bunkyo-Ku, Tokyo 113-0033, Japan, toyota@nagata.co.jp)

2aAAa82. Nagaoka Lyric Hall, Nagaoka, Japan. Yasuhisa Toyota, Chiaki Ishiwata, Katsuji Naniwa, and Minoru Nagata (Nagata Acoustics, Inc., Hongo Segawa Bldg. 3F, 2-35-10, Hongo, Bunkyo-Ku, Tokyo 113-0033, Japan, toyota@nagata.co.jp)

2aAAa83. Sapporo Concert Hall ''Kitara,'' Sapporo, Japan. Yasuhisa Toyota, Katsuji Naniwa, and Minoru Nagata (Nagata Acoustics, Inc., Hongo Segawa Bldg. 3F, 2-35-10, Hongo, Bunkyo-Ku, Tokyo 113-0033, Japan, toyota@nagata.co.jp)

2aAAa84. Suntory Hall, Tokyo, Japan. Yasuhisa Toyota, Akira Ono, Katsuji Naniwa, and Minoru Nagata (Nagata Acoustics, Inc., Hongo Segawa Bldg. 3F, 2-35-10, Hongo, Bunkyo-Ku, Tokyo 113-0033, Japan, toyota@nagata.co.jp)

2aAAa85. Akiyoshidai International Art Village, Yamaguchi, Japan. Keiji Oguchi, Makoto Inoh, and Minoru Nagata (Nagata Acoustics, Inc., Hongo Segawa Bldg. 3F, 2-35-10, Hongo, Bunkyo-Ku, Tokyo 113-0033, Japan, oguchi@nagata.co.jp)

2aAAa86. Ishikawa Concert Hall, Kanazawa, Japan. Keiji Oguchi, Satoru Ikeda, Masaya Uchida, and Minoru Nagata (Nagata Acoustics, Inc., Hongo Segawa Bldg. 3F, 2-35-10, Hongo, Bunkyo-Ku, Tokyo 113-0033, Japan, oguchi@nagata.co.jp)

2aAAa87. Kyoto Concert Hall, Kyoto, Japan. Yasuhisa Toyota, Keiji Oguchi, Chiaki Ishiwata, and Minoru Nagata (Nagata Acoustics, Inc., Hongo Segawa Bldg. 3F, 2-35-10, Hongo, Bunkyo-Ku, Tokyo 113-0033, Japan, oguchi@nagata.co.jp)

2aAAa88. Queensland Conservatorium of Music Theater, Brisbane, Australia. Keiji Oguchi, Yasuhisa Toyota, Katsuji Naniwa, and Minoru Nagata (Nagata Acoustics, Inc., Hongo Segawa Bldg. 3F, 2-35-10, Hongo, Bunkyo-Ku, Tokyo 113-0033, Japan, toyota@nagata.co.jp)

2aAAa89. Mitaka City Arts Center, Tokyo, Japan. Takayuki Hidaka (Takenaka R&D Institute, 1-5-1, Otsuka Inzai Chiba 270-1395, Japan, hidaka.takayuki@takenaka.co.jp)

2aAAa90. New National Theater, Tokyo, Japan. Takayuki Hidaka (Takenaka R&D Institute, 1-5-1, Otsuka Inzai Chiba 270-1395, Japan, hidaka.takayuki@takenaka.co.jp)

2aAAa91. Hamarikyu Asahi Hall, Tokyo, Japan. Takayuki Hidaka (Takenaka R&D Institute, 1-5-1, Otsuka Inzai Chiba 270-1395, Japan, hidaka.takayuki@takenaka.co.jp)

2aAAa92. Tokyo Opera City (TOC) Concert Hall, Japan. Takayuki Hidaka (Takenaka R&D Institute, 1-5-1, Otsuka Inzai Chiba 270-1395, Japan, hidaka.takayuki@takenaka.co.jp)

2aAAa93. Rocky Hill High School Auditorium—Rocky Hill, CT. Bennett M. Brooks (Brooks Acoust. Corp., 27 Hartford Turnpike, Vernon, CT 06066)

2aAAa94. Crowell Concert Hall, Wesleyan University Center for the Arts, Middletown, CT. Bennett M. Brooks (Brooks Acoust. Corp., 27 Hartford Turnpike, Vernon, CT 06066)

2aAAa95. Kavli Theater at Thousand Oaks Civic Arts Plaza, Thousand Oaks, CA. Ron McKay and Mark Rothermel (McKay Conant Brook, Inc., 5655 Lindero Canyon Rd., Ste. 325, Westlake Village, CA 91362, mrothermel@mcbinc.com)

2aAAa96. Royce Hall at University of California Los Angeles, Westwood, CA. Ron McKay and Mark Rothermel (McKay Conant Brook, Inc., 5655 Lindero Canyon Rd., Ste. 325, Westlake Village, CA 91362, mrothermel@mcbinc.com)

2aAAa97. Popejoy Hall at University of New Mexico, Albuquerque, NM. Ron McKay and Mark Rothermel (5655 Lindero Canyon Rd., Ste. 325, Westlake Village, CA 91362, mrothermel@mcbinc.com)

2aAAa98. Kaul Auditorium, Reed College, Portland, OR. Russ Altermatt (Altermatt Assoc., Inc., 522 SW Fifth Ave., Ste. 1200, Portland, OR 97204, raltermatt@altermatt.com)

2aAAa99. Acoustical renovation of The Orpheum Theatre, Vancouver, Canada. John O'Keefe (Aercoustics Eng. Ltd., 50 Ronson Dr., Toronto M9W 1B3, Canada), Gilbert Soloudre (Carleton Univ., Ottawa K1S 5B6, Canada), and John Bradley (Inst. for Res. in Construction, Natl. Res. Council of Canada, Ottawa K1A 0R6, Canada)

2aAAa100. The chamber music/lecture hall for the Royal Library in Copenhagen, Denmark. Anders Christian Gade (Acoust. Technol., Oersted DTU, Tech. Univ. of Denmark, Bldg. 352, DK-2800 Lyngby, Denmark)

2aAAa101. Radiohuset, Studio 1, Copenhagen. Jens H. Rindel, Anders C. Gade (Oersted-DTU, Acoust. Technol., DTU Bldg. 352, DK-2800 Lyngby, Denmark), and Niels V. Jordan (Jordan Akustik, Herslevvej 19, Gevninge, DK-4000 Roskilde, Denmark)

2aAAa102. Evans Concert Hall, Connecticut College. Timothy Foulkes (Cavanaugh Tocci Assoc., 327F Boston Post Rd., Sudbury, MA 01776)

2aAAa103. Performance Hall, Providence, RI. Lincoln B. Berry (Cavanaugh Tocci Assoc., 327 F. Boston Post Rd., Sudbury, MA 01776, berry@cavtocci.com)

TUESDAY MORNING, 5 JUNE 2001 RED LACQUER ROOM, 9:00 A.M. TO 12:00 NOON

Session 2aAAb

Architectural Acoustics, Engineering Acoustics and Noise: Acoustical Test Facilities (Poster Session)

Charles T. Moritz, Chair
Blachford, Inc., 1400 Nuclear Drive, West Chicago, Illinois 60185

Contributed Papers

All posters will be on display from 9:00 a.m. to 12:00 noon. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 9:00 a.m. to 10:30 a.m. and contributors of even-numbered papers will be at their posters from 10:30 a.m. to 12:00 noon.

2aAAb1. Johns Manville Acoustical Laboratory. Brandon D. Tinianov and Francis J. Babineau (Johns Manville Tech. Ctr., 10100 W. Ute Ave., Littleton, CO 80127)

Due to their long history in acoustics materials, the Johns Manville Corporation has maintained an acoustical laboratory in constant operation since the 1930s. The laboratory complex is accredited under NVLAP for a broad range of ASTM and ISO test methods. For classic architectural acoustics evaluation, the lab has a reverberant test suite with a 142-m^3 source chamber and a 316-m^3 receive chamber. The chambers' common opening allows for 2.8 by 4.3-m samples and is qualified from 100–10 000 Hz. Analysis of complex, three-dimensional systems conducted in a 176-m^3 hemi-anechoic chamber qualified to 160 Hz. Both intensity and pressure measurements are possible via a 5 degree-of-freedom robotic positioning system. Research of open office acoustics and office furniture is performed in a 111- (to the grid, 142 total) m^3 chamber designed to accept a suspended ceiling and vary its reverberation time from <0.3 to 2.3 s. In addition to these facilities, several benchtop and analytical resources are available.

2aAAb2. Acoustic test facilities of the Underwater Sound Reference Division. A. Lee Van Buren, Robert M. Drake, and Kirk E. Jenne (Naval Undersea Warfare Ctr., 1176 Howell St., Newport, RI 02841-1708)

The acoustic test facilities of the Underwater Sound Reference Division (USRD) of the Naval Undersea Warfare Center will be described. The USRD serves as the U.S. standardizing activity in the area of underwater acoustic measurements, as the National Institute of Standards and Technology does in other areas. It is the Navy's primary activity for underwater acoustic calibration, test, evaluation, and reference measurements on transducers and related devices and materials. In this role the USRD maintains specialized measurement facilities with the capacity to simulate real-world ocean environments. It also maintains transducer standards to help ensure the accuracy of measurements made both at USRD and elsewhere. This presentation will describe the Acoustic Pressure Tank Facility, the Low-Facility Facility (with its three test vessels), the Acoustic Open Tank Facility, the Long-Line Hydrophone Calibrator, the Conical Shock Tube Facility, the Leesburg Facility, and the Transducer Standards Loan Program.

2aAAb3. Acoustic measurement facilities at the Applied Research Laboratory. David L. Bradley (Appl. Res. Lab., Penn State Univ., P.O. Box 30, State College, PA 16804-0030)

Acoustic measurement facilities, ranging from a water tunnel that is also listed as an ASME Historic Landmark to a new ultrasonic measurement system, are illustrated and described with a brief explanation as to size, measurement equipment available, and frequency ranges that are typically employed. The facilities include a high-pressure tank for hydrophone tests ''at depth,'' the ultrasonics tank, an anechoic tank for hydrophone calibration, the water tunnel, an air acoustic anechoic room and a large, multipurpose underwater measurement facility.

2aAAb4. Energistics Laboratory facility. Victoria Cerami (Cerami & Assoc., Inc., 317 Madison Ave., Ste. 220, New York, NY 10017) and Jack Stegall (Energistics Lab., Houston, TX 77257)

Energistics Laboratory in Houston, Texas is a leading laboratory for the testing of HVAC equipment. For over 15 years, this facility has ensured the highest standards in leading-edge HVAC technology and architectural testing capabilities. Testing capabilities include both industry standard rating procedures, and mock-up testing to simulate field conditions. The laboratory is open to developers, owners, architects, engineers, general contractors, manufacturers and others who require independent component testing and evaluation.

2aAAb5. Designing acoustic facilities for education. Daniel R. Raichel (Raichel Technol. Group and CUNY Grad. Ctr., 2727 Moore Ln., Fort Collins, CO 80526)

An acoustic laboratory in a university generally serves two functions: education and research. In planning, budgetary considerations constitute a decisive factor. A low budget can result in the laboratory being relegated to fairly simple measurements but some degree of ingenuity can widen the scope of experimentation, as has been proven in Take Five demonstrations at ASA meetings. A more elaborate facility would enable measurements in the ultrasonic region in addition to those in the audio range. The cost of maintenance and eventual upgrading should be figured into the planning and design of any facility. Necessary instrumentation include means of measuring sound pressure levels and intensities, executing spectral analyses (preferably in real time), generating and synthesizing signals, performing fairly standard measurements for sound absorption coefficients; and it would be desirable that specific enclosure conditions be simulated (viz., through the use of anechoic and reverberation chambers). Research programs generally require more sophisticated instrumentation, but fortunately fairly recent advances in microelectronics permit integration of acoustic monitoring functions into personal computers at considerably lower cost than would be the case if individual measurement instruments were purchased separately.

2aAAb6. NASA Glenn Research Center Acoustical Testing Laboratory. Beth A. Cooper (NASA, John H. Glenn Res. Ctr. at Lewis Field, M.S. 86-10, 21000 Brookpark Rd., Cleveland, OH 44135, beth.a.cooper@grc.nasa.gov)

The Acoustical Testing Laboratory (ATL) at the National Aeronautics and Space Administration (NASA) John H. Glenn Research Center in Cleveland, Ohio supports the low-noise design of science experiment payloads developed for the International Space Station (ISS). The ATL consists of a 100-Hz vibration-isolated hemianechoic test chamber with 21-ft by 17-ft by 17-ft (h) interior working dimensions and removable floor wedges that allow the facility to be configured as either a hemianechoic or fully anechoic chamber. A separate, sound-isolated control room with outside dimensions of 23-ft by 11-ft by 12-ft (h) doubles as a noise control enclosure when testing articles that require remote connections to high-noise support equipment and services. These characteristics, along with very low design background sound levels, enable the acquisition of accurate and repeatable acoustical measurements on test articles, up to a full ISS rack in size, that produce very little noise. The ATL provides a comprehensive array of acoustical testing services, including sound power level in accordance with precision and engineering grade standards. These capabilities employ a PC-based acoustical data acquisition system that has been customized to perform simultaneous acquisition and real-time analysis of multiple channels of acoustical signals.

2aAAb7. USG Corp. Research and Technology Center's Acoustical Testing Facility. Stephen W. Payne, Jr. (USG Corp. Res. and Technol. Ctr., 700 N. Hwy. 45, Libertyville, IL 60048)

The Acoustical Testing Facility at the USG Corp. Research and Technical Center was commissioned in 1996. The Facility contains structurally isolated test chambers housed within an exterior shell. The Facility includes chambers for performing sound tansmission loss tests per ASTM E90, sound absorption tests per ASTM C423, ceiling attenuation tests per ASTM E1414 and interzone attenuation tests per ASTM E1111. In addition to testing per ASTM standards, tests can also be performed to ISO Standards. The chambers are designed to provide maximum isolation from external noise. The chamber walls, roof and floor are built of a minimum of 8-in. (200-mm) solid concrete or filled concrete block. In addition the reverberation chamber is mounted on an independent spring system. The Facility is accredited under the National Voluntary Laboratory Accreditation Program (NVLAP) and under the Underwriters Laboratories Client Data Test Program. The main function of this Facility is to provide data for new product and systems development and to provide testing under the UL Acoustical labeling program.

2aAAb8. Riverbank Acoustical Laboratories. David L. Moyer (1512 S. Batavia Ave., Geneva, IL 60134, dmoyer@iitri.org)

Riverbank Acoustical Laboratories (RAL) is located in Geneva, Illinois. Constructed in 1918, RAL became the world's first independent laboratory dedicated to the science of architectural acoustics. Riverbank owner, philanthropist Colonel George Fabyan, built the laboratory for its designer, professor Wallace Clement Sabine. The scientific efforts of this renowned Harvard physicist earned him two distinguished titles, acknowledged today as the father of the science of architectural acoustics and the first ''modern day'' acoustical consultant. The unit of absorption is a Sabin. When the professor died in 1919, his distant cousin, Dr. Paul Earls Sabine, became RAL's director. Retired in 1947, Dr. Sabine turned the operations over to the IIT Research Institute. RAL provides a wide range of acoustical testing services for over 600 clients. The majority of RAL tests are industrial related. RAL provides clients with an accredited, independent testing facility, not associated with any manufacturer or product. Part of RAL's outstanding credibility reflects upon their policies of maintaining strict confidence and proprietary rights for all clients. RAL is accredited by NVLAP, The City of Los Angeles, and others. Also of interest to many clients is RAL's museum dedicated to the Science of Architectural Acoustics. The ASA officially sanctioned the museum in 1989.

2aAAb9. Qualification of WEAL Reverberation Chamber. Gary E. Mange (Western Electro-Acoust. Lab., 1711 16th St., Santa Monica, CA 90404)

Western Electro-Acoustic Laboratory (WEAL) has recently qualified a new reverberation chamber. Qualification measurements were performed for absorption testing (ASTM C423) and sound power testing (ANSI S12.31 and S12.32). Results are shown at various stages of the qualification.

2aAAb10. Georgia Institute of Technology—Integrated Acoustics Laboratory. Kenneth A. Cunefare and Van Biesel (G. W. Woodruff School of Mech. Eng., The Georgia Inst. of Technol., Atlanta, GA 30332-0405)

The Integrated Acoustics Laboratory of Georgia Tech's School of Mechanical Engineering is a state-of-the-art facility dedicated to acoustic and vibration research. The current centerpiece of the lab is an anechoic chamber with interior dimensions of 17 ft $\times$ 17 ft $\times$ 12 ft. Laboratory instrumentation includes a 32-channel HP VXI data acquisition system, full LMS CADA-X data acquisition and analysis package, HP-VEE, a two-channel Siglab analyzer, Polytec PSV-200 scanning laser Doppler vibro-

meter, and a host of microphones, accelerometers, shakers, and load cells. The laboratories also feature a brake dynamometer. Two finite-element modeling stations are equipped with MSC/NASTRAN and COMET/Acoustics software for the complete modeling of vibration and acoustics. The laboratory will be expanded this year to include a reverberation room, which will be attached to the existing anechoic chamber by a hatch for transmission loss testing. The expansion will also add a semi-anechoic chamber to the lab.

2aAAb11. Noise and structural dynamics test facilities at Boeing. Wendell Miller, Donald Boston, and Charlie Pickrel (The Boeing Co., Seattle, WA 98124)

The noise and structural dynamics laboratories at Boeing provide a wide range of test and measurement services to the Boeing Company. Test data from these laboratories support all phases of the product life cycle across a diverse line of products and applications. The noise laboratory facilities include a low-speed free-jet acoustic wind tunnel, several anechoic and reverberation test chambers, a critical listening facility, and a materials test center. These facilities are supported with a network of data systems for in-lab testing and a variety of transportable data systems for field- and airplane-based testing. Structural dynamics laboratory facilities include large strongbacks and structural floors for component vibration testing, sonic fatigue test facilities, and vibration test facilities. These facilities are supported by a network of dedicated data systems for a wide range of modal, shock, vibration, and fatigue testing. Field tests are supported by a wide range of portable data systems and instrumentation trailers capable of large channel count measurements. This work will provide an overview of the test facilities and measurement capabilities of these laboratories.

2aAAb12. The new Blachford acoustics laboratory. Charles Moritz (Blachford, Inc., 1400 Nuclear Dr. W., Chicago, IL 60185)

In 1968, HL Blachford, Inc. constructed their first acoustics laboratory next to their Troy, Michigan plant. This laboratory consisted of a 216-m^3 reverberation room with an adjacent large hemianechoic room approximately 12 m long, 5 m wide, and 4.5 m high. After several ownership changes, a plant that used to belong to HL Blachford, Inc. was purchased in 1995 by HL Blachford Ltd. of Canada, and became the new headquarters of a new company called Blachford, Inc. It quickly became obvious that a new laboratory was required for vehicle testing and acoustical material development. The design of a new facility began in 1997 and was based on the old laboratory; however, the new design incorporates design changes to increase the capacity, capabilities, and quality of the facility. The new laboratory is currently under construction and features a larger hemi-anechoic room, approximately 15 m long, 9 m wide, and 6 m high, and a nonadjacent 200-m^3 reverberation room with several horizontal and vertical test sections for vehicle component and material testing.

TUESDAY MORNING, 5 JUNE 2001 PDR 9, 8:00 TO 11:50 A.M.

Session 2aAB

Animal Bioacoustics and Signal Processing in Acoustics: Role of Signal Processing in Understanding Echolocation

James A. Simmons, Chair

Department of Neuroscience, Brown University, Box 1953, Providence, Rhode Island 02912

Chair's Introduction—8:00

Invited Papers

8:05

2aAB1. Neuromorphic system for real-time biosonar signal processing. R. Timothy Edwards (Johns Hopkins Univ., APL 4-234, Laurel, MD 20723-6099, tim@stravinsky.jhuapl.edu) and Gert Cauwenberghs (Johns Hopkins Univ., Baltimore, MD 21218-2686, gert@jhu.edu)

Real-time classification from active sonar echolocation requires tremendous computational power, yet is performed effortlessly by bats and dolphins. Special-purpose, yet programmable and reconfigurable hardware tailored to the parallel distributed nature of the computation performed in biology, is presented. The system contains a front-end cochlear processor implemented as an analog filterbank, and a pattern recognition engine classifying the front-end time-frequency features. The reconfigurable front-end supports up to 34 bandpass channels spanning the 10 to 100 kHz range typical of bat and dolphin sonar. A support vector machine (SVM) classifies the features. In an alternative realization, an integrate-and-fire neural network uses the zero-crossings of the front-end bandpass channel outputs. The internal variables may be represented in numerous ways by reprogramming the hardware, giving the system great flexibility in representing features to match the methods used by various biosonar algorithms. Experimental results demonstrate the capabilities of the processor, classifying the CSS (Panama City) linear frequency modulation (LFM2) data in real-time. Implementation issues for alternative representations of biosonar signals are also discussed. [Work supported by ONR N00014-01-1-0315.]

8:25

2aAB2. Signal processing by echolocating dolphins. Whitlow W. L. Au (Hawaii Inst. of Marine Biol., P.O. Box 1106, Kailua, HI 96734)

Many different types of target discrimination experiments have been performed with echolocating dolphins in order to understand and model their echolocation capabilities. However, not many experiments have been performed that measured differential performance as a function of the magnitude of a single target variable, thus making it very difficult to extract limits of various auditory discrimination capabilities. Nevertheless, sufficient information exists that could provide some insight on important echo cues utilized

by dolphins and what the signal processing may be. Furthermore, other types of experiments emulating some dolphin experiments have also been conducted. Stretched versions of echoes from target used in dolphin experiments have been used in human listening experiments to determine salient cues present in the echoes. Neural network experiments have also been conducted using echoes from the same target used in the dolphin experiments. The most pertinent of these experiments involving dolphins, humans, and neural networks will be discussed in order to gain insight into important target cues and the manner in which these cues might be processed by echolocating dolphins. The discussion will be conducted in light of the results of a recent experiment in which the auditory filter shape of a dolphin for different frequencies was determined.

8:45

2aAB3. What kinds of auditory mechanisms explain biosonar perception by FM bats? James A. Simmons, Mark I. Sanderson, and Kyler M. Eastman (Dept. of Neurosci., Box 1953, Brown Univ., Providence, RI 02912, james_simmons@brown.edu)

A biological SCAT model of target ranging in FM-bat biosonar integrates (1) acoustic ecology with algorithms identified from (2) peripheral auditory representations and (3) central auditory processing to account for biosonar performance. The bat's auditory spectrograms incorporate a near-ideal time-frequency compromise and can outperform a simple matched filter for detection of natural targets. Neural coding of delay appears to have submicrosecond accuracy at brain-stem, midbrain, and cortical levels, and can explain ranging and azimuth performance for single targets. In contrast, inner-ear integration-time convolution required to match echo reception to the structure of natural targets blurs delay resolution to hundreds of microseconds, and this is retained in the recovery time of brain-stem responses. Analysis of field performance and laboratory tests both show the bat's delay resolution to be $\sim 2-5$ ms, however, midbrain responses constitute delay lines that can support tracking of prey and both coarse target ranging and deconvolution for fine resolution down to 6 ms. Cortical coincidence-detecting neurons test all combinations of pulse-echo delay and spectral profile contingencies to display coarse and fine range. Their characteristics predict bandwidth-squared dependence of submicrosecond hyperresolution, which is confirmed in jittered-echo experiments. The physical embodiment of individual targets in neural responses appears to be dynamic rather than by place.

9:05

2aAB4. Roles of neural oscillation in time domain processing in the bat inferior colliculus. Alexander V. Galazyuk and Albert S. Feng (Beckman Inst., Univ. of Illinois, Urbana, IL 61801)

Central auditory neurons in echolocating bats exhibit pulse-echo delay-tuned responses. Sullivan (1982) proposed that paradoxical latency shift (PLS), characterized by an increase in response latency to loud sounds, is important for this attribute. At present, the mechanism underlying PLS is unclear. The goal of the present study was to identify the mechanism underlying PLS. The responses of 92 neurons in the inferior colliculus of little brown bats to brief tone pulses at the unit's CF over a wide range of sound levels were studied. Of these, 16 neurons displayed unit-specific periodic oscillatory discharges at high sound levels with a characteristic period of 1.3–6.7 ms. The 27 neurons exhibited unit-specific PLS, with quantal latency shift of 1.2–8.2 ms. In 14 neurons showing PLS, unit's responses before, during and after iontophoretic application of bicuculline were investigated. Application of bicuculline abolished the PLS and transformed it into periodic discharges, suggesting that neural oscillation in combination with ordinary inhibition may be responsible for PLS. To further investigate whether intrinsic neural property was responsible for PLS, unit's responses to sounds having different durations were investigated. The result indicates that both intrinsic and extrinsic mechanisms are likely involved in creation of PLS. [Work supported by NIH R01DC00663.]

9:25

2aAB5. Temporal estimation by a model big brown bat. Janine M. Wotton, Michael J. Ferragamo, Timothy M. Sonbuchner (Biol. Dept., Gustavus Adolphus College), and Mark I. Sanderson (Brown Univ.)

The big brown bat, Eptesicus fuscus, uses echolocation to locate prey and displays extraordinary acuity in the perception of temporal cues in acoustic signals. Behaviorally the bat can detect changes at submicrosecond levels but individual neurons in the inferior colliculus (IC) and cortex operate with much less precision. Most of these cells are poor temporal markers with response variation on the order of a few milliseconds and some in tens of milliseconds. A temporal estimator was created incorporating the response properties of recorded neurons and behaviorally appropriate limitations on the number of echolocation emissions. The response of the neurons can be characterized as probability density functions in time and frequency. The characteristics of these neurons were used to create large simulated populations of IC and cortical neurons that show the full range of recorded variation. The connections between these two populations were simulated using a self-organizing neural network. If more than one IC neuron is required to trigger the response of each cortical neuron then the model operates with resolution of microseconds. Manipulating the firing threshold of cortical cells and the relative population sizes influences the errors in target estimation.

9:45

2aAB6. Binaural hearing in dolphins. Patrick W. B. Moore and Randall L. Brill (Marine Mammal Program, SPAWAR Systems Ctr., 49620 Beluga Rd., San Diego, CA 92152-6505)

Binaural hearing is an advantage of having two ears. Human benefits are evident in a 3-dB threshold difference, the ability to localize sound sources in space, and the ability to isolate primary sounds from corresponding echoes. The binaural capabilities of dolphins are relatively unexplored. Studies show that their localization of pure tones underwater is mediated by the same mechanisms observed in terrestrial mammals. Behavioral evidence from free-field localization studies supports reliance on time and intensity cues. Two studies have examined binaural hearing in dolphins using contact hydrophones to isolate the hearing mechanisms. They provided masking level differences (MLDs) comparable to humans and interaural time and intensity difference thresholds that were better than any recorded for terrestrial mammals. Neurophsyiological studies using evoked potentials investigated interaural sensitivity and intensity differences as a function of multiple frequency stimuli presented at various angles around the head of a dolphin. Recent

behavioral studies have mapped the acoustic sensitivity around the head and lower jaw. Those results suggest greater sensitivity forward of the panbone, thought to be the site of best reception, and an asymmetry in sensitivity that may be analogous to that observed in other animals such as the barn owl.

10:05–10:15 Break

10:15

2aAB7. Biomimetic chirplet transforms for environmentally adaptive sonar: A model of singing humpback whales. Eduardo MercadoIII (Ctr. for Molecular and Behavioral Neurosci., Rutgers Univ.–Newark, 197 University Ave., Newark, NJ 07102, mercado@pavlov.rutgers.edu) and L. Neil Frazer (SOEST, Univ. of Hawaii, Honolulu, HI 96822)

Animals that use echoes to identify and localize objects must be able to do so within a variety of acoustically complex environments. The neural processes underlying this ability can be characterized as an adaptive transformation from a one-dimensional space into an *n*-dimensional auditory parameter space. Neurophysiological data suggest that in mammals this transformation is highly overcomplete as well as species and experience dependent; such representations are modeled well by chirplet transforms. Chirplet-based representations of sonar signals and their echoes can be used to generate behaviorally relevant maps of acoustic features. These feature maps, which are similar to representational maps in the auditory cortex, can facilitate separation of echoes from predictable noise sources. A model of sound production and reception in humpback whales is presented to illustrate how such a feature map-based sonar might function. Singing humpbacks space themselves apart from and often approach distant, nonsinging whales, demonstrating that they are able to recognize and localize complex acoustic sources at long ranges in shallow water. Singers also modulate their sound sequences to match others they have heard, demonstrating that they are able to accurately and adaptively represent complex environmental sounds. These abilities suggest that humpback whales are well adapted for both passive and active sonar processing.

Contributed Papers

10:35

2aAB8. An investigation of the pulses produced by the least shrews (*Cryptotis parva*). Mersedeh S. Jalili and Jeanette A. Thomas (Lab. of Sensory Biol., Western Illinois Univ., Quad Cities, Moline, IL, mersedehj@aol.com)

Most echolocation studies have focused on bats and dolphins. Because of technological improvements in ultrasonic sensing and recording equipment, there now are cost-effective approaches towards examining ''other'' groups of mammals for possible echolocation abilities. In this study, we suggest that echolocation is a primitive characteristic, first appearing in insectivores, the ancestor of all other placental mammals. A few other studies and anatomical, behavioral, and physiological attributes suggest that shrews are likely to echolocate. In captivity, least shrews *(Cryptotis parva)* produce series of pulses. We used a board divided into an inner and outer box of equal area to run 8 least shrews through a set of foraging and orientation experiments. Experiments were in the dark and we monitored the circumstances under which shrews produced pulses using a night-vision video camera. An ANABATII bat detector monitored the presence of pulses, a Marantz cassette recorder documented the sounds, and Audioscope software on a laptop computer provided real time sonogram and oscillogram displays. The number of pulses and trains, the typical waveform, power spectrum, peak frequency, and bandwidth (-3 dB) were examined among the experiments. Results provided strong evidence that least shrews use simple pulses for both orientation and foraging.

10:50

2aAB9. The influence of flight speed on the ranging performance of bats using frequency modulated echolocation pulses. Arjan M. Boonman (Dept. of Animal Physiol., Univ. of Tuebingen, Auf der Morgenstelle 28, 72076 Tuebingen, Germany, arjan.boonman@uni-tuebingen.de), Gareth Jones, and Stuart Parsons (School of Biological Sci., Univ. of Bristol, Bristol BS8 1UG, UK)

Many species of bat use ultrasonic frequency-modulated pulses to measure distance to objects. The flight speed of the bat will result in a compression of the echo, and in compression of the time of flight of the signal, both leading to distortion of the perceived range. By comparing Doppler errors incurred with the SCAT-filterbank model and cross-correlation, it was established that the effects of the echo compression are virtually independent of the two receiver models. A cross-correlation model was used to investigate Doppler tolerance of the ranging estimation and range-Doppler coupling separately. Range-Doppler coupling is considered a potential error for bats since a reference time (*T*ref) does not seem biologically feasible. Without using a reference time, range-Doppler coupling due to echo compression was found to vary between 10 and 30 ms in six bat species. With additional range-Doppler coupling due to the compression of the time of flight (34 ms at 2 m from the target), this leads to perceived target displacements between 4.5 and 6.6 cm at a flight speed of 6 m/s. Range-Doppler coupling can be reduced by linearizing the curvature and by increasing the ratio highest/lowest frequency of the pulse. The study further revealed how pulse duration, bandwidth and harmonics influence Doppler tolerance and range-Doppler coupling.

11:05

2aAB10. Experimental simulation of binaural object classification by dolphins. Gerard J. Quentin (Laboratoire d'Imagerie Parametrique, UMR 7623, 15 rue de l'Ecole de Medecine, 75006 Paris, France) and Whitlow W. L. Au (Hawaii Inst. of Marine Biol., Kailua, HI 96734)

We used the experimental setup designed to simulate ultrasound transmission and binaural reception by dolphins at the Marine Mammals Research Program. Three composite piezoelectric transducers, located in the same plane, were placed at the corners of a triangle such that a distance of 14 cm separates the two receivers. This approximates the presumed spacing between the dolphin sound entrances. The two receiving channels approximate those of dolphins (e.g., similar spectral responses, different receiving patterns between the two receptors). Transmitted signals were similar to dolphin clicks and echoes were recorded simultaneously at the two receivers and digitized at 1 MHz using two data acquisition boards housed in a portable computer. We sought to identify different targets in these preliminary experiments: a 3-in. stainless steel spherical shell filled with water and two copper cylindrical shells with outer diameters equal to

1.63 and 1 in., respectively. The last two samples have the same thickness ($\frac{1}{16}$ in.) and are filled either with water or with air. The scattered signals are processed via a simple Fourier transform and analyzed using the resonance scattering theory. A preliminary study of the results is presented.

11:20

2aAB11. Matched field processing of existing binaural dolphinlike signals for MCM. A. Tolstoy (ATolstoy Sci., 8610 Battailles Court, Annandale, VA 22003, atolstoy@ieee.org) and W. Au (Hawaii Inst. of Marine Biol., Kailua, HI 96734, wau@hawaii.edu)

A new matched field processing (MFP) approach has been applied to existing (but limited) dolphinlike binaural data. The frequencies range from 25 to 200 kHz with the received quasi-monostatic echo lasting less than 4.0 ms. The data discussed consist of backscatter echoes from a mud bottom, a drum buried in the bottom, and a Manta-like object also buried in the bottom. The MFP approach (using NO modeling of any kind) was designed to extract differences between apparently similar signals. Since the available data show overly strong and different returns for all the targets, we first manipulate the data to bring them more into line with expectations. The new data were then processed to yield a target ''template'' of scattered returns as a function of time and frequency characterizing the target returns. These templates can easily be used to detect and identify scattering from targets on or in the bottom in low S/N situations. [Work supported by ONR.]

11:35

2aAB12. Consistency of echolocation signals of CF-FM bat measured by a telemetry system. Hiroshi Riquimaroux, Kenji Yamasumi, Yoshiaki Watanabe (Fac. of Eng., Doshisha Univ., Kyotanabe, Kyoto 610-0321, Japan), and Liang-Kong Lin (Tunghai Univ., Taichung 407, Taiwan)

The purpose of the present study was to examine the daily consistency of the constant frequency component (CF) of CF-FM bats and to test the significance of a telemetry microphone system (telemike) mounted on the bat's head. We examined sonar sound characteristics of leaf-nosed bats (Hipposiderous terasensis) when bats were at rest and during flight. Experiments were done within a steel-walled chamber. Their emitted pulses were simultaneously recorded by a fixed microphone at the wall and the telemike. The second harmonic of the CF component (CF2) at rest and during flight varied significantly over several months. When at rest, the fundamental frequency of the bat was about 35 kHz. The second harmonic was the most intense. There was little difference between data taken from two different microphones when at rest. However, during flight we found significant differences between data obtained by the microphone at the wall and the telemike. We could see a systematic change in CF2 frequency during flight through the telemike. Our findings may suggest that the bats compensated for the frequency of the CF component of their pulses by detecting Doppler-shifted CF of returning echoes from the target. We will discuss the individuality and consistency of the CF component during flight and at rest. [The present research was supported by Special Research Grants for the Development of Characteristic Education from the Promotion and Mutual Aid Corporation for Private Schools Japan, a grant to RCAST at Doshisha University from the Ministry of Education, Japan and Doshisha University's Research Promotion Fund.]

TUESDAY MORNING, 5 JUNE 2001 MONROE ROOM, 7:30 TO 11:55 A.M.

Session 2aAOa

Acoustical Oceanography and Underwater Acoustics: Benchmarking Range Dependent Numerical Models

Kevin B. Smith, Cochair
Department of Physics, Naval Postgraduate School, Code PH/SK, Monterey, California 93943

Alexandra I. Tolstoy, Cochair
ATolstoy Sciences, 8610 Battailles Court, Annandale, Virginia 22003

Chair's Introduction—7:30

Invited Papers

7:35

2aAOa1. Two approximate solutions for the 3-D field in the ASA benchmark wedge. Grant B. Deane and Michael J. Buckingham (Scripps Inst. of Oceanogr., La Jolla, CA 92093-0238, grant@mpl.ucsd.edu)

One of the current challenges to the underwater acoustics 3-D numerical modeling community is the development of benchmark geometries and solutions. Benchmark solutions to canonical problems provide a reference against which numerical models can be compared, as well as providing insight into regions or frequency regimes where the models fail. The 3-D penetrable wedge is a good candidate for a benchmark geometry because it corresponds to real-world environments (beaches and continental slopes), is tractable to a variety of numerical modelers, and has been studied analytically. Although there is no exact analytical solution available for the penetrable wedge, several approximate solutions have been developed, which may be regarded as secondary benchmarks. The results from 2 of these models for the 3-D extension of the ASA wedge benchmark will be presented.

7:55

2aAOa2. A two-way integral equation approach to a shelf break problem. David P. Knobles (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029) and Steven A. Stotts (Appl. Res. Labs., Univ. of Texas, Austin, TX 78713)

Benchmark solutions play an important role in the development, testing, and application of numerical models for range-dependent problems in underwater acoustics. To produce benchmark accuracy propagation solutions requires careful attention to both the theoretical formalism and numerical implementation. The focus of this research is a ''benchmark'' solution for a two-dimensional (2-D) shelf-break problem in ocean acoustics. The method employed to construct a solution is a two-way energy-conserving coupled mode approach in the form of coupled inhomogeneous integral equations. Solutions of the coupled equations are obtained using a powerful approach originally introduced in nuclear theory and also used to solve simple nonseparable problems in underwater acoustics. The basic integral equations are slightly modified to permit a Lanczos expansion to form a solution. The solution of the original set of integral equations is then easily recovered from the solution of the modified equations. Benchmark quality demands that the errors in each step of the numerical calculation be well quantified, so that the resolution in the transmission loss can be determined. Consistency checks are made in the form of using different numerical schemes to compute the adiabatic or free propagator and mode coupling matrix operator.

8:15

2aAOa3. PE solutions to some internal-wave benchmark problems. Gordon R. Ebbeson, David J. Thomson (Defence Res. Establishment Atlantic, P.O. Box 1012, Dartmouth, NS B2Y 3Z7, Canada, ebbeson@drea.dnd.ca), and Gary H. Brooke (A & T/Anteon, Victoria, BC V8V 3K3, Canada)

Benchmarking underwater acoustic propagation models is a necessary stage in the evolution of numerical modeling codes. Candidate benchmark problems should challenge the capabilities of existing models in order to promote the development of improved techniques for reliably simulating sound propagation in realistic oceans. In this paper, a parabolic equation (PE) model is applied to the suite of internal-wave test cases that are offered for numerical consideration by the organizers of the Benchmarking Range Dependent Numerical Models session. The PE calculations are carried out using a code that was originally developed for matched-field processing applications [G. H. Brooke *et al.*, ''PECan: A Canadian parabolic equation model for underwater sound propagation,'' J. Comput. Acoust. (2001)]. The capability of PECan to propagate sound accurately through the benchmark environments is examined for both tonal and broadband signals. The discussion will address issues related to reciprocity, interpolation of the internal-wave sound speed structure, and convergence of solutions as a function of Padé order and range and depth step sizes. Where possible, PECan will be validated against other well-known propagation models to assess its accuracy as well as to appraise the suitability of the proposed internal-wave benchmark problems to test the limits of existing propagation codes.

8:35

2aAOa4. Verified acoustic propagation predictions for a shelf-break case. Richard B. Evans (Sci. Appl. Int'l. Corp., 21 Montauk Ave., New London, CT 06320) and David B. King (Naval Res. Lab., Stennis Space Center, MS 39529-5004)

A comparison is made between transmission loss computed with stepwise couple modes and the parabolic equation for several shelf-break cases. The close agreement between these two diverse approximate solutions of the wave equation can be interpreted in one of two ways: Either both are producing the correct result or they are making identical errors. The low probability of the second coincidence is interpreted as verification of both results. The difficulties encountered in obtaining verified propagation results are discussed.

Contributed Papers

8:55

2aAOa5. Inaccurate results from benchmark accurate ocean acoustic propagation models: It's the implementation, not the model. Stanley A. Chin-Bing, David B. King, and Robert A. Zingarelli (Naval Res. Lab., Stennis Space Center, MS 39529-5004, chinbing@nrlssc.navy.mil)

Despite numerous publications that document the benchmark accuracy of several ocean acoustic propagation models, some users of these models continue to get inaccurate results. The problem is not with the model, but with the implementation of the model. The most common mistake in implementing range-dependent acoustic models (e.g., parabolic equation models) is the selection of model parameters (e.g., step size, grid spacing, attenuating bottom depth). Another common mistake is lack of sufficient environmental data to the model. The number of implementation errors could be significantly reduced, if the user would abandon the erroneous belief that the model is a ''black box'' that will always produce the correct answer in a single run. Knowledgeable users methodically vary the input parameters and make repeated runs until model convergence is evident. It is possible to preselect model parameters, and couple the acoustic model with its environmental databases, so that a general user can obtain an accurate result. The U.S. Navy has done this with its Navy Standard Propagation models for nearly a decade. The models and their revisions are documented, configuration managed, and verified by applications to over 500 test cases. Selected examples will be presented. [Work supported by ONR/NRL.]

9:10

2aAOa6. Complex ray solutions to the 3-D wedge ASA benchmark problems. Evan K. Westwood (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713)

In support of the ASA's effort to benchmark shallow-water, range-dependent acoustic propagation models, a complex ray approach is applied to the three-dimensional wedge problems. The ray model [Westwood, J. Acoust. Soc. Am. **92**, 2212–2222 (1992)] is designed specifically for the case of a point source in a 3-D wedge, where the receiver may lie cross slope, as well as downslope or upslope, from the source, and where the wedge of water and the bottom are isovelocity. The total field in the 3-D wedge is expressed as a sum of ray fields, each of which corresponds to a saddle point of a double integral over plane waves. The broadband case of the benchmarking problem is solved by finding frequency-dependent eigenrays at coarsely spaced frequencies and interpolating the eigenray characteristics. Eigenray diagrams provide an intuitive picture of the horizontal refraction effects associated with three-dimensional underwater acoustic propagation. [Work supported by ONR.]

9:25

2aAOa7. Geoacoustic inversion in range-dependent environments: A plane-wave reflection coefficient approach. Steven A. Stotts (Appl. Res. Labs., P.O. Box 8029, Univ. of Texas, Austin, TX 78713-8029) and David P. Knobles (Univ. of Texas, Austin, TX 78713)

One of the most challenging aspects of geoacoustic inversion in ocean waveguides involves range-dependent problems, such as in shallow water or continental shelf environments. Inversion calculations may require tens or hundreds of thousands of calls to an acoustic propagation model. When the forward propagation model is range dependent, the computation time can become extremely large. If, however; the range dependence is known, an approximation is introduced that significantly simplifies the problem. Using the measured bathymetry and range-dependent sound-speed profile, the geometrical optics approximation is used to compute the eigenrays between sources and receivers. Assuming a rigid bottom boundary condition, the MEDUSA algorithm is used for this purpose. This range-dependent ray theory calculation separates the propagation in the water column from that in the seabed. The modeled broadband pressure fields are obtained by computing the plane-wave reflection coefficient (PWRC) at the angles and frequencies as needed in the eigenray expansion. Namely, each perturbation of the seabed does not require a new range-dependent calculation of eigenrays, only a recalculation of the PWRC. Examples are shown that demonstrate the accuracy, efficiency, and limitations of the approach. [Research supported by SPAWAR and ONR.]

9:40

2aAOa8. Theory of acoustic radiation over a 3-D hyperbolic ocean ridge. Thomas K. Berger (Scripps Inst. of Oceanogr., Univ. of California, San Diego, 9500 Gilman Dr., La Jolla, CA 92093-0238)

An analytic solution to a 3-D problem of ocean acoustics involving horizontal refraction is presented. An expression for the acoustic field due to a harmonic point source in an isovelocity ocean overlying a hyperbolic ridge with perfectly reflecting boundaries is given as a finite modal sum of integrals. The integrals are in terms of Mathieu functions, the eigenfunctions of the reduced wave equation in the elliptical coordinate system. These eigenfunctions are approximated away from the low-frequency limit by standard WKB techniques and the resulting integrals are then estimated by first- and second-order stationary-phase asymptotics. This gives an approximation in terms of elliptical functions for which standard numerical routines are available. The approximation is very accurate in the shallow depth and gentle slope limits appropriate to the physical situation of ridges in an isovelocity ocean. The solution includes the locations of the caustics and shadow zones and predicts a complicated intramodal interference pattern resulting from the intersection of up to three rays in a given mode. The intermodal interference from the finite modal sum is also evident in the full solution. The explicit representation of the features arising from horizontal refraction makes this model useful as a 3-D benchmark solution.

9:55–10:10 Break

10:10

2aAOa9. Application of the enhanced ray theory to benchmark problems. Nick E. Maltsev (MVM Intl., 10678D Maplewood Rd., Cupertino, CA 95014)

Enhanced ray theory [Maltsev, J. Comp. Acoust. (in press)] was applied to benchmark problems. A new approach for searching of eigenrays, based on the structure of rays surface in phase space, was developed. The sound field is composed by numerous eigenrays, so the time and space structure of the field is preserved and available for interpretation.

10:25

2aAOa10. Broadband PE/SSF modeling in 2-D and 3-D. F. D. Tappert (Appl. Marine Phys., Rosenstiel School, Univ. of Miami, Miami, FL 33149)

Prediction of pulse propagation for the 2-D and 3-D test cases proposed for benchmarking range-dependent numerical models is accomplished with the c_0-insensitive version of the broadband parabolic equation with the split-step Fourier (PE/SSF) algorithm numerical model that takes advantage of efficient Fourier syntheses from the frequency domain to the time domain. Special attention is paid to the fast and accurate prediction of the travel times and amplitudes of the resolved multipath arrivals, as well as to the unresolved multipaths. The role of chaos in limiting accurate pointwise predictions in range-dependent (nonseparable) propagation problems is highlighted. [Work supported by ONR Code 321OA.]

10:40

2aAOa11. Some normal mode and PE propagation predictions for selected benchmark test cases. A. Tolstoy (ATolstoy Sci., 8610 Battailles Court, Annandale, VA 22003, atolstoy@ieee.org)

Results will be presented for the 3-D and shelf-break test cases. The 3-D problem results here will be for 2-D predictions only. The KRAKEN (adiabatic and coupled versions) and PE (energy conserving and nonconserving) models will be the primary codes investigated. It is found that it is *nontrivial* to run these codes even when one is familiar with them. It is also found that all the models generally agree, but there are some quantitative differences. [Work supported by ONR.]

10:55

2aAOa12. An implementation of a three-dimensional wave propagation model in a parallel processing computational environment. Thomas N. Lawrence (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713), Richard D. Pound, and Roy M. Jenevein (Dept. of Computer Sci., Univ. of Texas, Austin, TX 78712-1188)

A parallel processor implementation of a three-dimensional parabolic equation approach to acoustic wave propagation was developed using the University of Texas Cray T3E (Lone Star). Lone Star is currently an 84 processor machine. We will discuss results from modeling various underwater acoustic environments where three-dimensional results are required, and the computation speed-up achieved using a multiprocessor approach. High performance computing, i.e., computation on high speed multiple processor machines, is now entering the lower priced small computer era. Environmental acoustic problems heretofore ignored because of the computational intensity required can now be considered for operational field systems. Although the T3E is a mainframe machine, this model is being developed for a low cost clustered processor machine with a combination of shared and distributed memory, which we expect to have in operation this year. [Work sponsored under Independent Research and Development at the Applied Research Laboratories. Computer support on the T3E is funded by the National Partnership for Advanced Computational Infrastructure through the University of Texas Advanced Computing Center for Engineering and Science.]

11:10

2aAOa13. Numerical simulations with 3D WAPE considering shallow water range-dependent environments. Frédéric Sturm (INSA de Lyon, LMC-UMR CNRS 5514 et MAPLY–UMR CNRS 5585, 20 avenue Albert Einstein, 69621 Villeurbanne cedex, France)

A three-dimensional parabolic equation based model, 3-D WAPE, is run considering several shallow water range-dependent environments. This 3-D model is based on a higher-order Padé approximation in depth and a linear approximation in azimuth. Broadband modeling is handled using a Fourier synthesis approach. We focus on the numerical method chosen to solve the inherent initial- and boundary-value problem. When

one is concerned with solving sound propagation problems in any three-dimensional waveguides, one of the main issues is to reduce the azimuthal sampling so as to achieve 3-D computations in a more reasonable time with the same accuracy. The efficiency of using higher-order finite difference schemes in azimuth is discussed. Some numerical results obtained using 3-D WAPE on various test case scenarios are presented, including the ASA 3-D penetrable wedge-shaped waveguide benchmark. For each test case, CPU time comparisons are performed. Another interesting aspect of the marching algorithm used is its ability to be parallelized. Parallel computing appears to be very promising for solving large scale numerical problems. The advantage of using parallel computing is also discussed.

11:25

2aAOa14. 3-D benchmarking with a three-dimensional azimuthal wide-angle model in the wedge. Ying-Tsong Lin, Chifang Chen (Dept. of Naval Architecture and Ocean Eng., Natl. Taiwan Univ., 73 Chou-Shan Rd., Taipei, Taiwan, ROC), and Ding Lee (Yale Univ./Naval Undersea Warfare Ctr., CT)

In predicting wave propagation in either direction, the size of the angle of propagation plays an important role; thus, the concept of wide-angle is introduced. Most existing acoustic propagation prediction models do have the capability of treating the wide-angle but the treatment, in practice, is vertical. This is desirable for solving 2-dimensional (range and depth) problems. In extending the 2-dimensional treatment to 3 dimensions, even the wide-angle capability is maintained in most 3-D models, it is still vertical. This paper uses an azimuthal wide-angle wave equation. The 3-D benchmarking is done by a 3-dimensional azimuthal wide-angle model with real 3-D side-wall boundaries for up-slope calculations. The cross-slope calculations are also done for both subcases. [Work supported by National Science Council of Republic of China.]

11:40

2aAOa15. A research of sound rays propagated in sloping seabed ocean. Guanting Jiang and Shengming Guo (Zhongguancun Rd. 17, Beijing, PROC, jgt@oceana.ioa.ac.cn)

The property of sound waves transmitted through sea water always presents an important problem in acoustics. In this paper, sound rays propagated underwater on an irregular seabed were calculated and some interesting results were received. Sound transmitted amplitude is lower than in a plane seabed sea, and in a slope sea bottom the direction of the sound ray will turn deflexion along the bed. All of the results are obtained using the ray-trace method.

TUESDAY MORNING, 5 JUNE 2001 ADAMS ROOM, 8:30 A.M. TO 12:00 NOON

Session 2aAOb

Acoustical Oceanography: Acoustical Instrumentation for Water Column Measurements III

Kenneth G. Foote, Chair
Department of Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

Chair's Introduction—8:30

Invited Paper

8:35

2aAOb1. Fisheries applications of Doppler sonar. Len Zedel (Phys. and Physical Oceanogr. Dept., Memorial Univ. of New Foundland, St. John's, NF A1B 3X7, Canada, zedel@physics.mun.ca)

Signal processing inherent to broadband Doppler current profiling systems has great potential for applications in fisheries monitoring. The large bandwidth (25%–50% of the carrier frequency) contains information that can be used in target characterization, the system determines current profiles and there is the potential to make direct measurements of fish swimming speed. The capabilities of these systems are explored: field data collected with an RD Instruments 307-kHz Workhorse Acoustic Doppler Current Profiler (ADCP) demonstrate observations of large herring schools, performance is also evaluated using observations made in a tow-tank facility. For large concentrations of fish, speed and direction can be determined but for lower concentrations of fish new data processing techniques are required. For the system configuration as tested, single ping estimates have a standard deviation as low as 10 $\mathrm{cm\,s^{-1}}$. This accuracy is substantially better than the accuracy of $\simeq 40$ $\mathrm{cm\,s^{-1}}$ expected and suggests that for isolated targets the approach is equivalent to fully coherent Doppler processing. There is a difficulty discriminating discrete targets because of the long pulse code sequences transmitted. [This research was funded by the Natural Sciences and Engineering Council of Canada, and by Fisheries and Oceans Canada.]

8:55

2aAOb2. Inference of size and orientation of swim-bladdered fish from high-frequency acoustic echograms. Dezhang Chu, Andone Lavery (Woods Hole Oceanogr. Inst., 98 Water St., Woods Hole, MA 02543), and J. Michael Jech (NOAA/NMFS Northeast Fisheries Sci. Ctr., Woods Hole, MA 02543)

It is well known that the volume scattering strength is very sensitive to the size and orientation of scattering objects such as swim-bladdered fish. Such behavioral information is in general not always available. Inadequate knowledge of target orientations limits our ability to correctly estimate the size and abundance of fish acoustically. To extract orientation information of individual fish and/or fish schools, a scattering-model-based image processing method is presented. For fish aggregations within which individual fish can be resolved from the echogram, the geometric shapes and acoustic intensities of the identifiable highlights can potentially be used to infer the sizes and orientations of the individual targets. For fish schools within which individual fish cannot be resolved, the mean fish size and orientation can still potentially be inferred by examining the statistics of the amplitude distribution of the patches resulting from strong backscattering from fish schools. The scattering model is based on the Kirchhoff approximation. The requirements to improve the performance of the presented method and its limitations are discussed.

9:10

2aAOb3. Calibration and use of a broadband ADCP to measure zooplankton volume scattering strengths. Mark C. Benfield (Dept. of Oceanogr. and Coastal Sci./Coastal Fisheries Inst., Louisiana State Univ., 218 Wetland Resources, Baton Rouge, LA 70803, mbenfie@lsu.edu), Duncan E. McGehee (Ocean Sci. Res. Group, BAE Systems, San Diego, CA 92123-4333), Sean F. Keenan (Louisiana State Univ., Baton Rouge, LA 70803), and Kent L. Deines (DrDoppler.com, Poway, CA 92064)

A 1200 kHz broadband ADCP from RD Instruments was calibrated and used to collect quantitative volume backscattering data from zooplankton in the Gulf of Mexico. Voltages on the four receivers, reported in terms of ''counts,'' were linear with respect to received sound pressure level in dB *re* 1 μPa, with sensitivities of 2.376 to 2.434 counts/dB, depending on the receiver. Source levels ranged from 212.9 to 214.6 dB *re* 1 μPa at 1 m. The ADCP was mounted on the C/V Admiral Semmes and used to survey volume backscattering around an oil platform in the Gulf of Mexico. Volume backscattering strengths measured between -40 and -90 dB, which were generally consistent with independent observations of the numbers and sizes of scatterers present in the water column. These results suggest that, with careful calibration, ADCPs may be used for quantitative zooplankton research.

9:25

2aAOb4. New system for remotely monitoring the three-dimensional movement of acoustically tagged fish. Tracey W. Steig, Samuel V. Johnston, and Bruce H. Ransom (Hydroacoustic Technol., Inc., 715 NE Northlake Way, Seattle, WA 98105, support@htisonar.com)

A passive acoustic tag system was developed to monitor the three-dimensional movements of migrating fish with submeter resolution. The acoustic tag receiver monitors an array consisting of up to 16 omnidirectional hydrophones, with received signals synchronized to determine the arrival times for each pulse transmitted by the acoustic tag. Arrival times are then used to calculate the three-dimensional position of a tagged fish as it moves through the array. Algorithms were developed to precisely calculate the three-dimensional positions of the hydrophones, and of each acoustic tag. Over the last 4 years, this system was used at several dams in the United States. Most studies to date monitored downstream migrating juvenile salmonids as they approached and passed turbine intakes, spillways, and juvenile bypass systems at hydroelectric dams. Fish movement patterns were tracked in three dimensions over time, typically with submeter resolution. Tagged fish were 160–240 mm long. Acoustic tags were approximately 7 mm in diameter by 23 mm long, weighted 2 g, and transmitted at 307 kHz. Tag codes (up to 500), pulse width (typically 1–5 ms), and ping rate (typically 0.3–3 pings/s) were field programmable. Current tags incorporate signal encoding for an improved signal-to-noise ratio, and weigh as little as 1 g.

9:40

2aAOb5. Acoustically characterizing detritus with a 420-kHz echosounder. Diane E. Di Massa[a] (Univ. of Massachusetts–Dartmouth, II-116A, 285 Old Westport Rd., North Dartmouth, MA 02747, ddimassa@umassd.edu)

This paper presents the new application of hydroacoustically measuring and characterizing detritus (here defined as suspended decomposing plant material) for shallow-water environments such as bays, estuaries, or rivers. Detrital material was measured hydroacoustically using a 420-kHz digital echosounder, first in a controlled laboratory test and then in an open estuary. Controlled tests on both small matted masses and individual grass strands produced signals clearly visible (6–15 dB) above the background noise. Field tests were performed using ship-mounted downward looking sonar and echo integration software to determine the feasibility of using hydroacoustics to measure, characterize, and potentially quantify the detrital biomass. A vertical series of plankton nets, slightly downstream of the echosounder, and flow velocity measurements by a 1200-kHz acoustic doppler current profiler (ADCP) slightly upstream provided ground-truth information of local concentrations. Hydroacoustics proved successful in identifying variable material concentrations both spatially and temporally, with typical backscatter intensities of -70 db. Echo amplitude information from the ADCP showed a first-order correlation with backscatter detected by the echosounder and with net sample volumes. [a] Also affiliated with Woods Hole Group, 81 Technology Park Drive, East Falmouth, MA 02536.

9:55

2aAOb6. The new Simrad EK60 scientific echo sounder system. Lars Nonboe Andersen (Simrad AS, P.O. Box 111, N-3191 Horten, Norway)

A widely used tool for fish abundance estimation and other marine research is the scientific echo sounder. Acoustic information obtained from scientific echo sounders such as target strength and target position can be valuable for describing individual targets whereas measurements of volume backscattering strength are suitable for describing a large number of targets distributed over a volume of water. The new Simrad EK60 scientific echo sounder system provides high-quality target strength, target position, and volume backscattering strength measurements over a large dynamic range for a wide range of frequencies (18–200 kHz). An overview of the EK60 will be presented and operation of the echo sounder will be demonstrated. The presentation will include demonstration of the calibration procedure of the EK60.

10:10–10:20 Break

10:20

2aAOb7. A quantitative echo sounder with multiple beams, model KFC-3000. Kenji Kuratsu, Takahiro Satho (Kaijo Corp., 3-1-5, Sakae-cho, Hamura-shi, Tokyo 205-8607, Japan), Yoichi Miyanohana, Kouichi Sawada, and Yoshimi Takao (Natl. Res. Inst. of Fisheries Eng., Ebidai, Hasaki, Kashima-gun, Ibaraki 314-0421, Japan)

We have developed a quantitative echo sounder with various features such as effective combination of dual beam and split beam methods. This system is designed in accordance with the established designing method of a quantitative echo sounder. The method specifies source power, operating frequency, and so on for a given measuring object and range under the principle of minimum estimation error. Ideal omnidirectional and symmetrical simultaneous split beam methods detect the three-dimensional position of an individual target. An omnibus display on the multiwindows

provides full echo information. Raw data for post-processing is used for accurate biomass evaluation. The easy operation of this system offers a wider range of users more adequate and reliable data for fishery resource surveys.

10:35

2aAOb8. Hydroacoustic performance of fishery survey instruments. Hans P. Knudsen, Egil Ona, and Ingvald Svellingen (Inst. of Marine Res., P.O. Box 1870, Nordnes, N-5817 Bergen, Norway)

Accurate estimation of fish stock abundance by acoustic surveying methods is now essential for long-term assessment of many of the world's largest fish stocks and the status of their ecosystems. Significant improvements of the acoustic oceanic surveying methodology have been achieved over the last decade, in particular in calibration stability, receiver design, post-processing systems and transducer installations for oceanic, bad-weather operation. Experiences with introducing these new elements on four ocean-going research vessels, with examples, are presented and discussed. Since many of the potential survey errors related to the technical instrumentation seem to be overcome or minimized, the present focus for improvements must now be on biotic factors such as target strength variability, vessel avoidance and vertical distribution of the fish relative to the acoustic sampling volumes. The effect of significantly reducing vessel-emitted noise and improved observation methods in the echo sounder dead zones are discussed.

10:50

2aAOb9. Detection range of acoustic instruments for fisheries. Masahiko Furusawa, Jusam Park, Myounghee Kang, and Chumming Fan (Tokyo Univ. of Fisheries, Tokyo, Japan)

Detection ranges of acoustic instruments mainly used for fisheries and their research are derived as the range bordered by a certain signal-to-noise ratio (SNR) threshold. The SNR is depicted by several factors on transmitting and receiving, sound propagation, scattering by objects, and mainly self-ship noise. The detection ranges are shown for several fisheries instruments: fisheries echo sounder, quantitative echo sounder, and bio-telemetry system. The results can be used for designing the instruments, examining the capability of user's own instruments, and interpreting obtained data or echograms. Some examples of the results follow. Increasing transmitting power is not as effective for high frequencies as for low frequencies to increase the detection range. Comparison of volume backscattering strengths obtained by the quantitative echo sounder at several frequencies for the purpose of rough species' identification should be done within the same detection range. By applying the concept of the detection range for the bio-telemetry receiver beams, the number of the beams and the beamwidths can be determined.

11:05

2aAOb10. Experimental tests of a new method to monitor sea medium by analyzing ambient noise and reradiating it from a distant point. V. Furduev Alexander (Head of Lab. Andreyev Acoust. Inst., 4 Shvernik Str., 117036 Moscow, Russia) and D. Svet Victor (Head of Lab. Andreyev Acoust. Inst., 117036 Moscow, Russia)

In-sea experiments show a possibility to estimate parameters of the water column by analyzing space-time characteristics of the natural ambient noise. If the noise is reradiated by sound transponder located at a distance from the receiver, a regenerative monitoring scheme can be implemented. In such scheme a feedback is present that increases the contrast of the spectrum maximums to higher accuracy. By increasing the amplification gain of the whole circuit (including the feedback loop of the underwater channel), one comes to a self-sustained oscillator, its frequency deviations are indicating variations of the medium parameters. Experimental examples are presented to confirm feasibility of the proposed monitoring technique for measuring temperature variations, current velocity, periods of internal waves, and other features of the water column. The method is especially applicable to small sea areas like straits, harbors, and lakes to monitor their environmental equilibrium or manifestations of human activity. [Work supported by RFBR, Project No. 00-05-64226.]

11:20–12:00

Panel Discussion

TUESDAY MORNING, 5 JUNE 2001 SALONS III/IV, 8:00 TO 9:15 A.M.

Session 2aBBa

Biomedical Ultrasound/Bioresponse to Vibration: Ultrasound and Vibration in Musculoskeletal Structures

R. Glynn Holt, Chair

Department of Aerospace and Mechanical Engineering, Boston University, 110 Cummington Street, Boston, Massachusetts 02215

Contributed Papers

8:00

2aBBa1. Evaluation of vibratory coherence as an alternative to radiography in assessing bone healing after osteo-distraction. Tarek H. El-Bialy, Thomas J. Royston, and Akira Sakata (Univ. of Illinois at Chicago, Chicago, IL 60607, troyston@uic.edu)

Distraction osteogenesis is used in orthopedics to lengthen bones, by cutting or breaking the bone and gradually separating the two pieces as new bone fills the intervening space. There is a need for early assessment of the degree of bone healing that allows for normal functioning without unwanted side effects. This study compared different techniques used to evaluate the degree of bone healing during mandibular osteodistraction in 21 rabbits. For each rabbit, the mandible was cut in a surgical procedure and then 72 h later distraction began at a rate of 3 mm per day. Bone formation at the distraction site was assessed by *in vivo* photodensitometry on head radiographs, an *in vivo* (nondestructive) vibratory coherence test across the distraction site, a post-mortem, *ex vivo* (destructive) three-point bending mechanical test, and by post-mortem, *ex vivo* (destructive) histological examination. Statistical analyses included analysis of variance (ANOVA) and correlation coefficient tests. The findings revealed that the results of bone photodensity and the vibration technique are positively correlated to the results of the mechanical three-point test and histological examination. The use of the vibration technique may provide a substitute for or augment the routine use of radiography for *in vivo* evaluation and monitoring of bone healing.

8:15

2aBBa2. Physical mechanisms implied in ultrasound propagation through trabecular bones. Frederic Padilla and Pascal Laugier (Laboratoire d'Imagerie Parametrique, P6-CNRS UMR 7623, 15 rue de l'Ecole de Medecine, 75006 Paris, France, padilla@lip.bhdc.jussieu.fr)

Measurements of quantitative ultrasound parameters represent an established mean of skeletal status assessment in osteoporosis. However, to date, the exact physical mechanisms underlying ultrasound attenuation in cancellous bone have not been clearly documented. This study reviews the different mechanisms underlying propagation of ultrasound in trabecular bone, based on results obtained by our group and others. Bone is a porous, anisotropic, and scattering medium. Anisotropy of phase velocity has been explained with a multilayer model. Several attempts have been made to model the propagation with the Biot's theory for poroelastic media. Encouraging results have been obtained for velocity. However, the viscous absorption mechanism implied in this theory disagrees with the experimental attenuation data and with the observed influence of fluid viscosity. Negative dispersion has been recorded, in agreement with some multiple scattering ideas. Several models of scattering (analytical and numerical) have given good predictions of backscattering coefficient. Despite the good results observed with these models, none of these could explain the experimental data, and especially the attenuation data. In conclusion, a more complex model will be discussed, to describe both absorption and scattering effects, as in geophysics for rock studies.

8:30

2aBBa3. Measurements of attenuation and backscattering in trabecular bones over a large frequency range. Frederic Padilla, Sana Chaffai, Genevieve Berger, Pascal Laugier (Laboratoire d'Imagerie Parametrique, P6-CNRS–UMR 7623, 15 rue de l'Ecole de Medecine, 75006 Paris, France, padilla@lip.bhdc.jussieu.fr), and Francoise Peyrin (CREATIS, UMR INSA 502, 69621 Villeurbanne Cedex, France)

In order to understand the propagation of ultrasound through trabecular bones, measurements of attenuation and backscattering coefficients were performed. Attenuation was measured on 14 human calcanei, over a large frequency bandwidth (0.2–1.7 MHz). The experimental attenuation coefficient values were modeled with a nonlinear power fit $\alpha(f)=a_0+a_1 f^n$. The attenuation coefficient was found to increase as $f^{1.09\pm0.3}$. A highly significant relationship was noted between a_1 and BMD. No correlation was found between n and BMD. The backscatter coefficient was measured in 19 bones specimen in the frequency range 0.4–1.2 MHz. The experimental frequency dependence was found to be $f^{3.4\pm0.3}$. A twofold theoretical approach was then adopted. The analytical model of Faran for spherical and cylindrical elastic cylinders showed a qualitative agreement with experimental data. A better agreement was found with a weak scattering medium model, where the backscatter coefficient is related to the autocorrelation function of the propagating medium. These results first open interesting prospects for the investigation of the influence of bone microarchitecture on ultrasonic scattering; second it seems to indicate that scattering is not the main mechanism of ultrasonic attenuation, because of the two different frequency dependencies, as confirmed by a numerical fit (absorption + scattering).

8:45

2aBBa4. Cortical bone characterization by guided waves. Maryline Talmant, Emmanuel Bossy, Estelle Camus, Frederic Padilla (LIP, UMR CNRS 7623, 15 rue de l'Ecole de Medecine, 75006 Paris, France), and Pascal Laugier (LIP, UMR CNRS 7623, 75006 Paris, France, talmant@lip.bhdc.jussieu.fr)

The axial transmission method was first developed in 1958 to study the status of cortical bone during fracture healing. It has been subsequently used to investigate bone strength in osteoporosis. Two transducers are placed on the same side of the bone and the velocity of the first arriving signal on the receiver is measured. The present paper contributes to the understanding of the type of propagation involved in this technique when applied on bone. When using a propagation model for semi-infinite media the first signal is identified as the longitudinal lateral wave for a defined arrangement of transducers. However, the ratio of bone thickness over wavelength is of the order of unity and then normal modes of vibration give a non-negligible contribution to the reflected pressure. Whereas a lateral wave investigates the bone superficially, shell modes penetrate inside the material much more. Signal simulations, *in vitro* experiments and *in vivo* measurements using our prototype will be presented. Particular attention will be given to the influence of cortical thickness. The velocity of the first signal decreases with respect to the thickness, from the velocity of the lateral wave to the velocity of the extensional mode of the lowest order.

9:00

2aBBa5. Ultrasonic pulse transmission technique to evaluate biomechanical properties of ACL graft substitutes. Krishna V. Kambhampati (Dept. of Bioengineering, Aurobindo Pharma Ltd., Plot #2, Maithri Vihar, Ameerpet, Hyderabad, 500 038, India, rad@aurobindo.com)

The study aims at the identification of an ideal graft material for arthroscopy guided anterior cruciate ligament (ACL) reconstruction (ACLR). Ultrasonic pulse transmission technique was used to analyze the elastic properties (EP) of fresh human ACL, semitendinosus tendon (ST), middle one-third of patellar tendon (CPT) and cigarette rolled, distally based iliotibial band (CITB) grafts in an untwisted and incremented number of twists (two to eight times) at 1 MHz fundamental frequency. The results indicated that the strength of the graft depended upon its collagen density and the EP depended upon the pretensioning of the graft for ACLR. The EP were shown to have significant difference on selective pretensioning of the graft. The EP of CITB grafts changed after twisting. Rigidity of twisted CITB graft surpassed that of the ACL, ST, and CPT grafts. The CITB graft twisted five to seven times along its long axis is isoelastic with normal ACL and CITB graft twisted eight times is stronger than even CPT graft. Twisting of CITB graft provides initial strength and stiffness, necessary pretensioning and the balance between elasticity and rigidity is achieved to duplicate the EP of ACL. The success of ACLR depends upon using an isoelastic graft taken from a relatively, less important CITB.

Session 2aBBb

Biomedical Ultrasound/Bioresponse to Vibration: Topics in Medical Acoustics

Paul L. Carson, Chair

Department of Radiology, University of Michigan Medical Center, Kresge III, R3315, 200 Zina Pitcher Place, Ann Arbor, Michigan 48109-0553

Contributed Papers

9:45

2aBBb1. Active ultrasound imaging of breast tumors: Forward and inverse scattering methods. Yanqing Zeng, Zhongqing Zhang, and Qinghuo Liu (Dept. of Elec. and Computer Eng., Duke Univ., Box 90291, Durham, NC 27708-0291, qhliu@ee.duke.edu)

Ultrasonic imaging for the detection of breast tumors is an important technique to complement existing x-ray mammography. The potential advantages of the ultrasound imaging technique stems mainly from the relatively high contrast of acoustic properties between tumors and normal breast tissue. However, this high contrast also increases the difficulty of forming an accurate image because of the increased multiple scattering. To address this issue, we have developed fast forward methods based on a combination of extended Born approximation, conjugate- and biconjugate-gradient methods, and fast Fourier transform; we propose two nonlinear ultrasound imaging algorithms to improve the resolution for the high-contrast media encountered in ultrasound breast tumor detection. The measured data are simulated with a forward method. The inverse scattering algorithms account for the complicated multiple scattering (diffraction) effects. Numerical results show that our algorithms can accurately model and invert the high-contrast media in breast tissue. The outcome of the inversion is a high-resolution digital image containing the physical properties of the tissue and potential tumors. Compared with a conventional x-ray mammography, the ultrasound imaging system with an accurate imaging algorithm has the potential to provide better specificity between benign and malignant tissues.

10:00

2aBBb2. Modeling sound transmission through the pulmonary system and chest with application to diagnosis of a collapsed lung. Xiangling Zhang, Thomas J. Royston (Univ. of Illinois at Chicago, Chicago, IL 60607, troyston@uic.edu), Hussein A. Mansy, and Richard H. Sandler (Rush Medical College, Chicago, IL 60612)

A study was undertaken to demonstrate the feasibility of using audible-frequency vibro-acoustic waves for diagnosis of pneumothorax, a collapsed lung. The hypothesis was that the acoustic response of the chest to external excitation would change with this condition. External acoustic energy was introduced into the trachea via an endotracheal tube. For the control (nonpneumothorax) state, it is hypothesized that sound waves primarily travel through the airways, couple to the lung parenchyma, and then are transmitted directly to the chest wall. In contradistinction, when a pneumothorax is present the intervening air presents an added barrier to efficient acoustic energy transfer. A theoretical model of sound transmission through the pulmonary system and chest region to the chest wall surface is developed to more clearly understand the mechanism of intensity loss when a pneumothorax is present, relative to a baseline case. The model predicts a decrease in acoustic transmission strength of as much as two ordersof magnitude when a pneumothorax is present. This is approximately in agreement with experimental measurements on mongrel cannies. Development of the model and its comparison with experimental canine studies will be reviewed. [Research supported by NIH NCRR Grant No. 14250 and NIH NHLBI Grant No. 61108.]

10:15

2aBBb3. Detection of ultrasound generated contrast bubbles in a refluxing canine model. Emma Y. Hwang, J. Brian Fowlkes, Paul L. Carson (Dept. of Radiol., Univ. of Michigan, Kresge III, R3315, Ann Arbor, MI 48109-0553), and David A. Bloom (Univ. of Michigan, Ann Arbor, MI 48109-0330)

A surgical reflux procedure, performed on two canines, was used to test a less invasive method for diagnosing urinary reflux. After approximately 6 weeks of recovery, examinations tested for reflux over the next 7 weeks. The canines were positioned prone (slightly head down) on a tilted table with a hole exposing the depilated abdomen, where a 1.18 MHz high intensity focused ultrasound transducer could be coupled. The reflux condition was evaluated first using standard fluoroscopy and x-ray contrast instilled in the bladder by a gravity feed through a transurethral catheter. For high grade reflux, dilute contrast agent (Optison) was tested using harmonic color Doppler at 3 MHz on a Toshiba Powervision 8000 scanner imaging the kidney. Finally, aqueous solutions with elevated CO_2 were placed in the bladder and microbubbles generated using a 10 s application of 1.18 MHz ultrasound (25 cycle bursts with 0.5% duty factor and 8.4 MPa). Bubbles were detected in the kidney on two occasions with strong color Doppler signal, near that from dilute Optison. Methods can naturally produce evelated CO_2, and contrast bubbles should be producible transcutaneously for reflux diagnosis without ionizing radiation and catheterization. [Work supported in part by a USPHS grant 2 RO1DK42290.]

10:30

2aBBb4. High frequency backscatter and attenuation measurements of porcine red blood cell suspensions between 30 and 90 MHz. Subha Maruvada (Focused Ultrasound Surgery Group, Dept. of Radiol., Brigham and Women's Hospital, 221 Longwood Ave., LMRC 13, Boston, MA 02115), K. Kirk Shung, and Shyh-Hau Wang (The Penn State Univ., University Park, PA 16802)

There are now diagnostic ultrasonic imaging devices that operate at very high frequencies (VHF) of 20 MHz and beyond for clinical applications in ophthalmology, dermatology, and vascular surgery. To be able to better interpret these images and to further the development of these devices, knowledge of ultrasonic attenuation and scattering of biological tissues, such as blood, in this high-frequency range is crucial. VHF experiments were performed on porcine red blood cell (RBC) suspensions, which have been used in ultrasound experiments by many investigators but in a lower frequency range. Attenuation and backscatter from various hematocrit levels (6%, 10%, 15%, 20%, 25%, 30%) from 30 to 90 MHz were measured using focused transducers. The dependence of the attenuation coefficient from all suspensions followed linearly while the backscatter coefficient for low hematocrit suspensions was found to be a maximum between 10% and 15%. The higher hematocrits showed a decrease in their frequency dependence, possibly no longer indicating Rayleigh scattering since the wavelength is not much larger than the size of a porcine RBC. These results as well as the frequency limits of these type of scattering experiments will be discussed.

10:45

2aBBb5. Scattering by a red blood cell: the importance of particle shape and orientation. Constantin-C. Coussios and John E. Ffowcs Williams (Dept. of Eng., Univ. of Cambridge, Trumpington St., Cambridge CB2 1PZ, UK, ccc20@eng.cam.ac.uk)

Red blood cells, which normally present themselves as biconcave discs, have traditionally been modeled as spherical scatterers of equivalent volume. The simplicity and symmetry of the spherical model does not however account for the effect of the angle of incidence of the incident field onto the red blood cell. Intuition tends to suggest that the field scattered by an asymmetric particle would differ if the particle were impinged upon broad-side or thin-side, an effect that is likely to become important if a collection of scatterers are aligned in any particular plane. Our objective is to compare the scattering by a red blood cell *of given volume* modeled as a sphere, to the result obtained by approximating its shape as a disc of varying aspect ratio. Results strongly suggest that the spherical model provides a good description for frequencies up to 20 MHz, beyond which particle shape becomes important. Although this effect could go undetected if the scattered field is only measured at angles of 180° (backscattering) and 90° relative to the direction of the incident field, the importance of particle shape and orientation might nevertheless form the base of a novel detection technique.

11:00

2aBBb6. Ultrasonic characterization of proteins in complex fluids. Donato Valdez, Marcel Gindre, Jean-Yves Lehuerou, Marcel Waks (Laboratoire d'Imagerie Parametrique, 15 rue de l'ecole de medecine, F-75006 Paris, France), and Wladimir Urbach (Laboratoire de Physique Statistique, Paris, France)

Our aim is the study of the physicochemical protein properties (volume, compressibility) as a function of hydration by an ultrasonic technique. For this purpose, we have used a biomimetic medium: reverse micelles, a system where we can control precisely water concentration. The compressibility of a medium is obtained by measuring the density and the ultrasonic celerity. The difficulty is increased in a complex fluid: nanometric size inclusions (micelles) dispersed in an organic solvent. To reach this goal, we have custom-built in our laboratory an apparatus allowing the determination of ultrasonic celerity, in very small volumes, with a relative precision of about 10 ppm. Using the effective medium theory, we have determined the compressibility of these inclusions with a precision better than 1%. With a spherical model for micellar inclusions and with the hydrodynamic radius of micelles obtained by x rays, we have estimated the compressibility of the water within micelles. We have also evidenced a difference in compressibility between the central core water and water bound to the polar surfactant headgroups. Finally by applying the mean field theory, we have obtained an estimate of the compressibility of proteins inserted within the micelles, at various hydration levels.

11:15

2aBBb7. A hybrid model to simultaneously determine ultrasound wave velocity and thickness of multilayered media. Ana V. G. Sousa, Wagner C. A. Pereira, and Joo C. Machado (Biomed. Eng. Prog., COPPE, Federal Univ. of Rio de Janeiro, P.O. Box 68510, Rio de Janeiro, Brazil 21945-970)

This work presents a method to simultaneously obtain the wave velocity (wv) and layer thickness (h) in a stratified media, in order to generate a parametrical image. The influence of refraction is also measured and included in the propagation model. In biomedical ultrasound (US), this certainly would help diagnostic procedures, as it identifies different tissues that conventional US cannot discriminate, like benign cysts from malignant tissue. It can also be applied to determine the degree of corrosion of metal ducts and delamination of aircraft painting. The model is based on geometrical acoustics and uses two transducers to obtain experimental data (256 echoes per layer): one active (3.4 MHz) and a hydrophone, which is moved laterally through 15 positions and rotated. A two-layered phantom (alcohol 6.2 mm and acrylic 6.0 mm) is immersed in a water tank and insonified. The echoes are captured and processed using a cross-correlation-based method. The water layer parameters (wv and h) have an accuracy of 1% rms and 2% for precision; alcohol layer, 4% and 9%; and acrylic layer, 12% and 15%, respectively. The experimental results demonstrate that the model has potential to be investigated and will be applied to *in vitro* data.

TUESDAY MORNING, 5 JUNE 2001 PDR 18, 8:00 TO 11:10 A.M.

Session 2aEA

Engineering Acoustics and Underwater Acoustics: Transducers for Underwater Vehicle Sonars

Thomas R. Howarth, Chair

NAVSEA Newport, 1176 Howell Street, Building 1171, Code 20, Newport, Rhode Island 02871

Chair's Introduction—8:00

Invited Papers

8:05

2aEA1. New sonar concepts and the sensor-related developments they require. Brian Houston and Joseph Bucaro (Naval Res. Lab., Code 7130, 4555 Overlook Ave., SW, Washington, DC 20375-5350)

Recently, exciting new approaches to the underwater vehicle acoustic sensing problem have been proposed which, if realized, could revolutionize how we approach underwater sonar. Two such schemes are the so-called ''virtual sonar'' which is formulated around Helmholtz integral processing and ''wireless'' systems which transfer sensor information through radiated rf signals. The full implementation of such new conceptual approaches, while presenting exciting possibilities, will nevertheless require a number of sensor-related technology advances. In this presentation, we will discuss those sensors and sensor-related developments which are desired or required in order to make practical such new sensor system concepts, and we will indicate the status of such developments. Finally, we will present several underwater applications from the perspective of exploiting these new sonar concepts. [Work supported by ONR.]

8:30

2aEA2. Broadband transducer requirements and development for a synthetic aperture sonar. Jose E. Fernandez, James T. Christoff, Kerry W. Commander, and Daniel A. Cook (Coastal Systems Station, Code R21, 6703 W. Hwy. 98, Panama City, FL 32407-7001, fernandezje@ncsc.navy.mil)

Broadband acoustic signal transmission and reception are key to enhancing the performance of underwater acoustical sensor systems to support functions like underwater acoustic communications and high resolution underwater imaging. In particular, synthetic aperture sonar (SAS) is an attractive technology for high-resolution imaging because of its ability to produce better cross range resolution than that of a real aperture array sonar of the same length. In addition, SAS systems can produce a cross range resolution that is independent of frequency and range. When the advantages of a SAS system are coupled with those gained from going to a broadband signal space, the performance enhancements can be significant. Specifically improvements in clutter rejection, object classification, and area coverage rate will be achieved. A description of the transducer requirements and expected payoffs for a SAS system designed to operate and image proud and buried objects in shallow waters will be presented. [Work supported by ONR.]

8:55

2aEA3. Cymbal panels: Low-frequency acoustic projectors for underwater vehicle SONARS. James F. Tressler (Naval Res. Lab., 4555 Overlook Ave., SW, Washington, DC 20375) and Thomas R. Howarth (Naval Undersea Warfare Ctr., Newport, RI 02841)

The Naval Research Laboratory is developing low frequency SONAR systems for an unmanned underwater vehicle (UUV) for shallow-water applications. Included among the onboard SONAR systems are three thin, lightweight, low-volume occupation acoustic sources known as cymbal panels. Each panel features miniature Class V flextensional driver elements, called cymbals, sandwiched between two stiff, lightweight radiating plates where each of the three panels is designed with a tailored cymbal element parameter in order to cover a different frequency band. NRL is placing each of these panels next to each other on the UUV in order to cover the desired frequency range from 1 kHz to over 6 kHz. A duplicate set of three panels is also placed on the opposite side of the UUV to provide some directivity means. This presentation will provide an overview of the program objectives, the acoustic source objectives, the element and panel designs, and recent underwater electroacoustic calibration data. [Work supported by Jan Lindberg of the Office of Naval Research Code 321SS and Bruce Johnson of the Office of Naval Research Code 321TS.]

9:20

2aEA4. Wideband 1–3 piezoelectric composite parametric mode projectors for UUV/AUV applications. Kim C. Benjamin (Naval Undersea Warfare Ctr., 1176 Howell St., Newport, RI 02841)

Unlike conventional parametric mode projectors, the 1–3 piezoelectric composite-based parametric mode projector affords a broad frequency range of operation. Due to its intrinsically low mechanical quality factor, Qm, the 1–3 piezoelectric composite material allows a wide range of primary frequencies to be used to excite the parametric mode effect. Furthermore, unlike conventional parametric mode projectors, the 1–3 piezoelectric composite substrates do not require inherently narrow band tuning circuits to enhance their half-wave resonance frequency. This presentation will report on the design and measured results for two 1–3 piezocomposite-based parametric mode projectors. The first case consisted of a 23-mm-diam by 6-mm-thick active layer driven with primary frequencies centered about 200 kHz. Beamwidths of 3.5 to 4.0 degrees were realized at differences frequencies ranging from 5–50 kHz. In the second case, a 15-mm by 15-mm by 25-mm-thick active layer was driven with primary frequencies centered about 60 kHz. In this case beam widths of 8.5 to 9.0 degrees were realized at difference frequencies ranging from 1 to 10 kHz.

9:45–10:00 Break

10:00

2aEA5. Geometrically phased ''doily'' array. W. Jack Hughes (Appl. Res. Lab., Box 30, State College, PA 16804, wjh2@psu.edu)

The theory and development of geometrically phased arrays will be presented. The array has evolved from a multibeam PVDF receiver, where its shape looks like a ''doily,'' to a 1–3 composite pulse-echo imaging array that can be a 1-D or 2-D array capable of 3-D imaging. The main advantage of the doily array is that it can form a set of steered beams without the need for electronically applying phase shifts, drastically reducing the cost and complexity of a system. The steering angle is a function of frequency, allowing the design of multiple steered beams or an imaging array if an FM sweep or large bandwidth pulse is used. In a manner similar to geometric shaping of a transducer for low side lobes, to form steered beams the transducer uses an amplitude function shaped as a cosine and sine aperture, shifting one signal 90 deg. Several arrays have been fabricated using electroplated and etched electrodes on piezoelectric materials such as PVDF and 1–3 composite materials. Beam formation and imaging results will be shown. [Work supported by ONR and DARPA.]

10:25

2aEA6. Design of a steerable parametric-mode array. Steve Forsythe and Kim Benjamin (Naval Undersea Warfare Ctr., 1176 Howell Ave., Newport, RI 02871)

A design is presented for a steerable parametric array that operates in the range of 40–60 kHz. The array design is based on three technologies: (1) Parametric array technology: the nonlinear interaction of two narrow high-frequency beams used to create a virtual endfire array with a very narrow beam at a much lower frequency; (2) Frequency-steered array design based on the ideas developed at Penn State University [Hughes and Thompson, J. Acoust. Soc. Am. **59**, 1040 (1976)]; (3) Use of 1-3 piezocomposite as the active element of the array. The array is steered in two directions with minimum complexity of the required drive electronics. Steering along one axis is done by a simple change of primary frequencies, while steering along the other axis uses time-delay between elements. This allows the resulting parametric beam to be steered independently in two directions. The array is designed by discretizing the real and imaginary parts of the inverse DFT of the desired far-field patterns of the primaries. Scaling the 2-D DFT of the discretized array pattern allows an easy verification of the final array performance. Implementation issues, such as efficient element interconnection, drive electronics, and fabrication, will also be discussed.

10:40

2aEA7. Forward-looking acoustic mine sensing. R. Daniel Costley (Miltec, Inc., Natl. Ctr. for Physical Acoust., Univ. of Mississippi, University, MS 38677), James M. Sabatier, and Ning Xiang (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, University, MS 38677)

A mine sensing system based on acoustic-to-seismic coupling sonifies the ground with high amplitude (120 dB), broadband (80–300 Hz) sound and measures the resulting ground vibration with a scanning laser Doppler vibrometer (LDV). Images produced by these scans show a distinct contrast in several frequency bands between ground vibrations over a buried mine and not over the mine. In order to reduce the risk to the vehicle and its operators and to allow a safe stopping distance for the vehicle, it is desirable that the system looks a significant distance in front of the vehicle. Experiments were conducted to investigate the feasibility of a forward-looking system. In a forward-looking system, both the sound source and the LDV are moved farther from the scanned area. This configuration both reduces the sound pressure level at the scanned area and decreases the angle at which the LDV beam strikes the ground. These effects reduce the contrast between the over-mine and off-mine signals. In addition, the image is distorted at the shallower LDV-ground angles. Results from these experiments will be presented and discussed. [Work supported by U.S. Army Communications-Electronics Command, Night Vision and Electronic Sensors Directorate.]

10:55

2aEA8. Acoustic intensity scattered from an elliptic cylinder. Kang Kim and Gerald C. Lauchle (Grad. Prog. in Acoust. and Appl. Res. Lab., Penn State Univ., P.O. Box 30, State College, PA 16804)

A computational study is described where a 2-D elliptic cylinder is insonified by a plane acoustic wave. The elliptic cross section of the cylinder has a fineness ratio of 5:1, the incidence angle of the plane wave is 30 and 60 degrees relative to the major axis of the ellipse, and $ka=20$, where a is the major axis of the elliptic cross section and k is the acoustic wave number. The calculations are performed using the finite element method of solution for partial differential equations. The MATLAB formulation solves the Helmholtz equation with reflection-free conditions imposed on the computational outer boundary, and rigid conditions imposed on the surface of the scatterer. Of particular practical interest in this study is the spatial distribution of the total active acoustic intensity, i.e., the sum of the incident and scattered intensity components. Active intensity amplitude, and the phase between pressure and particle velocity, are computed and compared to pressure amplitude only. The results show that there is significant phase distortion in the forward scattered direction that could be useful in localizing targets in active bi-static operations if $p-u$-type acoustic intensity probes were employed. [Work supported by ONR, Code 321 SS.]

TUESDAY MORNING, 5 JUNE 2001 SALONS V/VI/VII, 10:00 A.M. TO 12:30 P.M.

Session 2aEDa

Education in Acoustics: Hands-on Demos with High School Students

Uwe J. Hansen, Chair

Department of Physics, Indiana University, Terre Haute, Indiana 47809

This session is intended to communicate the excitement of ''doing'' acoustics to high school students in the Chicago area. Individual demonstrations with associated instrumentation will be presented in a brief introduction. In the subsequent hands-on session, presenters will guide individual students to discover the fun of acoustics by letting them ''twiddle the knobs.'' Topics available for student participation will range in sophistication from simple wave observation to normal mode mapping on musical instruments. While attendance as well as participation by conference attenders is welcome, hands-on participation is primarily designed for the students attending this session by special arrangement.

Session 2aEDb

Education in Acoustics: Undergraduate Research Poster Session

E. Carr Everbach, Chair
Department of Engineering, Swarthmore College, 500 College Avenue, Swarthmore, Pennsylvania 19081-1397

Contributed Papers

All posters will be on display from 11:00 a.m. to 2:00 p.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 11:00 a.m. to 12:00 noon and contributors of even-numbered papers will be at their posters from 1:00 p.m. to 2:00 p.m.

2aEDb1. Pipe impedance and sound spectrum measurements of the khaen and the sheng. Daniel R. Hoover and James P. Cottingham (Phys. Dept., Coe College, Cedar Rapids, IA 52402, drhoover@coe.edu)

Although the Asian free reed mouth organs all employ metal free reeds mounted in resonating pipes, there are significant differences among different families of these instruments. The sheng employs the free reed at the end of a closed pipe resonator with both conical and cylindrical sections. The khaen pipe has an unusual configuration for a wind instrument, with the free reed placed at approximately one-fourth the length of an open cylindrical pipe. Impedance curves have been measured for several pipes from a sheng and for artificial khaen tubes constructed from PVC pipe. These are compared with the sound spectra produced when the pipes are played. The sounding frequencies for both types of pipes are typically found to be above the natural frequency of the reed and close to, but slightly above, the frequency of the first impedance peak of the pipe. For khaen pipes constructed with the fundamental frequency of the pipe substantially below that of the reed, the sounding frequency is close to the second harmonic of the pipe. These results are consistent with those of some earlier studies as well as with theoretical considerations.

2aEDb2. Using your ears to learn about the acoustics of metal tubes. Konrad Kaczmarek (Yale Univ., P.O. Box 206657, New Haven, CT 06520)

The goal of this experimental work was to ascertain to what degree the physical properties and oscillatory behavior of a material can be understood using only one's ability to hear and a few simple tools such as a ruler, a scale, and a piano. More specifically, by striking a metal pipe in various controlled ways, listening to the resonance frequencies of various modes, and then matching the pitches heard to those found on a piano, such physical properties as the Young's modulus and the velocity of sound in a medium (air and solid) can be derived. Similarly, more complex scenarios such as end corrections in air resonances and transverse modes of vibration can be described fairly accurately using only these prescribed tools. The accuracy of this method was checked using traditional instrumentation (microphone and spectrum analyzer). [Advisor, Professor Robert E. Apfel.]

2aEDb3. Nonlinear scattering of crossed ultrasonic beams in the presence of turbulence: Experiments performed with pulses. Rebecca A. Manry and Murray S. Korman (Dept. of Phys., U. S. Naval Acad., Annapolis, MD 21402, korman@nadn.navy.mil)

The nonlinear scattering of two finite-amplitude mutually perpendicular crossed beams—interacting in the presence of turbulence—generates a sum frequency component that radiates outside the interaction region. Experiments are reported where two primary pulsed ($f_1 = 2.35$ MHz and $f_2 = 1.65$ MHz) focused beams are generated by 2.54-cm-diam concave spherical transducer units (T_1 and T_2) of focal length 14 cm. The 4-MHz receiving unit (R) is a 2.54-cm-diam circular plane array. The turbulence is generated by a $D = 0.635$-cm-diam submerged circular water jet (nozzle exit velocity 7 m/s) whose orifice is located at $56D$ from the interaction region. All transducer beam axes and jet axis form a common plane. A scattering region is formed at the intersection of the focal points of the primary beams. While T_1 and T_2 rotate on radius arms—always keeping the beams perpendicular—R is fixed. Symmetry suggests a scattering angle θ_*, where $\theta_* = 0$ defines forward scattering. Ensemble averaged rms pressure spectra (near the sum frequency) are measured (1) as a function of angle and fixed pulse duration $\tau = 100$ μs, and (2) at $\theta_* = 0$ and variable τ in an effort to compare spectral broadening with our theory.

2aEDb4. Objective measurements in Strauss Recital Hall at the University of Nebraska. Jessica L. Rock and Lily M. Wang (Architectural Eng. Prog., Univ. of Nebraska–Lincoln, 200B PKI, Omaha, NE 68182-0681)

This poster presents measured acoustical data in a performing arts hall, obtained from three different measurement techniques and with various sound source and receiver positions. The room under study is Strauss Recital Hall, located on the campus of the University of Nebraska at Omaha. Sound decay measurements are taken between selected source and receiver positions in three ways: triggered by an impulse source such as a balloon pop; triggered by the end of steady-state white noise; and using a maximum length sequence (MLS) source. The measured responses are used to calculate the reverberation time, clarity index, strength, interaural cross-correlation coefficient, and other room acoustic measures. Analysis of the data focuses on how these objective measures change across the Hall and how they differ between measurement techniques. Additionally, the data obtained are compared to values from other renowned halls.

2aEDb5. Nonlinear scattering of crossed ultrasonic beams in the presence of turbulence: Experiments performed with continuous waves. Marlene Roush and Murray S. Korman (Dept. of Phys., U. S. Naval Acad., Annapolis, MD 21402, korman@nadn.navy.mil)

Experiments are reported where two primary cw ($f_1 = 2.35$ MHz and $f_2 = 1.65$ MHz) beams are generated by 2.54-cm-diam transducer units (T_1 and T_2) of focal length 14 cm. The 4-MHz receiving unit (R) is a 2.54-cm-diam circular plane array. Turbulence is generated by a $D = 0.635$-cm-diam submerged water jet (exit velocity 7 m/s) whose orifice is $56D$ from the interaction region. All transducer beam axes and jet axis form a common plane. The ensemble averaged spectrum (near 4 MHz) is measured (using a Tektronix digital oscilloscope) as a function of the

''symmetry'' scattering angle θ_*. After transferring the data to a computer, ''notebooks'' in Mathematica are developed to curve fit the Doppler shift $\langle f_d \rangle$ and the $n=2$nd, 3rd, and 4th spectral moments $\langle (f-\langle f_d \rangle)^n \rangle$ vs θ_* to the theory in an effort to measure the mean flow and turbulent velocity correlation functions of the turbulence [Korman and Parker, in *13th ISNA, Bergen 1993, Advances in Nonlinear Acoustics*, edited by H. Hobaek (World Scientific, Singapore, 1993), pp. 650–655]. For example, for $n=2$, 2nd order correlations $\langle v_x^2 \rangle$, $\langle v_x v_y \rangle$, and $\langle v_y^2 \rangle$ can be measured.

TUESDAY MORNING, 5 JUNE 2001 WABASH ROOM, 9:30 TO 11:45 A.M.

Session 2aNS

Noise, Architectural Acoustics and Psychological and Physiological Acoustics: Measurement Procedures in Social Studies on Soundscapes and Annoyance

Brigitte Schulte-Fortkamp, Chair
Physics-Acoustics, Carl von Ossietzky University, Carl Von Ossietzky Street, Oldenburg D-2611, Germany

Chair's Introduction—9:30

Invited Papers

9:35

2aNS1. Neighborhood soundscapes: Measurement and identification issues. Ronny Klåboe (Inst. of Transport Economics, P.O. Box 6110 Etterstad, N-0602 Oslo, Norway)

The term neighborhood soundscape has been coined to describe the amount of road traffic noise in the immediate neighborhood of a dwelling. A simple indicator of how much the road traffic noise level in the neighborhood exceeds the noise level at the dwelling has been developed using a geographical information system. By applying the indicator with data from several socio-acoustic surveys it has been shown that people living in a relatively noisy neighborhood are significantly and substantially more annoyed than follows from the residential road traffic noise level alone. However, this does not prove that the heightened annoyance is caused by the noise exposure in the neighborhood. Neighborhood road traffic also gives rise to traffic insecurity and local air pollution, and indicators of these environmental exposures are additional possible effect modifiers. To separate the different positively correlated modifying effects, it is therefore not enough to measure the neighborhood soundscape, but also measure relevant aspects of the environscape. To statistically separate the different modifying effects requires that the respective measures are sufficiently precise. To go beyond the purely statistical associations it is also necessary to link each measure to the respective psychscape impacts and model the relevant parts of the psychscape.

10:00

2aNS2. Impression of sound environments formed by verbal explanation. Kouji Abe, Toshio Sone (Akita Prefectural Univ., 84-4 Ebinokuchi, Tsuchiya, Honjo 015-0055, Japan, koji@akita-pu.ac.jp), Yôiti Suzuki (Tohoku Univ., Sendai 980-8577, Japan), and Kenji Ozawa (Yamanashi Univ., Kofu 400-8511, Japan)

Our research results suggested that additional verbal and visual information on sound environments may recall existing impressions of specific sounds and affect the auditory evaluation of environmental sounds (Y. Suzuki *et al.*, Proc. Internoise 2000, 2000, pp. 2285–2291). In order to verify this result, we conducted a new series of experiments as follows: At first, only a written verbal explanation of sound environments on sound fields and sound sources was presented to new subjects. In this experiment, we expected that some impression on the sound environments could be formed by the verbal explanation. Then, an experiment in which the stimulus sound only was presented as well as another experiment in which the stimulus sound was presented along with the verbal explanation were also conducted. The results from these experiments were analyzed by factor analysis. The results showed that the factors to evaluate the sound environments based only on the verbal information were similar to those in the evaluation with listening to the sounds. Moreover, analysis of the significant changes of factor scores among experiments showed that the performed impression on the sound environments by the verbal information affects evaluation of the sound.

10:25

2aNS3. How structural parameters values can act on subjects' preference judgments for sounds transmitted by a simple component of an exterior wall. Cathy Marquis-Favre and Julien Faure (LASH/DGCB, URA CNRS 1652, ENTPE, rue M. Audin, 69518 Vaulx-en-Velin, France, cathy.marquis@entpe.fr)

For a long time, the vibroacoustic behavior of structures was studied from a physical point of view with the goal of reducing sound levels and understanding physical phenomena. It has led researchers to elaborate upon physical models for sound radiation and transmission prediction, and also to establish vibroacoustic indicators such as radiation efficiency or sound transmission loss for acoustic performance criteria. Within the last decade, qualitative aspects of sounds have been integrated into the physical analysis to take into account sound perception and the expectation of subjects. This has yielded noticeable improvements in the sound quality of sources in various industries, mainly in the automotive industry. Our work aims at improving the acoustic performance of building structures which act as input filters to annoying sounds of the inner environment. The approach is based on coupling of the physical

and perceptual analyses. We will present how structural parameter values and their accuracy can act in a non-negligible manner on the similarity and preference judgment of subjects for broadband noises transmitted by a simple panel of an exterior wall. This is observed even in the case of structural parameter variations that induce no significant modification of vibroacoustic indicators such as sound transmission loss.

10:50–11:05 Break

11:05

2aNS4. Soundscapes in the sense of reaction to sound and vibration. Brigitte Schulte-Fortkamp (Dept. of Phys./Acoust., Oldenburg Univ., 26111 Oldenburg, Germany, brigitte@aku.physik.uni-oldenburg.de)

Evaluation of soundscapes in urban areas has to focus on different aspects like the structure of urban areas, people living in those areas, architectural and social parameters designing those areas, and acoustical and visual parameters. A measurement is needed which refers to objective and subjective parameters. The structure of the residential area, the combination of noise sources, and the soundscape are for the judgment of annoyance as well as important as subjective parameters which are relevant according to the people's point of view; moreover the relationship of both define the background for assessments. When carrying out a survey on noise and vibration different subject-centered methodological procedures have been taken into account to develop a suitable measurement procedure. The survey focusing on perception of sound and vibration was conducted with about 600 subjects from different European countries. Procedures and results of these tests will be presented with respect to improvement of social surveys, especially addressing the meaning of noise and vibration in a defined environment. [Work supported by BRITE-EURAM project IDEA PACI/BE97-4056.]

Contributed Paper

11:30

2aNS5. Correlation between acoustical sensory profiles and physical measures of product sound. David L. Bowen and Richard H. Lyon (RH Lyon Corp, 691 Concord Ave., Cambridge, MA 02138)

''Sensory profiles'' (SPs) are used extensively for judgments of taste, odor, and texture—situations where a useful physical metric is not available. An acoustical SP can be obtained using a panel of critical listeners, and can provide a bridge between descriptions of a sound and the sound quality (SQ)—judgments that consumers make about a product based on the sound it makes. But as descriptors, SPs also offer the opportunity to connect to physical measures of the sound. If the description is ''tinny'' or ''shrill,'' there is expectation of a high-frequency bias to the sound. If the description is ''throbbing'' or ''knocking,'' a modulation of the sound is expected. Since spectral balance and modulation are measurable, some kind of transformation between the SP and physical measurements is reasonable. This presentation focuses on the issue of the robustness of such a relationship, since it is desired that the physical measurements be able to predict or relate to a significant variation in the SP. Examples are given from studies of vacuum cleaner and washing machine sounds of the sensitivity of SQ to changes in the SP, and therefore of the variations in physical measures that are significant in terms of variations in SQ. [Work supported by NSF.]

TUESDAY MORNING, 5 JUNE 2001 PARLOR A, 8:15 TO 9:45 A.M.

Session 2aPAa

Physical Acoustics: Interfaces, Particles and Bubbles

Kerry W. Commander, Chair
Coastal Systems Station, Code R21, Panama City, Florida 32407-7001

Contributed Papers

8:15

2aPAa1. Passive stabilization of capillary bridges in air with acoustic radiation pressure and the equilibrium shape of bridges. Mark J. Marr-Lyon, David B. Thiessen, and Philip L. Marston (Dept. of Phys., Washington State Univ., Pullman, WA 99164-2814)

Long liquid cylinders ordinarily become unstable because of a capillary instability originally studied by Rayleigh and Plateau. In the present research liquid bridges in air were acoustically stabilized to 43% beyond the natural limiting length identified by Rayleigh and Plateau [Marr-Lyon, Thiessen, and Marston, Phys. Rev. Lett. (to be published)]. The stabilization was achieved by placing the bridge at the velocity antinode of a standing wave and selecting the wavelength such that the surface-averaged radiation pressure of the sound field increases with increasing bridge radius. The tests were performed on NASA's KC-135 aircraft in weightless conditions on bridges having volumes equal to that of a circular cylinder of the same length as the bridge. In the standing wave, the bridge becomes elliptical because of the spatial distribution of the radiation pressure. Calculations of the scattering of sound by elliptical cylinders using Mathieu functions show that elliptical deformation only weakly affects the average increase in the radiation pressure with the local bulging of a liquid bridge normally associated with the Rayleigh-Plateau instability. [Work supported by NASA.]

8:30

2aPAa2. Simulation of particle motions in an acoustophoresis device. Todd L. Brooks and Robert E. Apfel (Dept. of Mech. Eng., Yale Univ., New Haven, CT 06520-8286, todd.brooks@yale.edu)

Particles in a liquid host can be continuously fractionated by combining a laminar flow with an ultrasonic standing wave oriented perpendicular to the flow direction. This kind of separation method can separate particles based on their relative compressibilities as well as on mass density and size. Experimental work has focused on separating particles ranging in size from 5 to 55 μm in resonant devices operating from 100 kHz to 1 MHz. A numerical simulation based on a hierarchical tree method has been developed to simulate particle motion in such a device. The flow field, primary acoustic forces, and secondary interparticle acoustic forces are all taken into account. Full three-dimensional simulations can be performed. This method may be useful in other application areas where large numbers of particles, drops, and/or bubbles are under the influence of acoustic forces. [Work supported by NASA through Grant No. NAG8-1351.]

8:45

2aPAa3. Acoustic cavitation thresholds and stable oscillations of vapor-gas bubbles. Nail A. Gumerov (Inst. for Adv. Computer Studies, Univ. of Maryland, College Park, MD 20742)

The effect of rectified diffusion on the behavior of gas bubbles and the effect of rectified heat transfer to vapor bubbles are well established. These effects may be equally important for dynamics of vapor-gas bubbles and control the acoustic cavitation thresholds. To describe these effects a model of a spherical vapor-gas bubble is used. This model accounts for heat and mass transfer in the two-component system, including Soret–Dufour effects, kinetics of evaporation, surface tension, liquid viscosity, and compressibility. Equations for acoustic cavitation thresholds and stable bubble oscillations are obtained using asymptotic averaging techniques valid to second order in the amplitude of the acoustic field. Good agreement between the previous results for gas and vapor cavitation thresholds and the present computations is observed in the limiting cases of single-component bubbles. The effects of temperature, kinetics of evaporation, supersaturation, and others on the acoustic cavitation thresholds and stable bubble oscillations are studied and discussed. The existence of stable oscillations and cavitation thresholds of the bubble due to the second resonance of a vapor-gas bubble is predicted. These states occur due to the two-component nature of the bubble and appear at low inert gas concentrations near the liquid boiling point.

9:00

2aPAa4. Single bubble sonochemistry and sonoluminescence. Yuri T. Didenko and Kenneth S. Suslick (Dept. of Chemistry, Univ. of Illinois–Urbana-Champaign, 601 S. Goodwin Ave., Urbana, IL 61801)

An isolated, single vapor-gas bubble pulsating in a standing acoustic field can emit flashes of light. This peculiar phenomenon is known as single bubble sonoluminescence, SBSL. We find that these bubble pulsations in water are accompanied with the formation of radicals and molecular products. In this paper, the yields of hydroxyl radicals and nitrite ions formed inside the bubble were measured for the first time. The chemical and light efficiency of acoustic cavitation and the diffusion rate of nitrogen inside the bubble during its pulsation were calculated using experimental data. The energy efficiency of OH radical formation by a single bubble is comparable to that in multibubble cavitation. However, the energy efficiency of light emission is much higher for SBSL. The diffusion rate of nitrogen inside the bubble is in good agreement with that predicted by the dissociation hypothesis. [Work supported by NAVY/ DARPA.]

9:15

2aPAa5. Towards an intense bubble. Anthony Khong, Elizabeth Doschek, Jonathan Mori, Robert Apfel (Dept. of Mech. Eng., Yale Univ., 9 Hillhouse Ave., New Haven, CT 06520, anthony.khong@yale.edu), and Jeffrey Ketterling (Riverside Res. Inst., New York, NY 10036)

A main goal in SBSL remains the ability to form a bright and hot bubble. One recent postulate suggests that lowering the resonance frequency should favor a stronger luminescing bubble. It is also theorized that the chemical composition of the bubble can be streamlined to induce a more violent collapse. The water vapor content inside the bubble is considered. The premise of this conjecture is that water vapor molecules, present beyond equilibrium concentrations, act to siphon energy from the bubble, and in the process undergo thermal decomposition. Hence, reducing the water vapor content should eliminate such energy loss. The current study focuses on this issue. Keeping similar driving conditions, there is interest in using aqueous salts to depress the vapor pressure of the water in tandem with a reduction in the water temperature. A series of alkali metal halides will be examined. In a preliminary investigation, a 9% lithium chloride solution saturated with 1% argon gave stable SL at room conditions. With air, no SL was detected. A further extension of this work would be to determine the bubble temperature which has largely been inferred. [Work supported by DARPA through a subcontract from the University of Washington.]

9:30

2aPAa6. Focused pulse of pressure for boosting sonoluminescence. Jean-Louis Thomas, Yoel Forterre, and Mathias Fink (Lab. Ondes et Acoustique, CNRS UMR 7587, Univ. Paris VII, ESPCI, 10, rue Vauquelin 75231 Paris, France)

Single bubble sonoluminescence is characterized by a great concentration of energy conducting to the generation of UV photon with ultrasound of centimetric wavelength. Several teams have tried to again increase more of the energy concentration, but most of these trials have rapidly reached the boundaries of the narrow stability domain of sonoluminescence. The idea developed in this work is to overcome this limitation by adding a high-frequency pressure pulse to the monochromatic low-frequency field used in the now classical setup. For instance, the pressure in the neighborhood of the bubble can be dramatically increased during a single collapse of the bubble. This pressure pulse is generated with eight high-frequency transducers adaptively focused on the bubble and synchronized with its low-frequency cycle. With this setup, a transient amplitude of 8 atmospheres has been reached on the bubble, without destroying it, and a significant increase of the emitted light intensity has been observed on a single flash. This process is no longer limited by the domain of stability of sonoluminescence and could lead, in the future, to much more efficient inertial confinement of the gas contained in the bubble.

Session 2aPAb

Physical Acoustics: Sensors and Devices

Robin O. Cleveland, Chair
Department of Aerospace and Mechanical Engineering, Boston University, 110 Cummington Street, Boston, Massachusetts 02215

Contributed Papers

10:00

2aPAb1. Pulse-echo ultrasonics from the correlations of thermal phonons. Richard L. Weaver and Oleg I. Lobkis (Theoretical and Appl. Mech., Univ. of Illinois, Urbana, IL)

A high-sensitivity acoustic emission transducer in contact with a non-insonified solid body generates a noise signal, a significant fraction of which is due to thermal fluctuations in the body. This is established by comparisons of the spectral power density of its noise and the spectral power density consequent to the application of a calibrated source. This high sensitivity is some 30 dB greater than that of other ultrasonic transducers. As established elsewhere [Weaver and Lobkis, this meeting] the correlations of a diffuse field are related to the local density-of-ultrasonic-modes, and thus to local responses in pulse/echo. It is therefore hypothesized that the autocorrelation function of the noise in a high-sensitivity transducer should equal the pulse-echo signal generated by that transducer. The hypothesis is investigated by constructing that autocorrelation function (the autocorrelation appears to converge after only a few tens of milliseconds of data collection) and comparing it with the pulse/echo signal. The two waveforms are found to be substantially identical. The reasons for the differences are discussed, and possibilities for application are considered. [Work supported by NSF.]

10:15

2aPAb2. Diffraction correction and attenuation of high-frequency acoustic pulses in a relaxing medium. Yefim Dain and Richard M. Lueptow (Dept. of Mech. Eng., Northwestern Univ., Evanston, IL 60208, rlueptow@northwestern.edu)

The influence of intrinsic absorption in a relaxing medium and the diffraction correction for the magnitude of the acoustic pressure averaged over the surface of a receiver was investigated. An exact formula for the damped acoustic pressure averaged over the surface of the receiver was obtained for high-frequency modulated pulses in a mono-relaxing medium. The formulation is based on the solution of a modified wave equation in a relaxing medium. Based on asymptotic analysis, separate effects result for the attenuation of the pulse envelope and for the modulation of the carrier frequency for high-frequency pulses with an arbitrary envelope shape. The resulting distortion of the energy spectrum for arbitrary pulses depends on the medium's relaxation frequency and the acoustic carrier frequency. These results are useful in predicting the response of transducers in a relaxing medium. [Work supported by Ford Motor Company.]

10:30

2aPAb3. Cuts with a negative Poisson's ratio in Si. Svetlana Tokmakova (N. Andreyev Acoust. Inst., Shvernik Str. 4, Moscow 117036, Russia)

Poisson's ratio is the ratio of lateral strain in the direction n to the longitudinal strain during the stretching of a cylindrical rod along rod axis m. The Poisson ratio is a constant in an isotropic solid. The Poisson ratio of an anisotropic solid depends on directions m and n. In the paper the stereographic projections of Poisson's ratio for some crystals were computed (germanium, quartz, lithium niobate, langasite, zinc, copper, etc.). From the stereographic projections the Poisson ratio for any directions of m and n were calculated. Orientations of axes of a rod m and directions of lateral strain n with extreme (negative) values of Poisson's ratio were determined. Only for Si and HgCl was a decrease of rod cross-section under the uniaxial compression predicted for some orientations of the rod axis. For the Si rod the axis of the rod should lay within the 30-deg cone (the axis of the cone is the [001] direction of the Si crystal). Materials having a negative Poisson ratio may find a lot of applications, for example, as electrodes that amplify the response of piezoelectric sensors [Baughman *et al.,* Nature (London) **392**, 362–365 (1998)].

10:45

2aPAb4. A nonlinear analysis of piezoelectric coated laminae. M. Cengiz Dokmeci (Istanbul Tech. Univ., Gumussuyu, P.K.9, Istanbul 80191, Turkey) and G. Askar Altay (Bogazici Univ., Bebek, Istanbul 80815, Turkey)

In relation to vibrations of piezoelectric materials subject to strong electric fields and large strains, a system of 2-D nonlinear laminae equations is derived by the use of Mindlins kinematic hypothesis for each layer. The laminae is coated completely with perfectly conducting electrodes on both its faces, and it may comprise any number of bonded layers, each with a distinct but uniform thickness and electromechanical properties. By use of a recent variational principle [Altay and Dokmeci, Thin Walled Struct. **39**, 95–109 (2001)], the system of equations is consistently formulated in both invariant, differential and variational forms. The effects of elastic stiffnesses of, and the interactions between, layers of laminae are all taken into account. The system of equations governs the extensional, flexural, and thickness shear as well as coupled vibrations of piezoelectric laminae. Special types of vibrations, geometry, and material properties are recorded. Besides, the uniqueness is investigated in solutions of the initial-mixed boundary value problems defined by the linearized system of laminae equations. [Work supported in part by TUBA.]

11:00

2aPAb5. Crosstalk between acoustic sensors in a MEMS array. Allan D. Pierce, Harley T. Johnson, Robin O. Cleveland, Daniel N. Pascual, and Thomas G. Bifano (Dept. of Aerosp. and Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215)

The sensor array considered is in bare essentials a regular rectangular array of square holes in a slab of single crystalline silicon. The actual sensing of an incident pressure wave is achieved by optical sensing of the acoustically induced vibrations of recessed membranes (plates) within the holes. The present study addresses the crosstalk question by investigating the effect of the spacing between the holes (sensors) on the vibration of the individual membranes. The no crosstalk limit is taken as when adjacent holes are asymptotically very far apart. When they are closer together the amplitude changes because of two primary effects: (1) elastic waves originating at each of the individual holes and stimulating vibrations of the membranes in adjacent holes, and (2) reradiation of sound from one hole and subsequent diffraction into an anjacent hole. Complementary analyses

are carried out using a finite element program and the geometrical theory of diffraction. For the array under study the dominant effect results from diffraction through the fluid itself; structure-borne sound has a minor effect.

11:15

2aPAb6. Microfluidic amplification of ultrasound with application to MEMS hydrophones. Robin O. Cleveland, Daniel N. Pascual, Harley T. Johnson, Thomas G. Bifano, and Allan D. Pierce (Dept. of Aerosp. and Mech. Eng., Boston Univ., Boston, MA 02215)

Micro-electro-mechanical systems (MEMS) were used to fabricate a device with passive fluidic components that amplified the motion associated with ultrasonic waves in water. Two types of fluid components were etched in silicon, an acoustic horn, and an acoustic resonator. The fluid motion was detected by a small silicon diaphragm which moved with the sound wave. The diaphragm motion was measured with a laser Doppler vibrometer. Experiments were carried out for acoustic excitations in the frequency region of 0.5 to 5 MHz. Over this frequency range uniform thickness diaphragms were found to support multiple resonance modes and were found to be poor sensing surfaces. Composite diaphragms, with a stiff pedestal supported by compliant edge fixtures, provided a superior sensing surface. The motion of the diaphragm placed at the throat of an acoustic horn demonstrated broadband amplification with a gain of 6. An organ-pipe type resonator provided a narrow-band gain of 3. The horn and resonator structures could be incorporated into most types of MEMS hydrophone devices which use diaphragms to sense sound waves. [Work supported by DARPA.]

11:30

2aPAb7. The scaling of acoustic streaming for application in microfluidic devices. Kenneth Frampton (Vanderbilt Univ., Box 1592 B, Nashville, TN 37235)

Recent interest in MEMS devices in general, and in microfluidic devices specifically, has spurred a great deal of research into the behavior of fluids in very small-scale devices. Many novel techniques have been proposed for the propulsion of fluids in small-scale devices including peristaltic and electrokinetic. More recently, investigations have considered the use of acoustic streaming, that is, the imposition of steady fluid momentum via nonlinear acoustic effects. The purpose of this talk is to give an overview of the physics of acoustic streaming, to discuss the various physical phenomena which generate the effect, and to highlight the favorable scaling issues of acoustic streaming that make it a viable alternative in microfluidic devices.

TUESDAY MORNING, 5 JUNE 2001 GRAND BALLROOM, 8:45 TO 10:45 A.M.

Session 2aPPa

Psychological and Physiological Acoustics: Masking and Loudness

Mary Florentine, Chair
Department of Speech Language Pathology and Audiology, Northeastern University, (133FR), 70 Forsyth Street, Boston, Massachusetts 02115

Contributed Papers

8:45

2aPPa1. Effect of masker harmonicity on threshold: Auditory filter envelope-modulation rate variability versus equidistant spacing of partial frequencies. William C. Treurniet and Darcy R. Boucher (Commun. Res. Ctr., Box 11490, Sta. H, Ottawa, ON K2H 8S2, Canada, bill.treurniet@crc.ca)

In a previous paper [W. C. Treurniet and D. R. Boucher, J. Acoust. Soc. Am. **109**, 306–320 (2001)], a model was presented which successfully accounted for the lower threshold obtained using a harmonic masker instead of a similar inharmonic masker. For the harmonic masker, the frequencies of partials were separated by a fixed amount, so the envelope modulation-rates of auditory filter outputs remained the same across filters. However, for the inharmonic masker, the interval between adjacent partials was not fixed and this resulted in a decreased uniformity of modulation-rates across filters. The model proposed that the lower uniformity impedes detection of a probe-induced change in the modulation rates, thus accounting for the masked threshold difference. This paper shows that an inharmonic masker yields results similar to a harmonic masker provided that the modulation rates are uniform across affected auditory filters. Thus, the lowered threshold associated with a harmonic masker appears to arise from invariant modulation-rates across auditory filters, and does not require that all partial-frequencies be integer multiples of a fundamental.

9:00

2aPPa2. Upward spread of Schroeder-phase maskers. Jennifer Lentz, Marjorie Leek (Army Aud. & Speech Ctr., Walter Reed Army Medical Ctr., 6900 Georgia Ave., NW, Washington, DC 20307, jennifer.lentz@na.amedd.army.mil), and Laura Dreisbach (San Diego State Univ., San Diego, CA 92182-1518)

Harmonic complexes with phases selected according to the Schroeder algorithm can produce large differences in masking, depending on whether phases increase (positive Schroeder) or decrease (negative Schroeder) with frequency. This finding has been attributed to an interaction between the phase characteristics of auditory filters and the stimulus. The current study investigates effects of filter asymmetry on masking by harmonic complexes using an upward spread of masking paradigm. Maskers were Schroeder phase maskers with frequencies from 200 to 2000 Hz, and signal frequencies ranged from 1000 to 4000 Hz. When the signal fell within the masker bandwidth, it was added in-phase with the identical masker component. At signals below 2000 Hz, negative Schroeder maskers produced more masking than the positive maskers. When the signal frequency was above the masker, there was a rapid decrease in masking for both maskers, but for most subjects, the positive masker became more effective than the negative. This shift in the Schroeder masking effect may be related to phase changes occurring in the lower skirts of the auditory filters. [Work supported by NIH.]

9:15

2aPPa3. Relation between the slope of the rate-level curve, the maximum driven firing-rate, and the neuronal dynamic range in auditory primary afferents of the cat. Lance Nizami, JoAnn McGee, and Edward J. Walsh (Boys Town Natl. Res. Hospital, 555 N. 30th St., Omaha, NE 68131, nizami@boystown.org)

The power of the auditory primary afferent to indicate intensity change depends partly upon the rate-of-change-of-firing with sound-pressure-level, i.e., the slope of the rate-level curve. The latter was estimated for each of 778 rate-level-functions in the cat (McGee, MS thesis, Creighton, 1983) by fitting lines to the points on the curve that fall between (1) $Rs+[c/100](R\max-Rs)$ and (2) $Rs+[(100-c)/100](R\max-Rs)$, for $c=30$. Rs is spontaneous firing-rate and $R\max$ is maximum firing-rate. c specifies a pair of points on the curve, symmetric around the centroid. Each slope was divided by the respective maximum driven firing-rate, $R\max-Rs$, to obtain the adjusted slope (in 1/dB). Each adjusted slope was then plotted versus the neuron's dynamic range (in dB), whose endpoints, the threshold and the saturation level, are given by $c=1$. The resulting plot is smooth and hyperbolic. Threshold and saturation level have also been defined by expressions (1) and (2) in a logistic rate-level equation (Nizami, 24th ARO, 2001) that yields (3) adjusted slope $=[\ln[(100-c)/c]]/2(\text{dynamic range})]$. For $c=1$ and $c=2$, these curves neatly sandwich the empirical data. An afferent's power of discrimination is intimately related to both the maximum driven firing-rate and the neuronal dynamic range.

9:30

2aPPa4. Derivation of intensity-discrimination functions from loudness functions. William S. Hellman (Dept. of Phys. and Hearing Res. Ctr., Boston Univ., 590 Commonwealth Ave., Boston, MA 02215) and Rhona P. Hellman (Northeastern Univ., Boston, MA 02115)

Intensity-discrimination functions are computed from their concomitant loudness functions in a procedure where the only adjustable parameter sets the scale to match the measured intensity-discrimination data. The equation connecting loudness to intensity discrimination is determined from monaural loudness- and intensity-discrimination measurements at 1 kHz. Predictions of intensity-discrimination functions are then computed for a low-frequency tone masked by an adjacent high-pass noise, for broadband noise, and for high-frequency tones at 12.5 and 16 kHz. Results show that the derived intensity-discrimination functions closely capture the overall shape of the experimental data. A U-shaped function is obtained for the partially masked low-frequency tone, whereas a function approximating Weber's law is observed for broadband noise and also at 16 kHz. The good agreement between the calculated and measured functions implies that the form of the intensity-discrimination functions is strongly dependent on local variations in shape of the loudness functions. Together, these results provide further evidence to support the notion of an intimate relation between loudness and intensity discrimination. [Work supported in part by NEDO, Japan.]

9:45

2aPPa5. Loudness functions for long and short tones. Mary Florentine (Inst. for Hearing, Speech and Lang. and Dept. of Speech-Lang. Path. and Audiol., 133 FR, Northeastern Univ., Boston, MA 02115, florentin@neu.edu), Michael Epstein, and Søren Buus (Inst. for Hearing, Speech and Lang., Northeastern Univ., Boston, MA 02115)

This study tests the equal-loudness ratio hypothesis [Florentine *et al.*, J. Acoust. Soc. Am. **99**, 1633–1644 (1996)], which states that the loudness ratio between equal-SPL long and short tones is independent of SPL. The amount of temporal integration (i.e., the level difference between equally loud long and short sounds) is maximal at moderate levels. Therefore, the equal-loudness ratio hypothesis predicts that the loudness function is shallower at moderate levels than at low and high levels. Equal-loudness matches and cross-modality string-length matches were used to assess the form of the loudness function for 200- and 5-ms tones at 1 kHz. Results from nine normal listeners show that (1) the amount of temporal integration is largest at moderate levels in agreement with previous studies, and (2) the loudness function is shallowest at moderate levels. For eight of the nine listeners, the loudness ratio between the 200- and 5-ms tones is approximately constant, except at low levels where it tends to increase. The average data show good agreement between the two methods, but discrepancies are apparent for some individuals. These findings support the equal-loudness ratio hypothesis, except perhaps at low levels. [Work supported by NIH/NIDCD Grant No. R01DC02241.]

10:00

2aPPa6. Equal-loudness contours at high frequencies reconsidered. Rhona Hellman (Dept. of Psych., Northeastern Univ., 360 Huntington Ave., Boston, MA 02115), Hisashi Takeshima (Sendai Natl. College of Technol., Sendai 989-3124, Japan), Yôiti Suzuki (Tohoku Univ., Sendai 980-8577, Japan), Kenji Ozawa (Yamanashi Univ., Kofu 400-8511, Japan), and Toshio Sone (Akita Prefectural Univ., Honjo City 015-0055, Japan)

To add to the database and to clarify the spacing of the equal-loudness contours at high frequencies, equal-loudness relations determined from 1 to 16 kHz in a recent study [Hellman *et al.*, J. Acoust. Soc. Am. **103**, 2812 (1998)] are reevaluated. Relative to the linear loudness-level function with a slope of 1.0 observed for a standard 1-kHz tone, above 10 kHz the overall shapes and slopes of the loudness-level functions are both level and frequency dependent. Below 60 phons, the slopes of the loudness-level functions at 12.5 kHz and higher increase progressively with frequency from 1.31 at 12.5 kHz to 1.79 at 16 kHz. Conversely, above 60 phons the slopes decrease from 0.98 at 12.5 kHz to 0.74 at 16 kHz. The data imply that for frequencies between 1 and 10 kHz, the spacing between the equal-loudness contours is independent of loudness level. In contrast, above 10 kHz the equal-loudness contours are more closely spaced below 60 phons than at higher loudness levels. Moreover, in accord with the early work of Robinson [IRE Trans. Audio **1**, 6–13 (1958)], the higher the frequency, the more strongly does the spacing vary with loudness level. [Work supported by NEDO, Japan.]

10:15

2aPPa7. Effect of duration and level of the recalibration tone on the amount of loudness recalibration. Bärbel Nieder (Inst. of Hearing, Speech, and Lang. and Dept. of Speech-Lang. Path. and Audiol., 133 FR, Northeastern Univ., Boston, MA 02115, nieder@research.neu.edu), Søren Buus (Inst. of Hearing, Speech, and Lang. and Commun. and Dig. Sig. Proc. Ctr., Northeastern Univ., Boston, MA 02115), Mary Florentine (Inst. of Hearing, Speech, and Lang., Northeastern Univ., Boston, MA 02115), and Bertram Scharf (Inst. of Hearing, Speech, and Lang., Northeastern Univ., Boston, MA 02115)

The amount of loudness recalibration (i.e., the drop in loudness of a moderate-level tone caused by a preceding intense recalibration tone of the same frequency) was measured as a function of the recalibration tone's duration and level using a 2AFC procedure. The 500-Hz test tone—presented at 60 and 70 dB SPL—and the 2500-Hz variable-level comparison tone both lasted 200 ms. Results for 10 normal listeners show that 5-ms recalibration tones yielded only 3 (test tone at 60 dB SPL) to 4 dB (at 70 dB SPL) of recalibration, whether their level was 80, 95, or 110 dB SPL. In contrast, 200- and 500-ms recalibration tones at 80 and 95 dB SPL (200 ms only) yielded 6 (at 60 dB SPL) to 10 dB (at 70 dB SPL) of recalibration, again with no apparent effect of level. Note that 5-ms recalibration tones at 95 dB SPL yielded much less recalibration than 200-ms recalibration tones at 80 dB SPL, despite their nearly equal loudness. These data indicate that recalibration is not governed by the loudness of the recalibration tone and that recalibration-tone duration is a crucial parameter for recalibration of loudness. [Supported by NIH/NIDCD R01DC02241.]

10:30

2aPPa8. Further progress with loudness adaptation inside and outside the speech range. Sophia Boudouris, Kathleen Cross, Suzanne Boyce, Laura Kretschmer, David E. Sandman, and Ernest M. Weiler (ML #394, Psycho-Acoust. Lab., CSD, CAHS, Univ. of Cincinnati, Cincinnati, OH 45267-0394, Ernest.Weiler@uc.edu)

Measurement of loudness adaptation at 4000 Hz and below depends on the technique used [Weiler *et al.*, J. Acoust. Soc. Am. **101**, 3171(A) (1997); T. Maguire *et al.*, J. Acoust. Soc. Am. **106**, 2207 (1999)]. Further comparison of simple adaptation (SA) to the ipsilateral comparison paradigm (ICP) with repeated measures designs again shows adaptation for both techniques at 8000 Hz but stronger effects for the ICP now. At 6000 Hz only, violations of normal curve parameters were again observed but no SA. Observations at 250 Hz for the ICP and SA will be discussed. The difference in adaptation between the two techniques is extreme in the primary speech frequencies but intermediate at 8000 Hz for our repeated measures designs. Further investigation is planned.

TUESDAY MORNING, 5 JUNE 2001 GRAND BALLROOM, 11:00 A.M. TO 12:05 P.M.

Session 2aPPb

Psychological and Physiological Acoustics and Archives and History: History of Physiological Acoustics in ASA

Peter Dallos, Chair
Frances Searle Building, Northwestern University, 2299 Sheridan Road, Evanston, Illinois 60208

Chair's Introduction—11:00

Invited Paper

11:05

2aPPb1. History of physiological acoustics: A modeling perspective. Murray B. Sachs (Dept. of Biomed. Eng., Johns Hopkins Univ., 720 Rutland Ave., Baltimore, MD 21205)

Since the days of Helmholtz and Ohm in the nineteenth century, physiological acoustics has focused on models of mechanisms underlying human auditory perception. We review the history of three modeling streams: (1) models of cochlear mechanisms, (2) models that predict human psychoacoustic performance on the basis of neural representations of sound, and (3) models for processing of acoustic information by the central nervous system. We trace the evolution from early macromechanical models that were adequate to explain the broad tuning of the basilar membrane as reported by von Bekesy to recent micromechanical models needed to account for sharp basilar-membrane tuning as revealed by more sensitive experimental techniques. We review models of the molecular mechanisms underlying hair cell electromotility and its relation to sharp tuning. Because of the ''dynamic range problem'' early models of frequency and intensity discrimination based on auditory-nerve discharge rates gave way to models based on temporal properties of auditory-nerve firing patterns. Discovery of a population of fibers with broad dynamic ranges revived interest in rate-based models. Studies of information processing in the central auditory system have been shaped by models for the representation of the frequency, intensity, and spatial characteristics of acoustic stimuli.

A NOTE ABOUT THE ASA HISTORY LECTURE SERIES

In 1997, the ASA Committee on Archives and History conceived a plan for a series of invited lectures on each of the technical areas of the Society which would memorialize the significant achievements and milestones of each of its technical committees during the first three quarters of the Society's first century.

With the cooperation of the technical committees, distinguished individuals are selected to review the history of their particular technical specialty and present a lecture which shows how that activity has developed and has contributed to the Society at large and to the broad field of acoustics as well.

The invited lecturers have been asked to prepare a written manuscript of their lectures which will be published in a commemorative book for the 75th Anniversary of the Society to be celebrated in 2004. The Archives and History Committee and the individual technical committees/group welcome comments and suggestions on both the History Lecture Series and on the proposed ASA Diamond Anniversary Book. Volunteers to assist the committees would be most welcome too. Contact Henry Bass, Chair, Committee on Archives and History, pabass@sunset.backbone.olemiss.edu

Session 2aSAa

Structural Acoustics and Vibration: Vibroacoustics and Noise Control

Kenneth A. Cunefare, Chair
School of Mechanical Engineering, Georgia Institute of Technology, Graduate Box 268, Atlanta, Georgia 30332-0405

Contributed Papers

8:00

2aSAa1. Theoretical and experimental modal analysis of brake pads. Mario Triches, Jr., Samir N. Y. Gerges (Mech. Eng. Dept., Federal Univ. of Santa Catarina, Cx.P. 476-Florianopolis-SC, 88040-900, Brazil), and Shahram Tousi (MSC Laminates and Composites, Inc., Elk Grove Village, IL 6007-5995)

Disc brake squeal noise is a complicated dynamic problem that has confronted automobile manufacturers for decades. The reduction of brake squeal noise is an important technological subject in terms of making vehicles quieter. Two main mechanisms are correlated with squeal noise in disc brake systems. The first one is related to the velocity dependency of the friction coefficient. The second one is the modal coupling of the rotor and the pads. In this case, the modes of the pads and rotor are coupled and the system may become unstable. To extract the natural frequencies and the vibration modes of pads is of great importance for the forecast of the dynamic behavior of the complete system. Thus, it can predict the behavior of the pads with the noise generated by the system. In this work, results of experimental modal analysis of the pads are presented and compared with the results obtained through use of the finite-element method (FEM). Different conditions of temperature were applied to the pads and simulated by FEM, verifying the influence of the work temperature in the natural frequencies.

8:15

2aSAa2. Transfer matrix method for mufflers modeling and experimental verification. Fabio A. Thieme (Eberspcher Tuper Sistema de Exhausto Ltda., So Bento de Sul, SC, Brazil), Samir N. Y. Gerges (Federal Univ. of Santa Catarina, Florianopolis, SC, Brazil), and Jos L. B. Coelho (Insituto Superior Tecnico, 1049-001 Lisboa, Portugal)

Mufflers are very important elements in industry for the attenuation of exhaust noise in vehicles and machinery. Recent advances in modeling procedures, for accurate performance prediction, lead to the development of the transfer matrices method for more practical muffler components appearing in commercial designs. Muffler designers are looking for simple, quick modeling tools, especially in the preliminary evaluation stages of a design. Finite element (FEM) and boundary element (BEM) methods, in spite of giving results in a wide range of frequencies, are still expensive, need very skilled operators, and commercial software is still not user friendly. Plane wave modeling, using the transfer matrix form, offers a cheap and quick method for muffler designers to arrive at an initial prototype solution. In this paper, transmission loss calculations are presented for several practical muffler configurations in explicit formulas and comparison with experimental results are presented.

8:30

2aSAa3. Theoretical development and experimental verification of polyvinylidene flouride sensors for measurement of the local volume displacement of beams. Randall Rozema, Brian Zellers, Koorosh Naghshineh (Mech. and Aeronautical Eng. Dept., Western Michigan Univ., Kalamazoo, MI 49008), and Marcellin Zahui (Lake Superior State Univ., Sault Sainte Marie, MI 49783)

One method of reducing the sound radiated from vibrating structures at lower frequencies is to reduce the volume displacement of the vibrating surface (via active control). At these low frequencies, the volume displacement is directly proportional to the sound power emitted from a vibrating surface. To extend the effective frequency range of the active control system, several systems that reduce the volume displacement over localized areas of structural surface were employed. Thus, means of measurement of structural surface volume displacement become important. A traditional approach is to employ multiple point sensors (accelerometers). Recently, a single sensor made of polyvinylidene fluoride (PVDF) was utilized and found to represent an attractive solution. These sensors, which were designed to measure the structural volume displacement over a segment of structural surface, spanned the entire length of the structure. Such arrangement was found to be inefficient. In the work presented here an integrated sensor comprised of a shaped PVDF strip in addition to two point sensors located at the ends of this strip. The entire sensor does not extend beyond the segment of interest, thus, labeled a local sensor. Theoretical development of these sensors, fabrication techniques, and verification of these sensors will be presented.

8:45

2aSAa4. Investigation of local volume displacement sensors for rectangular vibrating plates. Mark Christensen, Ron Hoofnagle, Koorosh Naghshineh (Mech. and Aeronautical Eng. Dept., Western Michigan Univ., Kalamazoo, MI 49008), and Marcellin Zahui (Lake Superior State Univ., Sault Ste. Marie, MI 49783)

Further development and validation of a technique for measurement of local volume displacement is presented. This development supports the implementation of noise control techniques that are based on minimization of local volume displacements, velocities, or accelerations of a vibrating structure. In this work, we present a brief description of the methodology for designing such sensors fabricated using polyvinylidene fluoride (PVDF) film followed by the experimental verification of these sensors for a vibrating clamped–clamped plate comprising one side of an otherwise rigid box. These experimental results were then compared against predicted values.

9:00

2aSAa5. Experimental validation of the state-switched absorber for two-component harmonic forcing. Mark Holdhusen, Kenneth A. Cunefare, and Gregg Larson (The George W. Woodruff School of Mech. Eng., The Georgia Inst. of Technol., Atlanta, GA 30332-0405, gte165r@prism.gatech.edu)

A state-switched vibration absorber (SSA) is a semiactive device that is capable of instantaneously changing its dynamical state by altering its stiffness. However, the SSA can switch between more than one resonance frequency, providing a greater bandwidth as compared to classical tuned vibration absorbers. Modeling predictions indicate that with appropriate logic for switching between states, the SSA is more effective at vibration suppression than classical tuned vibration absorbers, when the system to be controlled is subject to a disturbance with more than one spectral component. This presentation considers the experimental validation of SSA performance for the suppression of vibration on a driven elastically mounted lumped mass system. The SSA considered here is implemented in a dynamical proof-of-concept demonstrator. State switching is effected

through the use of a magneto-rheological fluid clamp to connect or disconnect a coil spring in parallel with another coil spring. The two coil springs, one switchable, the other not, constitute the spring element of an otherwise passive mass-spring-damper vibration absorber. The stiffness state is controlled by actuation of the MR clamp. The experimentally observed performance of the SSA will be presented and compared to modeling predictions. [This work is based upon work supported by, or in part by, the U.S. Army Research Laboratory and the U.S. Army Research Office under Grant No. 38955-EG, Dr. Gary Anderson, Technical Monitor.]

9:15

2aSAa6. A passive means for cancellation of structurally radiated tones. Jeffrey A. Zapfe and Eric E. Ungar (Acentech, Inc., 33 Moulton St., Cambridge, MA 02138-1118)

The mass of a spring–mass system whose spring is driven at a given frequency vibrates in opposite phase to the excitation, provided that the system's natural frequency is lower than the excitation frequency. If this natural frequency also is near the excitation frequency, then the amplitude of the mass can be considerably greater than that of the driving motion. These ideas were applied to construct radiation-canceling devices that act like resiliently supported masses with significant sound radiating areas, so that radiation from these could cancel the sound radiated from considerably larger vibrating surfaces to which these devices may be attached. The sound radiation from such devices and the attendant reduction in the sound radiated from sample structures were analyzed, some experimental prototype devices were built, and their effects were measured. Devices of this type are thought to be useful for reducing the sound radiation from equipment items, such as transformers, which vibrate and radiate noise at a fixed frequency. [Patent pending.]

9:30

2aSAa7. Sound insulation analysis of solid walls using the finite element method. Lutz Ackermann and Heinz Antes (Inst. of Appl. Mech., Tech. Univ. of Braunschweig, Spielmannstr. 11, 38106 Braunschweig, Germany, l.ackermann@tu-bs.de)

In times of increasing noise pollution, the numerical simulation of sound transmission through solid walls, e.g., of masonry, is a challenging building acoustics topic. For an adequate building design tool it is of great importance to take arbitrary geometrical and acoustical boundary conditions as well as the air–structure interaction into account. Here, a model based on the finite element method is presented for the calculation of sound propagation in layered air/wall/fiber systems. The structure-borne sound is, on the one hand, influenced by the bending waves and, on the other hand, by the in-plane waves, if also the flanking transmission is investigated, where the bending waves are modeled by the Mindlin plate theory and the in-plane waves by the dynamic elastic disk equation. The acoustic behavior of air is described by the Helmholtz equation. The fluid–structure interaction is performed by using the principal of virtual work, and, as a final result, a completely coupled methodology is derived which allows the determination of sound fields and of the transmission loss of conventional solid walls. Numerical results are compared with measured values and the influence of various parameters is studied.

TUESDAY MORNING, 5 JUNE 2001 CRYSTAL ROOM, 10:00 A.M. TO 12:00 NOON

Session 2aSAb

Structural Acoustics and Vibration: Analytical and Numerical Methods in Vibrations

Linda P. Franzoni, Chair

Department of Mechanical Engineering and Material Science, Duke University, Box 90300, Durham, North Carolina 27708-0300

Contributed Papers

10:00

2aSAb1. Wave vibration analysis and control of a transversely vibrating Timoshenko beam. Carole Mei (Dept. of Mech. Eng., The Univ. of Michigan–Dearborn, 4901 Evergreen Rd., Dearborn, MI 48128)

Apart from the modal description, vibrations in structures can also be described in terms of waves. In this paper, wave vibration analysis and control of a transversely vibrating Timoshenko beam are studied. The wave reflection and transmission characteristics under general discontinuity and general boundary conditions are presented. The reflection and transmission matrices for incident waves upon such discontinuities and boundaries are derived. Both free and forced responses of the Timoshenko beam are analyzed from wave reflection and transmission standpoint. The Timoshenko and the Euler–Bernoulli beam model are compared. Numerical examples of the free and forced vibrations of both slender and nonslender beams are presented to show the similarities and to signify the differences between the two beam models. Vibration suppression strategies from wave point of view are addressed.

10:15

2aSAb2. A high-frequency theory for thermo-viscoelastic beams. G. Askar Altay (Dept. of Civil Eng., Bogazici Univ., Bebek 80815, Istanbul, Turkey) and M. Cengiz Dokmeci (Istanbul Univ., Istanbul, Turkey)

To predict the macromechanical response of beams of uniform cross section, a system of 1-D equations is consistently formulated in both differential and variational forms within the frame of the 3-D theory of thermo-viscoelasticity. First, the basic equations of linear, nonisothermal 3-D theory of thermo-viscoelasticity are expressed in terms of the Laplace transformed field variables by a recent variational principle [Altay and Dokmeci, Acta Mech. **143**, 91–111 (2000)]. Next, the temperature increment and displacement fields are expressed by the series expansions of the aerial coordinates of beam cross section. Then, a system of higher-order beam equations is derived by means of the variational principle together with the series expansions of field variables. The system of equations governs the extensional, flexural, torsional, as well as coupled vibrations of thermo-viscoelastic beams at both high and low frequencies. Certain cases involving special types of vibrations, geometry, and material properties are recorded. The uniqueness is investigated in solutions of the system of beam equations. [Work supported in part by TUBA.]

10:30

2aSAb3. Special properties of higher-order symmetric Lamb waves on plates. Michael F. Werby (NRL Code 7180, Stennis Space Center, MS 39529 and Dept. of Phys., Catholic Univ. of America, Washington, DC 20064) and H. Überall (Catholic Univ. of America, Washington, DC 20064)

Exact linear elastodynamic theory is used to derive expressions for all of the modes on a two-dimensional unloaded plate in an effort to understand certain features generated from the symmetric modes on plates and shells. In that context we explain the unusual nature of all S_1 modes including the negative slope of the phase velocity with respect to wave number at critical frequency, the amplitude modulated nature of the transient solution in that frequency zone, and the fact that one type of elastic material has an infinite number of such unusual modes (materials with a Poisson ratio of exactly 1/3) while the remaining class only appears to have that property for the S_1 mode. We review the asymptotic nature of all of the plate modes and show that the symmetric Lamb modes always have a plateau region in phase velocity relative to wave number and how many of the properties of the symmetric Lamb modes may be used for classification and detection of shells and plates.

10:45

2aSAb4. Reflection of sh waves from an irregular free-end in a semi-infinite elastic plate. Nahil A. Sobh (Theoret. and Appl. Mech., Univ. of Illinois–Urbana-Champaign, Urbana, IL 61801) and Yagoub N. Al-Nassar (KFUPM, Dhahran 31261, Saudi Arabia)

The problem of anti-plane-wave reflection off an inclined free-end in a semi-infinite elastic plate is investigated. An artificial internal boundary is constructed which divides the plate into two regions. One region contains the irregular free-end while the other contains the rest of the semi-infinite plate. Each admits a separable solution of the wave equation and satisfies all its physical boundary conditions. The full wave solution is finally constructed by matching displacement and stress continuities at the internal boundary. A number of numerical studies illustrate the variation of the reflected modal energy as a function of the free-end inclination and the incident modal energy.

11:00

2aSAb5. An energy finite element formulation for high-frequency vibration analysis of plate structures subjected to heavy fluid loading. Weiguo Zhang and Nickolas Vlahopoulos (Dept. of Naval Architecture and Marine Eng., Univ. of Michigan, 2600 Draper Rd., Ann Arbor, MI 48109-2145)

An energy finite element formulation for computing the high-frequency behavior of plate structures in contact with a dense fluid is presented. The heavy fluid loading effect is incorporated in the derivation of the governing differential equations and in the computation of the power transfer coefficients between plate members. The validity of the new formulation is demonstrated through comparison to an analytical solution and to results computed by conventional finite element models comprised by a large number of elements. The influence of the heavy fluid on the dynamic response of two marine structures is identified by comparing results computed with and without the effect of fluid loading. The impact of the heavy fluid loading on the computation of the power transfer coefficients is also identified. [Research supported by the Michigan Seagrant.]

11:15

2aSAb6. 2.5-D FEM for cochlear-based transducers: Theory and experiment. John M. Dodson and Karl Grosh (Mech. Eng. Dept., Univ. of Michigan, 3124 G. G. Brown Lab., 2350 Hayward St., Ann Arbor, MI 48109-2125, grosh@engin.umich.edu)

The mammalian cochlea is an extraordinary organ exhibiting outstanding sensitivity and frequency discrimination capabilities. This tiny organ senses acoustic input via a coupled pressure-displacement wave along a pair of ducts separated by a partition known as the basilar membrane (modeled as a variable impedance plate) which exhibits spatial-frequency dependent resonances. The cochlea provides a template for designing devices which utilize this unique mode of transduction and holds potential for a new class of transducer. A reduced-order hybrid analytical/finite-element method (FEM) is used to allow computation of this coupled fluid-orthotropic variable width plate interaction problem. In this method, a 3-D FEM problem is reduced to 2-D discretization by utilization of a limited set of analytical pressure/structural modes orthogonal to the direction of wave propagation. Simple and complex assumed structural mode shapes were investigated to predict the plate response. Comparison to experiments revealed that the deflection of plate supporting structures is important to achieve a good correlation between theory and experiment. The validation program was conducted for isotropic and orthotropic plates mounted on a fluid-filled duct excited by a shaker-driven piston. Design studies for a cochlear-based transducer will be presented.

11:30

2aSAb7. Dynamic response of *in vacuo* prolate spheroidal shells excited by surface force fields. Sabih I. Hayek (Dept. of Eng. Sci. and Mech., Penn State Univ., University Park, PA 16802-6812, sihesm@engr.psu.edu) and Jeffrey E. Boisvert (Naval Undersea Warfare Ctr. Div. Newport, Newport, RI 02841)

The equations of motion for nonaxisymmetric vibration of prolate spheroidal shells of constant thickness were derived using Hamilton's principle. The thin shell theory used in this derivation includes three displacements and two changes of curvature. The effects of membrane, bending, shear deformations, and rotatory inertias are included in this theory. The resulting five partial differential equations are self-adjoint and positive definite. The shells are excited by axisymmetric line forces. The axisymmetric modal solutions are expanded in an infinite series of comparison functions. These include associated Legendre functions in terms of the prolate spheroidal angular coordinate. Numerical results for the frequency response were obtained for several shell thickness-to-length ratios ranging from 0.005 to 0.1, and for various diameter-to-length ratios, including the limiting case of a spherical shell. Sample mode shapes were obtained at selected resonant frequencies. [Work supported by ONR and the Navy/ASEE Summer Faculty Program.]

11:45

2aSAb8. Recursive Green function technique applied to acoustic wave propagation. Mauro S. Ferreira, Gerrit E. W. Bauer (Dept. of Appl. Phys., Delft Univ. of Technol., Lorentzweg 1, 2628 CJ Delft, The Netherlands), and Kees Wapenaar (Delft Univ. of Technol., 2628 RX Delft, The Netherlands)

Guided by similarities between electronic and classical waves, a numerical code based on a formalism proven to be very effective in condensed matter physics has been developed, aiming to describe the propagation of elastic waves in stratified media (e.g., seismic signals). This so-called recursive Green functions technique is frequently used to describe electronic conductance in mesoscopic systems. It follows a space discretization of the elastic wave equation in frequency domain, leading to a direct correspondence with electronic waves traveling across atomic lattice sites. An inverse Fourier transform simulates the measured acoustic response in time domain. The method is numerically stable and computationally efficient. Moreover, the main advantage of this technique is the possibility of accounting for lateral inhomogeneities in the acoustic potentials, thereby allowing the treatment of interface roughness between layers. Without those lateral variations, wave propagation across perfectly layered media with planar sources are equivalent to 1-D environments, and therefore should present localization effects. Inclusion of those variations breaks the symmetry and turns the wave propagation into a truly 2-D or 3-D problem, where localization is weaker or may not occur. Numerical techniques without lateral inhomogeneities may inadvertently overestimate localization effects.

Session 2aSC

Speech Communication and Psychological and Physiological Acoustics: Special Session Honoring Harry Levitt

Arlene C. Neuman, Chair
Center for Research in Speech and Hearing, Graduate Center, City University of New York, 365 Fifth Avenue, New York, New York 10016

Chair's Introduction—8:00

Invited Papers

8:05

2aSC1. Adaptive procedures in psychoacoustics and speech research. Marjorie Leek (Army Audiol. & Speech Ctr., Walter Reed Army Medical Ctr., Washington, DC 20307-5001, marjorie.leek@na.amedd.army.mil)

Over the past 30 years, adaptive procedures have become ubiquitous in psychoacoustic and speech perception research. The increasing use of these procedures may be traced in part to Harry Levitt's paper describing the flexibility of up–down testing methods to determine the point on a psychometric function associated with a given level of performance [H. Levitt, J. Acoust. Soc. Am. **57**, 55–62 (1971)]. Most psychoacousticians have, at one time or another, consulted this article to inform their experimental procedures. However, Harry's contributions to psychometric testing do not begin or end with the publication of this seminal article. Here I will discuss the far-reaching impact of the transformed procedures that formed the basis of the 1971 article, point out some caveats to be considered when using these methods, and review other developmental work that Harry has accomplished to improve psychometric testing, including adaptive methods of fitting hearing aids and of testing speech recognition. [Work supported by NIH (DC00626).]

8:30

2aSC2. Harry Levitt legacies: Studies of speech perception and intelligibility in children with cochlear implants. Emily A. Tobey (Callier Adv. Hearing Res. Ctr., 1966 Inwood Rd., Dallas, TX 75235, etobey@utdallas.edu) and Ann Geers (Central Inst. for the Deaf, St. Louis, MO)

Harry Levitt has contributed a substantial body of knowledge regarding the speech perception and production skills of hearing-impaired children. In this paper, Dr. Levitt will be honored by describing the speech perception, speech intelligibility, and acoustic characteristics of 136 cochlear implanted children between the ages of 8 and 9 who have at least 4 years experience with a Nucleus multichannel cochlear implant. Half of the children used auditory-oral modes of communication and the other half used total communication. The questions addressed are as follows. (a) Do children implanted before 5 years of age develop intelligible speech? (b) Are speech perception skills reflected in the acoustic measurements of speech in children with cochlear implants? (c) What factors are associated with intelligible speech in children with cochlear implants? The presentation will describe the population demographics, speech perception measures (including open and closed set speech perception performance), and measures of speech production (including speech intelligibility, communication breakdowns, parental rating scales, and acoustic characteristics). Factors contributing to high levels of speech perception and intelligibility will be described and contrasted. Implications for the development of intelligible speech will be discussed, particularly as it relates to mode of communication. [Work supported by NIH-NIDCD.]

8:55

2aSC3. Reordering disordered speech: The search for speech prostheses. H. Timothy Bunnell (duPont Hospital for Children and Univ. of Delaware, 1600 Rockland Rd., Wilmington, DE 19803)

The development over the last quarter century of computer-based speech processing has allowed investigators to (a) modify disordered speech as a means of gaining a better understanding of its nature, and (b) look toward the possibility of altering disordered speech to enhance its perceived naturalness or its intelligibility. Harry Levitt and his colleagues [H. Levitt, IEEE Trans. Audio. Electroacoust. **21**, 269–273 (1973); M. J. Osberger and H. Levitt, J. Acoust. Soc. Am. **66**, 1316–1324 (1979)] were among the first to examine the possibility of enhancing disordered speech via computer speech processing. Since then, a number of other investigators have taken up this line of research, employing a variety of speech processing techniques to examine issues raised by Levitt (1973), particularly regarding the relative importance of temporal versus spectral distortions in determining the intelligibility or naturalness of disordered speech. In this work, I will trace some of the history of this line of research, including work that Grace Yeni-Komshian and the author have conducted [J. Acoust. Soc. Am. **104**, 637–647 (1998)], and subsequent related studies.

9:20

2aSC4. Harry Levitt and hearing aids. Jont B. Allen (AT&T Labs, Rm. E161, Florham Park, NJ 07974)

Levitt has always been interested in technology transfer. In 1980 he became a proponent of digital hearing aids and fitting systems. I first got to know Levitt during the 1983 AT&T venture into a multiband compression hearing aid, proposed to Waldhauer by Villchur, based in the ideas of Steinberg and Gardner (1938). One of the first things we did was to visit Levitt and pepper him with questions. He saw it as a huge opportunity, but it was difficult because he was consulting for Nicolet. Eventually AT&T decided to return to its ''core'' business, and the venture was sold to Resound Corp. At CUNY we also developed the loudness growth in 1/2 octave bands (LGOB) test, which was popular at one time. Soon after, Levitt convinced me to port our research distortion product analyzer, called CUB^{e}DIS, to the PC platform. It was then further developed by Jeng (Levitt's student), and is presently Starkey's DPOAE/TOAE DP-2000 infant hearing screener. All this technology transfer would not have taken place if it were not for Harry Levitt—so for that we must graciously thank him.

9:45

2aSC5. Acoustic and psychoacoustic benefits of adaptive compression thresholds in hearing aid amplifiers that mimic cochlear function. Julius L. Goldstein (Washington Univ., Campus Box 1127, St. Louis, MO 63130), Michael Valente (Washington Univ. Med. School, St. Louis, MO 63110), and Roger D. Chamberlain (BECS Technol., Inc., St. Louis, MO 63132)

Automatic gain control of linear amplifiers dominates advanced hearing aid design, and has been extensively studied [Levitt *et al.*, IEEE (1980); Dillon, Ear Hear. **17**, 287–307 (1996)]. The normal cochlea uses essentially nonlinear, rapidly compressive amplification under efferent control [Kiang *et al.*, Hear. Res. **22**, 171–182 (1986); Robles *et al.*, J. Acoust. Soc. Am. **80**, 1364–1374 (1986)], whose salient characteristics have been modeled [Goldstein, Hear. Res. **49**, 39–60 (1990)] and are presently adopted for multichannel hearing aids. Signal transduction in each channel is linear at low and high sound pressure levels (SPL), and smoothly joined by a compressive range. The transition from linear to compressive response at low SPL is controlled by an adaptive compression threshold, whose quiescent level provides the gain needed for weak sounds. For sustained sounds at higher SPL the compression threshold adapts upwards, providing effectively linear response and better acoustic quality for speech in noise. Pilot psychoacoustic experiments with a design simulation demonstrated that both normal and impaired hearing subjects perceive the improved quality, and comprehend speech in noise at least as well as with advanced hearing aids [Valente *et al.*, J. Am. Acad. Audiol. **9**, 342–360 (1998)]. [Work supported by NIDCD.]

10:10–10:25 Break

10:25

2aSC6. Feedback cancellation in hearing aids. James M. Kates (AudioLogic/Cirrus Logic, 2465 Central Ave., Ste. 100, Boulder, CO 80301, jkates@audiologic.cirrus.com)

Harry Levitt was an early advocate of digital signal processing for hearing aids. His interest included not only algorithms for compression and speech enhancement, but also more practical issues such as obtaining the desired hearing aid versus frequency response at the ear drum. Acoustic feedback is one problem that can limit the maximum gain possible in a hearing aid. Feedback cancellation, in which the acoustic feedback signal is estimated and subtracted from the microphone input, allows for greater hearing-aid signal amplification, and feedback cancellation was included in the work that Harry supported in his research group. In this presentation, the effects of feedback on the hearing-aid response will be reviewed. Digital adaptive filter techniques for feedback cancellation will then be presented, along with measurements indicating the limitations of feedback cancellation.

10:50

2aSC7. Speech intelligibility and listener preferences for amplitude-compressed speech: Speech-based STI predictions and measurements on listeners with simulated hearing loss. Peninah F. Rosengard, Louis D. Braida (Res. Lab of Electron., MIT, Cambridge, MA 02139), and Karen L. Payton (Univ. of Massachusetts–Dartmouth, North Dartmouth, MA 02747)

Relations between objective intelligibility scores, subjective pleasantness ratings, and estimates of the STI for speech processed by multi-band amplitude compression systems were studied in normal-hearing listeners with simulated hearing loss. STI estimates were based on modulation spectrum changes in the processed speech signals [Payton and Braida, J. Acoust. Soc. Am. **106**, 3637–3648 (1999)]. Linear amplification and two syllabic compression conditions were tested with and without two backgrounds: Speech-spectrum noise and restaurant babble. Signals were compressed independently in four nonoverlapping frequency bands with compression ratios of two and three, and attack and release times of 20 and 200 ms, respectively. The NAL-R formula determined output frequency-gain characteristics. Flat, 50 dB, sensorineural hearing losses were simulated in normal-hearing listeners via multiband expansion [Duchnowski and Zurek, J. Acoust. Soc. Am. **98**, 3170–3181 (1995)]. Speech intelligibility and pleasantness ratings were obtained. All conditions were also evaluated using the modified STI. The STI predicted speech in restaurant babble would be more intelligible than speech in noise for each compression condition. The data reflected this for the two compression conditions. The STI predictions were also consistent with pleasantness ratings: Linear amplification was the most pleasant compression condition. [This work supported by NIDCD.]

11:15

2aSC8. Recent advances in audio and video text-to-speech synthesis. Ann K. Syrdal (AT&T Labs Res., 180 Park Ave., Florham Park, NJ 07932), Joern Ostermann, Eric Cosatto, and Hans Peter Graf (AT&T Labs Res., Middletown, NJ 07748)

Harry Levitt characteristically combines a statistical or data-driven engineering approach with extensive scientific knowledge when tackling fundamental research questions (such as how to measure co-articulation) and practical problems (such as evaluating intelligibility). Some of Harry's expansive research interests and contributions have focused on audio and video text-to-speech (TTS) synthesis. Some major recent improvements in these technologies that have resulted from applying similar approaches are discussed. The unit selection method of synthesis described differs from both the earlier techniques of formant synthesis and of concatenative synthesis from a diphone inventory. In unit selection synthesis, speech or video units are selected from among multiple candidates in a large database by optimizing a cost function. The cost functions used for unit selection are estimates of perceptual distances. The improvement in naturalness for both audio and visual TTS achieved through the unit selection technique is demonstrated, and some of its current limitations and future work needed in this area is discussed.

Contributed Paper

11:40

2aSC9. Adaptation algorithms for noise reduction in hearing aids based on auditory models. Birger Kollmeier, Jürgen Tchorz, Thomas Wittkop, and Volker Hohmann (Medizinische Physik, Universität Oldenburg, D-26111 Oldenburg, Germany, birger.kollmeier@uni-oldenburg.de)

Noise reduction algorithms for hearing aids usually employ certain assumptions about the ambient noise and the target (speech) signal that are only met in certain acoustical situations. In order to reduce unwanted artifacts whenever the acoustical real-life situation does not meet these assumptions, two approaches are introduced and tested that classify the acoustic environment. Both approaches are motivated by auditory models and are used to control the parameters of noise-reduction schemes. The first approach uses the amplitude modulation spectrogram (AMS) in combination with a neural net to estimate the speech-to-noise ratio either on the broadband signal or within narrow bands. It can be used for single-channel (i.e., monaural) noise reduction even for fluctuating background noise and unfavorable signal-to-noise ratios. The second approach derives the ''degree of interaural coherence'' from the input signals to both ears that decreases with increasing reverberation and increasing number of active sound sources. It is combined with several binaural noise reduction techniques, i.e., directional filtering, dereverberation, and adaptive beamforming. The advantage of using these approaches will be presented and discussed on the basis of speech intelligibility and subjective quality assessment data.

TUESDAY AFTERNOON, 5 JUNE 2001 RED LACQUER ROOM, 1:00 TO 5:00 P.M.

Session 2pAA

Architectural Acoustics: Halls for Music Performance . . . Another Two Decades of Experience 1982–2002

Christopher A. Storch, Cochair
Artec Consultants, Inc., 114 West 26th Street, 9th Floor, New York, New York 10001

Timothy J. Foulkes, Cochair
Cavanaugh Tocci Associates, Inc., 327F Boston Post Road, Sudbury, Massachusetts 01776

Ian B. Hoffman, Cochair
The Talaske Group, Inc., 105 North Oak Park Avenue, Oak Park, Illinois 60301

Posters from Session 2aAAa will remain on display until 5 p.m.

Session 2pAB

Animal Bioacoustics and Psychological and Physiological Acoustics: Comparative Aspects of Auditory System Development

Andrea M. Simmons, Chair
Department of Psychology and Neuroscience, Brown University, Box 1853, Providence, Rhode Island 02912

Chair's Introduction—1:15

Invited Papers

1:20

2pAB1. Confocal imaging of sensory organ formation during *Xenopus* inner ear development. Elba E. Serrano and Quincy A. Quick (Biol. Dept., New Mexico State Univ., Las Cruces, NM 88003)

During *Xenopus* larval development, the otic vesicle gradually gives rise to several auditory and vestibular sensory organs that are innervated by the eighth cranial nerve. Confocal imaging methods were used to examine the appearance of stereociliary bundles in the various sensory epithelia during *Xenopus* development (stages 28–50), and to gather comparative developmental data that show how the inner ear gradually forms distinct auditory and vestibular compartments discernable by larval stage 50. A BioRad MRC-1024 confocal microscope was used to collect digitized images from inner ears labeled with Alexa 488 phalloidin to detect the actin cytoskeleton, and propidium iodide to highlight DNA in cell nuclei. Confocal images from whole inner ears clearly show the developmental emergence of uniquely organized sensory epithelia populated by hair cell bundles of varying morphology, as well as the position of the sensory ganglia relative to the endorgans. Images from sectioned tissue provide additional information about cell morphology and the arrangement of the sensory epithelium in relation to neural innervation. Taken together, the image data illustrate the complex structural changes that underlie inner ear organogenesis and morphogenesis. [Work supported by grants to EES (NIH NIGMS, NASA) and awards to QQ (NASA NMSGC Fellowship, NIGMS RISE).]

1:40

2pAB2. Developmental changes in GABA expression across metamorphosis. Andrea M. Simmons, Seth S. Horowitz, and Judith A. Chapman (Dept. of Psych. and Neurosci., Brown Univ., Providence, RI 02912, andrea_simmons@brown.edu)

Gamma-aminobutyric acid (GABA) expression in the developing brain has been implicated in factors such as process outgrowth, cell proliferation, cell migration, and synaptogenesis. Immunofluorescent staining techniques were used to characterize changes in GABA distribution in medullary and cerebellar regions of the bullfrog's brainstem across metamorphosis. Auditory and vestibular nuclei show profound developmental changes in distribution; the cerebellum and cerebellar nuclei show little change, being strongly stained at all stages examined. In early stage tadpoles, the dorsolateral nucleus (DLN), vestibular nucleus (VN), and anterior lateral line (LLa) nucleus show widespread punctate and diffuse label, but few discrete GABAergic cells. Little or no staining is observed in the superior olivary nucleus (SON). This basic pattern persists until just before metamorphic climax stages, when discrete GABAergic cells first begin to appear in the SON; more discrete cell staining is also visible in the DLN and VN. At metamorphic climax, the DLN, VN, and SON show clear populations of GABAergic cells. In postmetamorphic froglets and adults, GABAergic cells and GABA-labeled fibers exist in sharply delineated regions of these nuclei, with limited diffuse/puncta label. These changes are consistent with the idea that GABA plays important roles in regions of the brainstem undergoing developmental transformation.

2:00

2pAB3. Experience and auditory brainstem development. Edwin W. Rubel (VMB Hearing Res. Ctr., Box 357923, Univ. of Washington, Seattle, WA 98195, rubel@u.washington.edu)

Understanding the cellular basis of experiential influences on auditory system development involves characterizing (1) the intercellular signaling molecule(s) responsible for activity-based long-term changes in targeted neurons; (2) the cascade of cellular events which alters the phenotype of the targeted neurons; and (3) the biological basis of developmental periods of enhanced brain plasticity, critical periods. I will summarize a series of experiments examining neuronal integrity in the cochlear nucleus of birds and mammals following manipulations of eighth-nerve activity. Comparisons of conductive hearing loss and TTX blockade of eighth-nerve activity along with *in vitro* experiments indicate that glutamate release and activation of group 1 metabotropic glutamate receptors (mGluRs) are essential for normal survival of cochlear nucleus neurons. The cascade of intracellular events determining survival or death of

cochlear nucleus neurons involves regulation of intracellular free calcium by one or more mGluRs. Finally, new studies using transgenic mice are beginning to reveal biological mechanisms responsible for developmental changes in cellular responses to afferent deprivation. I will present recent data showing that manipulations of bcl-2 gene expression dramatically alter the critical period of susceptibility to afferent deprivation. [Research supported by grants from NIH/NIDCD.]

2:20

2pAB4. Thyroid hormone influences on the development of auditory function in mammals. Edward J. Walsh and JoAnn McGee (Boys Town Natl. Res. Hospital, 555 N. 30 St., Omaha, NE 68131)

The essential role of the thyroid hormone in development is one of the most clearly established principles of biology. Although its importance as a factor affecting the maturation of the inner ear is unquestionable, the precise mode of its actions and the identity of its intracochlear targets remain issues of conjecture. Although both alpha1 and beta forms of the thyroid hormone receptor (TR) are expressed in the developing cochlea of mammals, permanent auditory deficits occur only when TR-beta1, which is distributed widely in the brain and inner ear, is defective. The presence of multiple forms of thyroid hormone receptors with different functions suggests that the mode of thyroxins action in the inner ear is complex. When thyroxin is unavailable during development, as in Tshr (hyt) mutant mice or animals rendered hypothyroid by treatment with goitrogens, a distinctive constellation of functional deficits are observed, suggesting that the primary, enduring locus of pathology lies within the domain of the cochlear amplifier. We will discuss the physiological basis underlying this position and propose a mechanism that accounts for the phenomenon, as well as the wide range of anatomical abnormalities that have been observed in the cochleae of hypothyroid mammals. [Work supported by NIDCD grant DC00982.]

2:40

2pAB5. Development of perceptual dimensions lags the ability to perceive details. Lincoln Gray, Joshua Breier (Univ. of Texas, Houston Med. School, 6431 Fannin St., Houston, TX 77030), Randy Diehl (Univ. of Texas, Houston, TX 77030), Peggy Haardt, and Christine Turley (Univ. of Texas Medical Branch, Houston, TX 77030)

Studies with humans and animals show that the ability to discriminate fine changes along an auditory continuum develops before the ability to appreciate the continuum. The visual analogy would be seeing details before the ''big picture.'' This trend is discussed with data from normal newborn chickens responding to changes in frequency and azimuth, from patients after corrected congenital hearing loss, and from children and chickens exposed to lead. In each instance, subjects respond to small changes before they respond consistently to large changes. Multidimensional scaling can reveal the ''attribute'' or perceived dimension that distinguishes a set of stimuli. The attribute is compared to the difference limen. These data suggest that discriminations can emerge before the appropriate perception of a sensory attribute. Perceptual development may not always involve a process of increasingly fine discriminations along continua that are established at an early age, but instead may involve the ability to process global pattern. Some theories suggest that perceptions ''sharpen'' over development. Sharpening does not seem to occur in situations we have studied: over normal development, after early deprivation, or after early lead poisoning. If confirmed in other situations, this could suggest novel approaches to ameliorate some early perceptual processing deficits.

3:00

2pAB6. Neural representation and perception of speech sounds in children with learning problems. Nina Kraus (Depts. Commun. Sci., Neurobiology, Otolaryngol., Northwestern Univ., 2299 N. Campus Dr., Evanston, IL 60201) and Cynthia King (Northwestern Univ.)

An inability to process auditory information, especially speech, characterizes many children with learning and attention problems. These speech–sound perception problems may arise, at least in some cases, from faulty representation of the speech signal in central auditory centers. Our working hypothesis is that acoustic–phonetic disorders are abnormalities in preconscious neurophysiologic representation of sound structure by central auditory pathway neurons and is reflected by subcortical and cortical aggregate neural responses. Brainstem and cortical potentials (ABR, FFR, cortical P1/N1/N2, MMN) reflect activity from different anatomic sources and represent different aspects of auditory function. An obvious characteristic distinguishing these responses is their development time course. Results indicate that while normal and learning impaired children do not differ in auditory system development per se, they develop different listening strategies which affect the neural representation of sound structure along the auditory pathway. [Work supported by NIH RO1 DC 01510.]

Contributed Papers

3:20

2pAB7. The auditory system as a window onto the processes of prenatal development. Dennis McFadden (Dept. of Psych. and Inst. for Neurosci., Mezes Hall 330, Univ. of Texas, Austin, TX 78712)

Otoacoustic emissions (OAEs), auditory evoked potentials (AEPs), and hearing sensitivity all exhibit sex differences of several types in humans, and these sex differences are highly similar in newborns and adults. The strong implication is that these differences originate during prenatal development and persist through life. The most parsimonious explanation is that these sex differences are produced by androgenic processes operating on the ear and brain of the developing fetus. Certain special populations of humans have been shown to have OAEs and/or AEPs that differ from the sex-typical values, and again the most parsimonious explanation is that the degree of androgen exposure is the basis for these differences. Certain nonauditory measures from these special populations are in accord with this interpretation. The data from special populations will be summarized and a number of alternative explanations considered. Implications for research on the development of hearing will be discussed. [Work supported by NIDCD.]

3:35

2pAB8. Pattern adaptation in midbrain and medullar auditory units of the frog. Nikolay G. Bibikov (N. N. Andreyev Acoust. Inst., Shvernik St. 4, Moscow 117036, Russia)

The responses of medullar and midbrain auditory cells to prolonged amplitude-modulated tone stimulation were recorded extracellularly in the dorsal medullar nucleus, superior olive and torus semicircularis of curarized grass frogs (*Rana temporaria*). The majority of the cells with tonic response to tone bursts showed a significant adaptation in their firing rate within first 20 to 30 s and then stabilized gradually. The temporal course of the rate adaptation was approximated by a single or double exponent plus a steady component corresponded to sustained firing. The dependence of the time constant of the rate decrease and the sustained firing value upon mean carrier intensity, modulation depth and modulation frequency have been studied for 128 medullar and 105 midbrain units. Generally, adaptation was stronger for midbrain versus medullar units. In each auditory nuclei the adaptation decreased with the increase in carrier intensity and modulation depth. The weakest adaptation (the highest value of sustained rate) was usually observed when a low-frequency noise was used as a modulating waveform. The comparison to the mammalian central auditory units is discussed.

3:50

2pAB9. The fish swimbladder as an acoustic waveguide. David T. I. Francis (School of Mathematics and Statistics, Univ. of Birmingham, Edgbaston, Birmingham B15 2TT, UK) and Kenneth G. Foote (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

The possible role of the fish swimbladder as a sound-sensing organ is investigated using numerical models based on the boundary-element method. A number of gadoid specimens are considered. For each specimen, wire-frame models of the swimbladder surface are derived from morphometric data. The response of each swimbladder to a continuous incident wave is calculated at discrete frequencies up to 50 kHz. Pressure distributions within the swimbladders and velocity distributions on the surfaces are presented. Results reveal high-pressure and high-velocity regions around the anterior horns of the swimbladders. [This work originated in European Community RTD—Contract No. MAS3-CT95-0031 (BASS).]

TUESDAY AFTERNOON, 5 JUNE 2001 SALONS III/IV, 1:00 TO 5:30 P.M.

Session 2pBB

Biomedical Ultrasound/Bioresponse to Vibration and Signal Processing in Acoustics: Novel Imaging Techniques in Biomedical Ultrasound

Christy K. Holland, Chair
Department of Radiology, University of Cincinnati, 234 Goodman Street, Cincinnati, Ohio 45219-2316

Chair's Introduction—1:00

Invited Papers

1:05

2pBB1. 3-D compound imaging with refraction and motion correction. Paul L. Carson, Jochen F. Kruecker, Gerald L. LeCarpentier, Charles R. Meyer (Dept. of Radiol., Univ. of Michigan, Kresge III, R3315, Ann Arbor, MI 48109-0553), and J. Brian Fowlkes (Univ. of Michigan, Ann Arbor, MI 48109-0553)

The early goal of multiview imaging was delineation of quasi-specular image boundaries. Medical compound imaging was partially replaced because of resolution loss by refraction, misregistration, and tissue motion, and completely replaced due to incompatibility with real-time imaging. Three potential improvements in: boundary delineation, volumetric contrast-to-noise ratio, and Doppler anisotropy reduction (BC&D) have remained an attraction of compounding. Now marketed, in-plane compounding in real-time and in extended field of view imaging offer the potential for in-plane, image-based reregistration of multiple views. Image volume-based registration (IVBaR) of separate angular views in 3-D data sets allows full 3-D displacement corrections. Compounding out of the image plane (elevational compounding) increases the number of independent views, increasing potential yield from improvements BC&D. Our work has been in elevational compounding with conventional linear arrays, in contact with, and partially moving the proximal tissues. IVBaR using the mutual information metric and thin plate spline interpolation of warped volumes has produced exceptional reclamation of spatial resolution along with the expected improvements BC&D. Full-volume, warped registration, and a faster, subvolume registration have been compared *in vivo* and with realistic beam aberrations in test objects. [Work supported in part by PHS Grant No. R01HL54201 from the NHLBI.]

1:25

2pBB2. Artifact reduction in medical ultrasound. Gary A. Schwartz (ATL Ultrasound, P.O. Box 3003, Bothell, WA 98042, gary.schwartz@philips.com)

Interpretation of medical ultrasound images is confounded by the presence of significant acoustic imaging artifacts. These artifacts result from diffraction, propagation, and scattering effects and regional variations in these effects. Commercial diagnostic imaging systems have taken various design approaches to mitigate the artifacts. Anatomic features widely vary in size relative to the imaging wavelengths, resulting in angular scattering variations and speckle-limited resolution. Tissue attenuation varies greatly resulting in shadowing and enhancement. This review will look at the origins of certain acoustic imaging artifacts and discuss methods that have been applied to address their contribution to image quality. Included in the discussion will be acoustic methods (such as spatial compounding), and signal processing methods (such as attenuation compensation and frequency compounding). Design approaches and clinical image examples will be presented.

1:45

2pBB3. Higher order nonlinear ultrasonic imaging. Bruno Haider (GE Corporate R&D, KW C1315, One Research Circle, Niskayuna, NY 12309) and Richard Y. Chiao (GE Medical Systems, Milwaukee, WI 53219)

The processing of second harmonic echoes from both biological tissue and contrast agents has generated new diagnostic methods in medical ultrasound. The work presented here demonstrates the extraction of higher order nonlinearities. The underlying idea is to model the nonlinear wave propagation or reflection from a contrast bubble by a polynomial expansion of some basis waveform. When this model is excited by a number of transmit pulses which only differ in their amplitude and phase then the coefficients of this polynomial model can be extracted through least squares inversion. The coefficients correspond to the individual nonlinear components. An important feature of the method is the evaluation of nonlinear components whose spectra are folded back into the transmission band. All odd order nonlinearities can create such echo components. The reception of these components eliminates the high bandwidth requirements encountered in second harmonic imaging. Higher-order even harmonics may also be detected by taking advantage of the harmonic fold-back process. Folded frequency components will be centered around DC and at two times the transmit frequency ($2\,f0$). This still requires a bandwidth sufficient to detect signals at $2\,f0$ but eliminates the reception at higher multiples of $f0$.

2:05

2pBB4. B-mode blood flow (B-Flow) imaging. Richard Y. Chiao (GE Medical Systems, P.O. Box 414, Milwaukee, WI 53201)

B-Flow is a new technique that extends the resolution, frame rate, and dynamic range of B-mode to simultaneously image blood flow and tissue. B-Flow relies on coded excitation to boost weak signals from blood scatterers and on tissue equalization to simultaneously display flowing blood and tissue without threshold decision and overlay. Various classes of codes such as Barker and Golay may be used. Clinical B-Flow cineloops demonstrate $3\times$ resolution and frame rate improvement over color flow, which, together with over 60 dB of display dynamic range, is able to image hemodynamics and vessel walls with unprecedented clarity.

Contributed Papers

2:25

2pBB5. A two-dimensional amplitude-steered array for real-time volumetric imaging. Catherine H. Frazier (Johns Hopkins Appl. Phys. Lab., 11100 Johns Hopkins Rd., Laurel, MD 20723), W. Jack Hughes (Penn State Univ., State College, PA 16801), and William D. O'Brien, Jr. (Univ. of Illinois, Urbana, IL 61801)

Real-time three-dimensional acoustic imaging is difficult in water or tissue because of the medium's slow sound speed. Conventional pulse-echo data acquisition, which uses at least one transmit pulse per line in the image, does not allow for the real-time update of a volume of data at practical ranges. Recently, we presented a linear amplitude-steered array [J. Acoust. Soc. Am. **107**, 2430 (2000)], based on an earlier concept [J. Acoust. Soc. Am. **59**, 1040 (1976)], that allows the collection of a plane of data with a single transmit pulse by spatially separating frequencies in the lateral direction. The linear-phased amplitude-steered array uses frequency separation to determine vertical position and conventional beamforming to determine horizontal location. To image a volume of interest with a single transmit pulse, the received signal must contain information to give vertical, horizontal, and range position of the target. The target range is obtained from the time elapsed until the reflected signal is received. The vertical position information is determined by the returned signal's frequency. The target's horizontal position is found by using conventional, linear phased array processing. In this study, we describe the volumetric imaging system, giving the two-dimensional array design, and describing data acquisition and image processing strategies. [Work supported by DARPA's Sonoelectronics Program.]

2:40

2pBB6. Ultrasound image based on ultrasound characterization of tissue microstructure of spontaneous rat mammary tumors. Michael L. Oelze, William D. O'Brien, Jr. (Dept. of Elec. and Computer Eng., Univ. of Illinois, 405 N. Mathews, Urbana, IL 61801, oelze@brl.uiuc.edu), and James F. Zachary (Univ. of Illinois, Urbana, IL 61802)

Experimental results of acoustic backscatter from spontaneous mammary tumors in rats are obtained over the frequency range of 1 to 15 MHz. The power spectrums (dB) from the experimental results are compared to theory in order to obtain information about tissue microstructure. The theoretical power spectrum assumes an isotropic 3-D Gaussian function as the spatial autocorrelation function used to describe the distribution of scatterers in the tissues. The theoretical power spectrum can be shown for

small scatterers (diameter less than 0.2 mm) to be proportional to the Rayleigh spectrum times a second-degree polynomial in frequency. Independent values of the average scatterer diameter and scatterer concentration are obtained from the coefficients of the second-degree polynomial. Scatterer sizes and concentrations measured by ultrasound backscatter are compared to histological data. B-mode images are made of the tumors and surrounding tissues with superimposed regions-of-interest (ROIs) parameterized by the scatterer sizes and concentrations. [Work supported by NIH CA 079179.]

2:55

2pBB7. Spatial sampling resolution study on BAI-mode imaging for defect detection. Xiangtao Yin (Dept. of Elec. and Computer Eng., Univ. of Illinois, 405 N. Mathews, Urbana, IL 61801, xyin@uiuc.edu), Osama Nayfeh, Scott A. Morris, and William D. O'Brien, Jr. (Univ. of Illinois, Urbana, IL 61801)

The principal concern with regard to defect detection in hermetically sealed flexible food packages is the safety and shelf life of the food. The spatial sampling issue of the pulse echo backscattered amplitude integral (BAI) mode imaging technique for defect detection is investigated. In our previous spatial sampling study, it has been shown that for channel defects, the contrast to noise ratio (CNR) degrades as a function of scanning step size on each dimension of the image. To further understand spatial sampling resolution of this method, BAI imaging technique is applied to rectilinear grided dot samples with different grid sizes (distance between adjacent dots). Data is collected with the transducer scanned in a zigzag raster pattern. Quantitatively, the CNR and ΔBAI values are assessed to evaluate the image quality versus the changing spatial interval between dots. At a given operating frequency and a fixed spatial grid size, the CNR and ΔBAI value degrade as a function of the scanning step size on each dimension. Not only the scanning step size, but also the ultrasound beam spot size affect the spatial sampling resolution in the BAI imaging technique for defect detection. [Work supported by the C-FAR program, University of Illinois.]

3:10–3:30 Break

3:30

2pBB8. Spatial resolution performance of a one-dimensional linear amplitude-steered array in an attenuating medium. Sarah A. Hartleben, Virginia Lassale, and William D. O'Brien, Jr. (Dept. of Elec. and Computer Eng., Univ. of Illinois, 405 N. Mathews, Urbana, IL 61801, hartlebe@uiuc.edu)

The amplitude-steered array concept [J. Acoust. Soc. Am. **59**, 1040 (1976)] introduced the idea of steering the maximum response of a linear array by amplitude weighting the output signals of the elements. Recently, an evaluation of the linear amplitude-steered array assessed its performance such as lateral and axial resolution in a lossless medium [Frazier, Hughes, and O'Brien, J. Acoust. Soc. Am. **107**, 2430 (2000)]. In the present work, the array's performance is evaluated in an attenuating medium similar to biological tissue (1 dB/cm MHz). Similar to the array's performance in a lossless medium, the length of the array limits the axial and lateral resolution in the attenuating medium. As the steering direction is decreased, which corresponds to increasing frequency, the array's performance with attenuation has a worse resolution when compared with its performance without attenuation. For example, using a 10-cm-length array and a 6-deg steering direction (4.6 MHz), the axial resolution (-3-dB point spread function width) with attenuation was 7.9 mm, compared with 6.4 mm without attenuation. When the steering direction was increased to 12 deg (2.3 MHz), the axial resolution with attenuation was 12.8 mm compared with 11.5 mm without attenuation. In the attenuating medium there is increased importance on the steering direction or frequency. [Work supported by NIH CA09067.]

3:45

2pBB9. Elasticity imaging with time-resolved pulsed elastography. Laurent Sandrin, Micka Tanter, Stefan Catheline, and Mathias Fink (Laboratoire Ondes et Acoustique, Supérieure de Physique et Chimie Industrielles de la Ville de Paris, Université Denis Diderot, CNRS UMR 7587, 10 rue Vauquelin, 75005 Paris, France)

Time-resolved pulsed elastography is a promising technique for imaging the shear modulus of soft tissues. It is based on the investigation of a low-frequency transient shear wave using an ultrafast ultrasonic imaging system (up to 10 000 frames/s) composed of 128 channels sampled at 50 MHz and having 2 Mbytes. The system is connected to a linear array of transducers. Displacements induced by the propagating shear wave are measured using cross correlation of the ultrasonic signals. A low-frequency vibrating device was designed. The linear array of transducers is placed between two rods fixed to electromagnetic vibrators. The rods are either orthogonal or parallel to the active surface of the transducer array. The low-frequency shear waves are sent using the rods which are placed in such a way that the lobes of the induced shear waves superimpose in front of the transducer array. Large displacements are observed in the tissues which makes deeper investigations possible. Movies of the low-frequency (50–200 Hz) shear wave propagation through soft tissues can be used either to estimate the shear modulus distribution in the medium by direct local inversion, or to localize visually lesions of unexpected elasticity. In vivo measurements in human breast will be presented and discussed.

4:00

2pBB10. Quantitative elasticity imaging with ultrasound. Paul E. Barbone (Dept. of Aerosp. and Mech. Eng., Boston Univ., Boston, MA 02215 and Inst. of Cancer Res., Sutton, Surrey, UK) and Jeffrey C. Bamber (Inst. of Cancer Res. and Royal Marsden Hospital, Sutton, Surrey, UK)

For the past 10 or so years, various researchers have proposed different ways to measure strain distributions *in vivo*. The measured strains result from various sources including external palpation, low frequency vibration, or internal motion. By examining the relative strains in adjacent tissues, it is thought that one can infer the relative distributions of tissue stiffness. In this presentation, we discuss the process of creating a quantitative stiffness image from a given measurement of tissue strain. We show that the strain image by itself is insufficient information to infer the elastic stiffness. By examining the well-posedness of the inverse problem, we determine what information is needed to supplement the strain image in order to quantitatively infer the elastic stiffness distribution. Methods for obtaining the required information *in vivo* are currently being developed. We show examples of incorrect and misleading stiffness reconstructions in the absence of the required data. [Work supported by BU, NIH, Fulbright Foundation, ICR, and NSF via CenSSIS ERC.]

4:15

2pBB11. A new technique for real-time freehand ultrasonic elasticity imaging. Yanning Zhu, Timothy Hall, and Larry Cook (Dept. of Radiol., Univ. of Kansas Med. Ctr., 3901 Rainbow Blvd., Kansas City, KS 66160, yzhu@kumc.edu)

A method for high-speed, freehand ultrasonic elasticity imaging will be described. The method is based on a modified block matching technique. The modification serves two purposes. First, in order to achieve real-time performance a forward prediction mechanism is used to reduce the computational load that a conventional block matching technique requires. Second, in order to accommodate the waveform decorrelation associated with freehand scanning a statistical error detection and correction method was developed. This prevents the displacement estimation errors from propagating to affect a larger area. This new algorithm has been implemented on a Siemens Elegra (a high-end clinical ultrasound imaging system). The initial tests show that the new system can perform *B*-mode and strain display task at real-time frame rates. The strain images have low

noise and excellent contrast. Some results obtained with this system will also be presented. [Support by USAMRAA DAMD17-00-1-0596, NSF BES-9708221, and Siemens Medical Systems is gratefully acknowledged.]

4:30

2pBB12. Comparison of stiffness estimated with vibration elastography to dynamic mechanical analyzer results. Travis Oliphant (Elec. and Computer Eng. Dept., Brigham Young Univ., Provo, UT 84602), Randy Kinnick, James Greenleaf, Alex Dresner, and Richard Ehman (Mayo Foundation, Rochester, MN 55905)

Several pathological conditions induce altered tissue mechanical properties. The goal of elastography is to noninvasively generate images and/or measurements of tissue mechanical properties. Vibration elastography is a technique for measuring stiffness using low-frequency (10–1000 Hz) shear waves to vibrate a region of interest. Ultrasonic time-delay estimation techniques or phase-contrast magnetic resonance methods measure the amplitude and phase of one or more components of the harmonic vibration everywhere in the region. Under a linear, isotropic, and incompressible tissue model, the displacement field in a locally homogeneous region satisfies the vector-Helmholtz equation and stiffness estimation becomes an inverse-scattering problem where the field is known *inside* the region of interest. Using an optimal inversion algorithm recently developed by the authors and a data-dependent noise-model, an estimate of local stiffness along with its uncertainty is reported at frequencies from 200–600 Hz for three gels of varying concentrations. For each gel, the displacement is measured using either ultrasound or magnetic resonance. The vibration elastography estimate of stiffness for the three gels is less than the complex modulus measured over a frequency range of 10–250 Hz using the shear sandwich clamp on a dynamic mechanical analyzer (DMA 2980, TA Instruments) by an approximate factor of 3.

4:45

2pBB13. Response on the surface of a viscoelastic medium due to subsurface acoustic sources with application to vascular medical diagnosis. Yigit Yazicioglu, Thomas J. Royston, and Francis Loth (Univ. of Illinois at Chicago, Chicago, IL 60607, troyston@uic.edu)

The vibratory response at the surface of a viscoelastic material due to buried spherical and cylindrical acoustic sources is considered theoretically, numerically, and experimentally. An analytical solution that accounts for the compression, shear, and surface wave response to the buried source in a halfspace is formulated and compared with numerical finite-element simulations and experimental studies on finite-dimension phantom models. The focus here is on developing a better understanding of how biological soft tissue affects the transmission of vibro-acoustic energy from biological acoustic sources below the skin surface, such as a turbulent regime in an artery caused by a partial blockage. Such an understanding could catalyze the development of noninvasive procedures using an array of inexpensive acoustic sensors on the skin surface for the identification and monitoring of vascular blockages, a precursor to many serious cardiovascular diseases. [Research supported by NIH NCRR Grant No. 14250.]

5:00

2pBB14. Efficient three-dimensional cylindrical-geometry ultrasound imaging. Mark A. Haun, Douglas L. Jones (Dept. of Elec. and Computer Eng., Univ. of Illinois, 1308 W. Main, Urbana, IL 61801), and William D. O'Brien, Jr. (Univ. of Illinois, Urbana, IL 61801)

A number of new imaging modalities collect data from cylindrical platforms. *In vivo* imaging needles and intravascular ultrasound imaging catheters are examples of this geometry, where imager rotation and translation parallel to the cylinder axis are the only allowed motions. Efficient three-dimensional ultrasound image formation in these cases can be challenging when the aperture is small and/or highly curved. A frequency-domain imaging algorithm is obtained by approximating the free-space point spread function in cylindrical coordinates and obtaining its Fourier transform by analogy with the equivalent problem in Cartesian coordinates. We further propose an effective use of limited aperture by placing a focused transducer across the aperture, thereby creating a virtual source at the focus which is treated as a real, unfocused source by the imaging algorithm. This approach retains the simplicity and potential angular resolution of a small single element, yet permits full use of the available probe aperture and a higher energy output. Computer simulations and experimental ultrasonic results with wire targets show that this imaging technique attains the theoretical resolution dictated by the operating wavelength and transducer characteristics. [Work supported by NIH CA 079179.]

5:15

2pBB15. Experimental results of real-time freehand elasticity imaging. Timothy Hall, Yanning Zhu, Candace Spalding, and Larry Cook (Radiol. Dept., Univ. of Kansas Med. Ctr., 3901 Rainbow Blvd., Kansas City, KS 66160, thall@kumc.edu)

A system for high-speed calculation of tissue elastic properties using a clinical ultrasonic imaging system will be presented. The hypothesis driving this development is that real-time feedback of elasticity images is essential in obtaining high-quality data (consecutive images with high spatial coherence). Extensive experience with laboratory fixtures and off-line processing of elasticity data showed that problems occurring in data acquisition often resulted in poor elasticity image quality. The delay in observing the resulting images, due to off-line processing, resulted in slow progress in developing experimental techniques and signal processing strategies. When high-quality data were obtained, high contrast-to-noise images were available. Initial experience with real-time freehand elasticity imaging shows that images with similar high contrast to noise can be obtained. Results in volunteer patients have shown that high-quality elasticity images are easily obtained *in vivo* for a variety of breast pathologies, and the changes in breast tumor elasticity during the course of chemotherapy can be monitored. Video of real-time elasticity imaging will demonstrate the ease with which these results are obtained. [We are grateful for the support by USAMRAA DAMD17-00-1-0596, Siemens Medical Systems, and NSF BES-9708221.]

Session 2pEA

Engineering Acoustics: Analysis and Modeling of Acoustical Systems

Stephen C. Thompson, Chair
Knowles Electronics, 1151 Maplewood Drive, Itasca, Illinois 60143

Chair's Introduction—1:00

Contributed Papers

1:05

2pEA1. Numerical model of acoustic interactions between class IV flextensional transducers using a modal method. John B. BlottmanIII (Naval Undersea Warfare Ctr., Newport, RI 02841, blottmanjb@npt.nuwc.navy.mil)

Acoustic projectors assembled in an array experience an interaction effect as a result of the coupling of their individual radiated powers through the acoustic medium. The acoustic interaction may be expressed as a superposition of modal–mutual radiation impedances. A hybrid modeling technique is presented that combines the boundary integral method with a piece-parts equivalent circuit. The circuit consists of assembling a motional branch for the piezoelectric driver and for a set of flexural modes of the flextensional shell. The *in vacuo* eigenmodes of the shell are determined using the ATILA finite element method. Modal radiation impedances are generated using the CHIEF boundary integral equation method. The resulting modal–mutual impedances are applied to the motional branches through a set of source-coupling terms. The circuit then provides a set of modal participation factors. Results compare well to measurements and to full finite element models of small close-packed arrays. [Work supported by ONR and NUWC ILIR.]

1:20

2pEA2. Extending nearfield acoustical holography past intermediate sources. Edward Zechmann and J. Adin MannIII (Aerosp. Eng. and Eng. Mech., Iowa State Univ., 2271 Howe Hall, Ames, IA 50011)

With planar nearfield acoustical holography (NAH), the complex acoustic pressure and the acoustic velocity vector can be estimated on a plane near a complex source surface to identify individual noise sources. However, there are often intermediate sources between the measurement plane and the primary source plane. The NAH technique in its classical form cannot be extended past the intermediate source. A method will be presented to remove the intermediate sources so that NAH can be extended to the surface of the primary source. The normal velocity at the surface of an intermediate source is used to estimate the complex pressure due to the intermediate source on the measurement plane, which is then subtracted from the measured data, providing an estimate of the measured sound pressure if the intermediate source was not present. NAH is then used to project to the surface of the primary source. The source removal process was tested with simulated and measured data. Results for two methods to estimate the pressure caused by the intermediate source, inverse NAH and a point source approximation, will be presented. The results show the effectiveness and limitations of the source removal process. [Work supported by Ford Motor Company.]

1:35

2pEA3. A trimodal directional transducer. Alexander L. Butler, John L. Butler (Image Acoustics, Inc., 97 Elm St., Cohasset, MA 02025), and Joe Rice (Naval Postgrad. School, Monterey and SSC, San Diego, CA)

A cylindrical transducer, which achieves a directional horizontal beam through the addition of the first three extensional modes of radial vibration is presented. The three modes are the omni, dipole, and quadrupole modes, related in resonant frequency by the formula $f_n = f_0(1+n)^{1/2}$, where n=0, 1, and 2, respectively. The addition of the quadrupole mode to the omni and dipole modes yields a frequency independent horizontal beam that has greater directionality than the commonly used two-mode scheme, first described by S. L. Ehrlich and P. D. Frelich [U. S. Patent 3,290,646 (1966)]. Theoretical and measured beam pattern results on two different trimodal cylindrical transducer designs are presented and compared. One of the transducers, fabricated from rings of the same diameter, operates in the vicinity of 20 kHz and combines modes with three different resonant frequencies. The second transducer, fabricated from rings of different diameters, operates in the vicinity of 10 kHz and combines modes with the same resonant frequency. It is shown that steered beamwidths of 90° and front to back ratios of 10 dB can be readily obtained from comparatively small diameter transducers. [SBIR work supported by ONR and SPAWAR.]

1:50

2pEA4. Development and validation of a computational process for pass-by noise simulation. Zhidong Zhang, Nickolas Vlahopoulos (Dept. of Naval Architecture and Marine Eng., Univ. of Michigan, 2600 Draper Rd., Ann Arbor, MI 48109-2145), T. Allen, and K. Y. Zhang (Ford Motor Co.)

The indirect boundary-element analysis is employed for developing a computational pass-by noise simulation capability. An acoustic field reconstruction process is developed in order to generate the definition of the main pass-by noise sources in a computational model. Numerical boundary-element models that characterize the individual sources are combined in order to develop a system model for the pass-by noise simulation. The acoustic field reconstruction process is validated initially by employing analytical solutions. The numerical techniques are also validated through comparison between numerical results and test data for component-level and system-level analyses. Specially, the source definition capability is validated by comparing the actual and the computationally reconstructed acoustic field for an engine intake manifold. The overall pass-by noise simulation capability is validated by computing the maximum overall sound-pressure level for a vehicle under two different driving conditions. [Research supported by the Ford University Research Program.]

2:05

2pEA5. Development of a model for acoustic liquid manipulation created by a phased array. LeAnn E. Faidley and J. Adin MannIII (Iowa State Univ., 2271 Howe Hall, Ames, IA 50010, lfaidley@iastate.edu)

Acoustic Liquid Manipulation (ALM) is the application of the nonlinear effects of high-powered ultrasound to move buoyant objects, create fluid flow, or manipulate fluid–fluid interfaces. Work being done at NASA Glen Research Center indicates that this technique has the possibility to be used to control the location of air bubbles in space vehicle fuel tanks, keeping them away from outlet lines. In order to determine whether ALM is feasible for this and many other possible applications, a model that

predicts the nonlinear behavior of sound created by a variety of source configurations needs to be developed. This was the goal of this project. In this paper the development of a model for radiation pressure, acoustic streaming, and acoustic heating due to a phased array configuration will be outlined. The results of the model will be discussed, and a comparison will be made with experimental data.

2:20

2pEA6. Optimization of micro-loudspeaker radiation via finite element modeling. Anthony R. Bontomase and Courtney B. Burroughs (Grad. Prog. in Acoust., The Pennsylvania State Univ., State College, PA 16802)

Measurements and predictions from numerical models for the response of and radiation from a micro-loudspeaker are compared. The numerical models are then used to examine various means for improving the performance of the loudspeaker radiation by varying the material properties of the loudspeaker. A commercial finite element code uses measured material properties to predict the diaphragm velocity distribution of lower order modes. Noncontact laser vibrometer measurements on the excited speaker are compared with the FEM predictions. The radiation from the loudspeaker, at frequencies of resonance, is predicted by a boundary element model. Diaphragm material properties (density, thickness, and elastic modulus) were varied and the sensitivity of the radiation to the material properties of the diaphragm determined.

2:35

2pEA7. Modeling of perforated absorbing silencers. Ahmet Selamet, Iljae Lee (The Ohio State Univ., 930 Kinnear Rd., Columbus, OH 43212, selamet.1@osu.edu), and Norman T. Huff (Owens Corning, Inc., Granville, OH 43023)

The acoustic characteristics of single-pass perforated ducts filled with absorbing material are studied experimentally and analytically. The transmission loss predictions from three models [lumped, one-dimensional decoupled, and three-dimensional boundary element method (BEM)] are compared with experimental data from an extended impedance tube setup in the absence of mean flow. Experimentally determined complex characteristic impedance and propagation constant are used to account for the wave propagation through the absorbing material. The perforation impedance suggested by Sullivan [J. Acoust. Soc. Am. **64**, 207–215 (1978)] is modified to incorporate the effect of absorbing material. The results show that: (1) the effect of absorbing material on the perforation impedance needs to be taken into account for accurate predictions; (2) silencers filled with absorbing material exhibit significantly higher attenuation at higher frequencies than those without the filling material; (3) the three-dimensional BEM shows good agreement for the overall frequency range of interest (0–3200 Hz), while the one-dimensional approach is reasonable at relatively low frequencies; and (4) the lumped model, which treats the filled silencer as a Helmholtz resonator, may provide an approximate frequency for the location of peak attenuation.

2:50–3:05 Break

3:05

2pEA8. Interior noise prediction using a broadband high-frequency boundary-element method based on energy and intensity variables. Linda Franzoni, Brett Lussier, and Donald Bliss (Mech. Eng. and Mater. Sci., Duke Univ., Durham, NC 27708-0300)

The analysis of high-frequency broadband sound fields is computationally intensive if it is done on a frequency-by-frequency basis. In addition, traditional numerical methods require that the boundary-element or finite-element model be discretized into elements that are fractions of a wavelength in size. Despite the high-resolution numerics, in many cases the resulting broadband mean-square pressures are spatially fairly uniform. In this work a boundary element approach has been developed that uses uncorrelated broadband energy sources whose strength and directional properties satisfy certain conditions at the boundary. This method considers all frequencies at once and the element size is large compared to a typical wavelength in the frequency band. There is good agreement between the energy-based boundary-element method and exact solutions for a three-dimensional enclosure with an interior sound source. Spatial correlation effects, which are not in the current model, are also discussed. [Work sponsored by NSF.]

3:20

2pEA9. Broadband, multimode, free-flooded, baffled circular ring projectors. Boris Aronov, Tetsuro Oishi, Lawrence Reinhart, and David A Brown (Acoust. Res. Lab., Elec. Eng., Univ. of Massachusetts–Dartmouth, North Dartmouth, MA 02747 and BTECH, 1445 Wampanoag Tr., S-115, East Providence, RI 02915)

In many underwater applications, it is required to have broadband unidirectional projectors with a small electrical Q and large electrical power factor. Cylindrical transducers made of circular piezoceramic rings with a part of the surface baffled can be employed to achieve unidirectionality in the horizontal plane over a broad frequency range. In our presentation, it is shown theoretically and experimentally that the operational frequency range of a projector can be extended by simultaneously exciting the 0 (breathing) and 1 (dipole) modes of ring vibration. In addition, the corresponding resonance modes of the internal volume of the free-flooded transducer can further enhance the frequency response. The results of an experimental investigation of the electroacoustical parameters of multimode fluid-filled cylindrical transducers are presented. [This work was supported in part by SBIR N99-011: ONR321SS,ONR36-(J.Rice), BTECH(IRaD), and ONR321SS(Lindberg).]

3:35

2pEA10. Modeling absorbent finite element for numerical disipation. M. A. Picard Lopez, P. E. Solana Quiros, and J. V. Arizo Serrulla (Univ. Politecnica de Valencia, E.T.S. Ing. Industriales, E.T.S. Ing. Caminos Canales y Puertos, Camino de Vera S/N, 46022 Valencia, Spain)

The modeling of the absorbent materials is of great importance for accurate results of the sound field in irregular enclosures using different models by means of finite elements. In this paper different models of porous materials are proven, some already well known due to Craggs, for different geometries of enclosures. The computed results are compared with experimental results obtained demonstrating the reliability of the method and their precision, as well as the validity of the modelization for the absorbent elements by means of their placement in different positions inside the enclosure and for different frequencies. The physics properties used by the absorbent materials are the porosity, the mass density, the flow resistance, and the structure factor. The kinetics, potential, and dissipation energy were considered by the pattern variational used in the FEM. For the discrete distribution of points of the enclosure and the results obtained, averaging of the average residuals has been made. On the other hand, between computed and measured sound pressure values, a Pearsons correlation coefficients has been carried out. There are about 0.8 in numerous cases, an intermediate frequencies. With the absorbents and the numeric method employed, good approaches have been reached.

3:50

2pEA11. Study on acoustic transmission loss characteristics using artificial neural network approach. A. Ramachandraiah (Civil Eng. Dept., IIT Madras, Chennai-36, India) and G. Maheswari (IIT Madras, Chennai-36, India)

The sound transmission losses of multilayer walls are determined by the physical properties of the component materials. Various analytical methods exist to evaluate the transmission loss of sound in materials. Each method has its own limitations resulting in differences of the predicted and the experimental values. Of late the neural networks have emerged as powerful pattern recognition techniques. In this approach the network learns the similarities among the patterns and infers solutions from the complex, distorted data, which generally the conventional approaches fail to do. Transmission loss of materials in building acoustics seems to be one field where the neural network can be addressed. The transmission loss characteristics of a partition are generally governed by factors like density, thickness of partition, presence of absorptive material, etc. The relation between the input and output parameters is a complex one and may at times be difficult to determine by conventional methods. A neural network has been developed to predict the transmission loss and transmission class values of composite partitions with reasonable precision. The generalization of the network is also tested. The performance of the neural network is observed and the results obtained through neural network approach are analyzed and discussed.

4:05

2pEA12. Simulating an audio beam: A second-generation wavelet collocation method to numerically solve the Khokhlov–Zabalotskaya–Kuznetsov equation. Kelvin C. M. Lee (DSP Lab., School of Elec. and Electron. Eng., Nanyang Tech. Univ., Nanyang Ave. S., 639798 Singapore), Yew-Hin Liew, and Woon-Seng Gan (ecmlee@ntu.edu.sg)

The Khokhlov–Zabalotskaya–Kuznetsov (KZK) equation is a nonlinear parabolic wave equation which describes an intense finite-amplitude directional sound beam and accounts for the combined effects of diffraction, absorption, and the nonlinearity behavior of the sound beam in air. A general exact solution has not been found for the KZK equation. Nevertheless, it is usually solved by approximate methods or numerical techniques. A second-generation wavelet collocation method using a lifting scheme is used to numerically solve the KZK equation in a highly accurate and efficient manner and has been found to have significant improvements over the conventional finite-difference scheme. Using a new approach to construct wavelets on the interval in the spatial domain (independent of the Fourier transform), the method uses a computational grid which adapts dynamically in time to allow for solution refinement in local regions with sharp transitions, e.g., development of shocks. It is also found to be particularly effective in the treatment of nonlinear terms and general boundary conditions in the equation. Furthermore with a lifting scheme, the wavelet transform can be computed in place of the original signal without the need for auxiliary memory.

TUESDAY AFTERNOON, 5 JUNE 2001 SALON FOYER, 1:00 TO 2:00 P.M.

Session 2pEDa

Education in Acoustics: Undergraduate Research Poster Session

E. Carr Everbach, Chair
Department of Engineering, Swarthmore College, 500 College Avenue, Swarthmore, Pennsylvania 19081-1397

Contributed Papers

All posters from Session 2aEDb will be on display from 1:00 p.m. to 2:00 p.m. Authors of even numbered papers will be at their posters from 1:00 p.m. to 2:00 p.m.

TUESDAY AFTERNOON, 5 JUNE 2001 SALONS VII/VIII, 2:00 TO 4:00 P.M.

Session 2pEDb

Education in Acoustics: Take Fives for High School Presentations: Sharing Ideas for Teaching Acoustics

Victor W. Sparrow, Chair
Graduate Program in Acoustics, Pennsylvania State University, 157 Hammond Building, University Park, Pennsylvania 16802

Chair's Introduction—2:00

Do you have a novel demonstration, a new laboratory experiment, a favorite video, a recorded sound example, or a new idea for teaching acoustics which you are willing to share with your colleagues? At this session a sign-up board will be provided for scheduling presentations. No abstracts are printed. Presenters are encouraged to have handouts to distribute. Multiple presentations are acceptable (not consecutively). Presentations are limited to 5 minutes. Keep them short! Keep them fun!

Session 2pMU

Musical Acoustics: Experimental Musical Instruments

James P. Cottingham, Chair
Department of Physics, Coe College, Cedar Rapids, Iowa 52402

Chair's Introduction—1:00

Invited Papers

1:05

2pMU1. Contemporary ideas in musical instrument making: An overview of recent trends. Bart Hopkin (Exp. Musical Instruments, P.O. Box 784, Nicasio, CA 94946)

In this talk I will discuss some particularly inventive ideas that have arisen in acoustic musical instrument making in recent years. We'll look at broad trends in instrument design as well as specific instrument types. We will cover acoustic chorusing instruments such as Sharon Rowell's Huaca and Richard Water's Waterphone, musical uses of longitudinally vibrating strings such as Ellen Fullman's Long String Instrument, just intonation and harmonics-oriented instruments such as Hans Reichel's Pick-Behind-the-Bridge Guitar and Jacques Dudon's Photosonic Synthesizer, balloon-mounted and bowed-metal instruments from the Baschet brothers and Tom Nunn, and many more. There will be sound recordings and visuals, and I may be able to bring one or two smaller instruments along for show and tell.

1:35

2pMU2. New and unusual percussion instruments. Thomas D. Rossing (Phys. Dept., Northern Illinois Univ., DeKalb, IL 60115)

Many new and unusual percussion instruments have been developed in recent years, and more are in the experimental stage. What is often termed ''contemporary sound'' makes extensive use of percussion instruments. We will describe the acoustical properties of a number of new percussion instruments made of wood, metal, glass, and stone. Normal modes of vibration and sound radiation will be emphasized.

2:05

2pMU3. Sounding clay: Pre-Hispanic flutes. Susan Rawcliffe (P.O. Box 924, San Pedro, CA 90037)

Ms. Rawcliffe makes and plays ceramic flutes and sound sculptures. Many were inspired by her explorations into ancient and wonderful wind instruments. To build a better flute, she studied construction methods of both contemporary and pre-Hispanic artists. The laws of acoustics dictate the possibilities for instrument construction within which design decisions are made according to cultural and individual preferences. She makes acoustical copies, learns to play them, then reinvests her insights, evolving through stages into new instruments, which inspire with their wonderful scales and evocative timbres. For 30 centuries, societies from the Olmecs to the Mayans and Aztecs developed a unique flute organology in a great diversity of form, timbre, and tunings. Many of the most complex innovations of the pre-Hispanic ceramists resulted in instruments of restricted pitch but rich timbre. They are not so much the ''remains of a bygone art, as the sacred sound symbols of a now vanished cult.'' In reviving this lost craft and science, Ms. Rawcliffe is able to play sounds that may not have been heard for 1000 years. For this presentation, using slides of her own and of ancient flutes, and performances on her flutes, she will discuss and illustrate the above issues.

2:35

2pMU4. The development of ''21st century'' acoustic and orchestral instruments. Patrick Ozzard-Low (Ctr. for New Musical Instruments, London Guildhall Univ., 41 Commercial Rd., London E1 1LA, UK)

The work of the Centre for New Musical Instruments (CNMI), recently established at London Guildhall University, is reported. In addition, a wide range of original slides is shown, providing an overview of international innovations relevant to the development of new versions of existing acoustic orchestral instruments, especially those offering new forms of expression to composers and performers. In view of the overwhelming majority of developments of electronic as opposed to acoustic instruments during the 20th century, the purpose of CNMI is to encourage innovations of the latter. Acoustic instruments designates not only those traditionally referred to by this term, but also those in which elements of sound generation may be electromechanical or electroacoustic, although sound diffusion is not through a loudspeaker. Three areas of innovation are discussed: (i) purely mechanical instruments; (ii) electromechanical and electroacoustic hybrids; and (iii) instruments designed for alternative tuning systems. All the orchestral families are considered. The mutually reinforcing possibilities and limitations of these elements, together with the evolving aesthetic issues of contemporary and other musics, suggests the importance of specific areas of future research and instrument making. Amongst these is the necessity of interdisciplinary perspectives and collaboration.

3:05–3:15 Break

3:15

2pMU5. The Woodstock Gamelan. Lydia Ayers and Andrew Horner (CS Dept., HKUST, Clear Water Bay, Kowloon, Hong Kong)

This paper considers the spectral properties of the Woodstock Gamelan, a 3-octave set of tubular chimes built by Woodstock Percussion in upstate New York. One of the main features of the instrument is expandability, and it includes 75 microtones in the middle octave. The justly tuned instruments of Harry Partch inspired the instrument. The Woodstock Gamelan has two types of aluminum tubes. Racks support the tubes down to Eb4, and these have a vibraphonelike timbre. The larger hanging tubes go down to G3, and have larger diameters than the rack tubes. The sound of the hanging tubes is similar to that of orchestral chimes. The Woodstock Gamelan has five exponentially decaying partials, and their frequency ratios are the same in the low and high registers. The frequency ratios measured were about 1, 2.69, 5.15, 8.38, and 12.08. The ratios are close to the just ratios of 1, 2.667, 5.333, 8.533, and 12. A Csound model for the Woodstock Gamelan has been developed. Listening tests show that the model produces tones nearly indistinguishable from the original. The model produces attractive related timbres by simple changes to the parameters.

3:45

2pMU6. New woodwind instruments. Lewis Jones (Musical Instrument Technol., London Guildhall Univ., 41 Commercial Rd., London E1 1LA, UK, liones@lgu.ac.uk)

Progress towards the development of woodwind instruments designed to play fluently in tuning systems other than 12-note equal temperament is described. The problems of closing many toneholes with only ten digits are outlined, and past solutions, depending chiefly on rod and lever keywork, are briefly reviewed. The alternative of largely keyless toneholes, as embodied in Jones and Armitage's 19ET recorders (1999) is considered. The electromechanical means of tonehole closure introduced in Giles Brindley's ''logical bassoon'' is taken as a point of departure for a range of woodwind instruments including flutes dividing the octave into as many as 36 steps, others of exceptionally deep pitch, and others, narrow but of great length which, by exploiting the higher harmonics of the air column can, without impossibly many toneholes, produce divisions of the octave into 72 or 96 steps. Softare mediates between practicable fingerings and at times complex patterns of tonehole movement. A scheme of fingering is proposed which is intended to enable a single player to learn to play in many alternative tuning systems while preserving as many essential parallels between fingering and sound as possible.

4:15

2pMU7. Conventional harmonic behavior associated with the equal tunings of 17 and 19 notes. Easley Blackwood (5300 S. Shore Dr., Chicago, IL 60615, eblackwood@interaccess.com)

It will be demonstrated how diatonic behavior associated with the equal tunings of 17 and 19 notes is virtually identical to that of the standard 12-note equal tuning. However, levels of discordance are noticeably different among the three tunings. The reasons for this will be shown in detail. It will also be shown how the conventional musical notation—the five-line staff with treble and bass clef, as well as accidentals—is an accurate and comprehensible representation of the tunings of 17 and 19 notes. Musical illustrations in the form of brief (4-minute) compositions, electronically created, can be provided. Funding for this study was provided by a grant from the NEH, and administered through Webster College, St. Louis.

4:45–5:45

Informal Performance and Demonstration of Experimental Instruments

Session 2pNS

Noise and Architectural Acoustics: Motor Vehicle Interior Noise Control

Daniel R. Raichel, Chair
2727 Moore Lane, Fort Collins, Colorado 80526

Chair's Introduction—1:00

Invited Papers

1:05

2pNS1. Customizing vehicle interior noise to reflect the brand image. Hans P. Schedl (Audi AG, 85045 Ingolstadt, Germany)

Noise control engineers in their endeavors to make vehicles as quiet as possible have succeeded thanks to major strides in measurement techniques. The different noise sources in vehicles are now quite well known. Additionally, sounds may be simulated and easily modified on simple PCs. Through psychoacoustic means, we are able to evaluate and understand the effects of noise on human beings. A totally new approach to vehicle interior noise now exists. Engine noise conveys meaningful information to the driver, and thus, noise characteristics constitute a design attribute. Manufacturers are increasingly establishing design rules for the sound of their cars to meet special requirements and preferences of their potential customers. The design rules are meant to yield a vehicle brand-specific sound. This paper analyzes the positive and negative effects of noise on both driver and passengers in a vehicular environment and how sound is used as a marketing and public relations instrument. We also examine briefly the way sound-effect design affects a project's progress, cost and time limits. An example will demonstrate how the interior noise of a basic car is modified to either sound like an upper-class luxury sedan or like a high-performance sports car.

1:30

2pNS2. Quantifying customer perceptions of impulsive vehicle noise. Mike Blommer, Scott Amman, and Deanna Hoffman (Ford Motor Co., SRL Bldg., MD 2115, P.O. Box 2053, Dearborn, MI 48121, mblommer@ford.com)

Sounds such as spark and diesel knock, squeaks and rattles due to body and suspension components, gear rattle, and other impulsive events, can occur in vehicles and be a major source of customer dissatisfaction. It is desirable to not only know the detection thresholds of these impulsive events, but also the relative annoyance they impart on the customer once they become audible. This work describes research addressing both aspects. The first part of the paper presents a generalized detection model of impulsive events in common vehicle background noises (e.g., wind, road, and powertrain noise). Important properties of the model are the combination of impulsive event information across frequency and also the effect of overall background noise level on detection thresholds. Application of the model to predict detection thresholds for spark knock and also squeaks and rattles is presented. Comparisons are made to measured subjective thresholds. The second part of the paper presents research in objectively quantifying the relative annoyance of impulsive sounds as a function of their temporal and amplitude distributions, as well as their loudness. Comparisons are made to objective metrics used for characterizing single impulsive events.

1:55

2pNS3. Prediction of vehicle passenger compartment noise due to turbulent flow excitation. Sean Wu and Guoming Wu (Dept. of Mech. Eng., Wayne State Univ., 5050 Anthony Wayne Dr., Detroit, MI 48202)

An engineering computer model is developed for estimating noise transmission into a vehicle passenger compartment, given the characteristic dimensions of the vehicle and turbulent flow excitation. In this model, the effects of window offset and leakage, and those of power spectral densities due to separated and reattached flows are taken into account. Empirical formulations are developed to estimate the reattachment line and the power spectral densities of separated and attached flows. The effect of leakage is accounted for by a correction factor. The computer model thus developed is validated on a greenhouse vehicle model. Experiments are conducted in a wind tunnel at the Chrysler Technology Center. Pressure fluctuations due to separated and attached flows on the side window are measured and compared with the calculated values. These pressure fluctuations are taken as input to a computer program, called VibroAcoustic Payload Environment Prediction System based on Statistical Energy Analysis, to calculate the noise spectrum, given the A-pillar angle, window size and offset, vehicle yaw angle, and wind speed. The predicted noise spectra inside the passenger compartment are compared with the measured ones under various flow speeds. Satisfactory agreements are obtained in all cases. [Work supported by a Chrysler Challenge Fund.]

2:20

2pNS4. Future trends in the automotive sound package industry. Pranab Saha (Kolano and Saha Engineering, Inc., 3559 Sashabaw Rd., Waterford, MI 48329, prsaha@kandse.com)

With various challenges in the automotive industry, there are enormous opportunities in sound package treatments. Although the fundamentals of the sound package practices have changed very little, the technology has changed significantly, and this has brought about new opportunities. More and more the supplier companies are asked to join OEM companies in actively assisting with the

optimization of sound package treatments in vehicles. This environment encourages innovation and suppliers now can bring in new concepts and products that were not thought of before. This adds more value which is a combination of performance, cost, and weight. New products that are coming to the industry are products like fine fibers and barriers for airborne noise, and sprayable damping materials for structure-borne noise. In addition, there are new test methods that are intended to validate these products which will help acoustics in the vehicle. This paper discusses these trends, and associated opportunities and challenges in the automotive sound package industry in the years to come.

2:45–2:55 Break

2:55

2pNS5. Computational methods for prediction and control of automotive interior noise. S. T. Raveendra (Collins & Aikman, 47785 W. Anchor Court, Plymouth, MI 48170)

Prediction of noise and vibration characteristics is important to the automotive community. Typically, noise and vibration testing are performed on prototypes of vehicles during pre-production stages of the development. Although testing is an essential tool, it is not the most effective tool since it is done at the later stages of the development. As a result it is not possible to make major design modifications that may be necessary for the effective treatment of the noise and vibration problems. On the other hand, computational modeling and simulation is feasible at the early stages of the design and is thus effective for the management of noise and vibration problems. This presentation will demonstrate how boundary-element method based techniques can be used to predict and effectively control noise in the interior of an automobile. Initially, a technique based on acoustic contribution analysis that intuitively allows the reduction of noise in the interior of an automobile will be demonstrated. This will be followed by a demonstration of an integrated experimental/computational simulation technique based on the near-field acoustical holography that permits the identification of panel vibrations that contribute to the noise in the interior of an automobile.

3:20

2pNS6. Effects of brake pressure and temperature on squeal noise. Alessandro Mattiuzi Balvedi (MSC Laminates and Composites, Inc. and Univ. of Santa Catarina, Santa Catarina, Brasil) and Ahid Nashif (MSC Laminates and Composites, Elk Grove Village, IL 60007-5995)

Brake squeal noise has been an important issue for the automotive industry. This phenomenon is related to the modal coupling of the rotor and the pads of a brake system. When a mode of the pad matches with a mode of the rotor, the system becomes unstable. The consequence is the generation of additional vibration energy that cannot be dissipated by the system. In this manner, a solution is to improve the damping capacity of the brake system. This is accomplished by use of metal and viscoelastic materials laminates, where some of the vibrational energy is dissipated by shearing or extending the viscoelastic layer. Since the dynamic behavior of brake systems is a function of frequency, temperature, and brake pressure (friction force), it is of utmost importance to verify the influence of such parameters in order to design and select a proper multilayer treatment. This work presents a procedure to verify the influence of these parameters in the coupling of rotor and pads modes. Different conditions of pressure and temperature are applied on a brake system and the dynamic of the system related to squeal noise is evaluated.

Contributed Papers

3:45

2pNS7. Burst mode dither control of automotive brake squeal. Mawuli Dzirasa, Kenneth A. Cunefare (The Georgia Inst. of Technol., Atlanta, GA 30332-0405), Victor Rastelli, and Nila Montbrun (Universidad Simon Bolivar, Caracas 108, Venezuela)

Brake squeal is an annoying phenomenon, generally occurring when vehicles brake at slow speeds. Significant effort is expended by the automotive industry to avoid squeal during the development of brake systems. Nonetheless, expensive warranty claims still occur when such efforts are not completely successful. Recent research demonstrated that introducing dither control into a brake is an effective means of suppressing squeal. Dither control is a method by which high-frequency disturbances are introduced into a system. This suppresses the friction induced squeal response at lower frequencies, thus eliminating the audible brake squeal. Dither control, when applied at a 100% duty cycle, is an effective means of suppressing brake squeal. This presentation focuses on the potential to use burst mode dither control for squeal suppression. Burst mode dither control is characterized by duty cycles of less than 100%. Dither is introduced to a brake by placing a piezoelectric stack actuator in the piston of a floating caliper brake. The piezoelectric stack actuator is driven by a 20-kHz burst mode signal, and the impact upon the signal is assessed. Burst mode dither with duty cycles as low as 50% are shown to be as effective in suppressing brake squeal as dithering at a 100% duty cycle.

4:00

2pNS8. An experimental investigation of techniques for reducing flow-induced cavity resonance. Paul Zoccola (Carderock Div. Naval Surface Warfare Ctr., 9500 MacArthur Blvd., West Bethesda, MD 20817, zoccolapj@nswccd.navy.mil)

Excitation of cavity resonance by flow over an aperture is often a source of unwanted noise in aerospace, automotive, and marine applications. An experimental investigation of three resonance reduction techniques was conducted. These were: a fence at the upstream edge, fluid injection, and a new technique developed by the author, in which fluid from the boundary layer is diverted into the cavity. Spectra of the pressure in the cavity were obtained for various flow speeds. Results show that the fence and the boundary-layer diversion technique have the effect of reducing the Strouhal number of the flow-excited sheartones, delaying resonance to higher speeds. Reduction at the resonant speed is also observed. The boundary layer diversion technique was much more effective than the fence. The Strouhal number reduction effect is not observed with fluid injection. The effect on resonance reduction by fluid injection as a function of several parameters, including the rate of fluid injection, is discussed.

4:15

2pNS9. Visualizing acoustic field inside a passenger vehicle compartment. Moondra Manmohan and Sean Wu (Dept. of Mech. Eng., Wayne State Univ., 5050 Anthony Wayne Dr., Detroit, MI 48202)

The objective of this study is to examine the effectiveness of the HELS method [S. F. Wu and J. Yu, J. Acoust. Am. **104**, 2054–2060 (1998); S. F. Wu, *ibid.* **107**, 2511–2522 (2000)] in visualizing the areas that are prone to noise transmission into a full-size vehicle passenger compartment due to engine, powertrain system, tires, and wind excitations. The input to the HELS formulation is the acoustic pressure measured inside the compartment. No vehicle geometry and dimensions are necessary. The optimum number of expansion functions is determined by minimizing the sum of the squared errors with respect to the measured data. Once the HELS formulation is established, the acoustic pressure anywhere including the vehicle interior surface can be determined. The normal component of the surface velocity can be reconstructed in a similar manner. Once these quantities are specified, the normal component of the time-averaged acoustic intensity and acoustic energy flow inside a vehicle passenger compartment can be visualized. This three-dimensional acoustic image can provide valuable insight into vehicle interior noise reduction. The effects of number and locations of measurements on the accuracy of reconstruction are investigated. [Work supported by the Daimler-Chryster Challenge Fund.]

4:30

2pNS10. Fiber optic sensor system for structural acoustics control. Miao Yu and Balakumar Balachandran (Dept. of Mech. Eng., Univ. of Maryland, College Park, MD 20742, mmyu@eng.umd.edu)

Recent work on acoustic measurements using a fiber-tip-based Fabry–Perot sensor system is presented. A single Fabry–Perot sensor using a path matched Mach–Zehnder interferometer is developed, and by taking advantage of an integrated optical circuit phase modulator, a digital demodulation scheme based on the phase stepping technique is developed. It has been determined that this system can be used to detect acoustic fields in the frequency range of 20 Hz–6 kHz with a sensitivity of 0.9 rad/Pa. This sensor is designed to be used in a multiplexed architecture to provide inputs to a structural acoustic control system. A series of experiments are performed to investigate the possibility and potential use of this sensor system for acoustic noise detection. In this study, initial test data from the prototype optical sensor microphone are presented and the envisioned multiplexed sensor scheme and control system are illustrated.

TUESDAY AFTERNOON, 5 JUNE 2001 CRYSTAL ROOM, 1:00 TO 3:35 P.M.

Session 2pPAa

Physical Acoustics: Extraterrestrial Acoustics

George Mozurkewich, Chair
10671 Canton Center Road, Plymouth, Michigan 48170

Chair's Introduction—1:00

Invited Papers

1:05

2pPAa1. ''Sonoluminescence'' from the early Universe. Michael Turner (Dept. of Astron. & Astrophysics, Univ. of Chicago, 5640 South Ellis Ave., Chicago, IL 60637 and Fermi Natl. Accelerator Lab.)

The tens of microkelvin variations in the temperature of the cosmic microwave background (CMB) across the sky are snapshots of the Universe at 400 000 years of age and a Rosetta Stone for unraveling its earliest history. During the first 400 000 years the lumpy distribution of dark matter excited gravity-driven acoustic oscillations of the ordinary matter. Pressure provided by the photons in the CMB, which were tightly coupled to the ordinary matter, acted as the restoring force for these compressional modes. During an oscillation, photons were alternately compressed and heated, and rarefied and cooled. Ordinary matter and photons decoupled at 400 000 years, and CMB photons streamed freely to us, creating a snapshot of modes caught at maximum compression (hot spots) and maximum rarefaction (cold spots) at this early time. The pattern of hot and cold spots on the CMB sky today is being used to determine the curvature of the Universe, the Hubble constant, the total amount of ordinary matter and of dark matter, and to test theories of the early Universe including inflation.

1:35

2pPAa2. Helioseismology—sound inside the sun. Frank Hill (Natl. Solar Observatory, 950 N. Cherry Ave., Tucson, AZ 85719)

In 1960, the solar surface was found to be moving radially with a period of 5 min. The cause of this effect was found to be the trapping of acoustic waves in the solar internal thermal gradient. Since the sound fills a resonant cavity, only specific oscillatory modes are present whose characteristics depend on the plasma properties. Thus, the solar interior can be remotely sensed by observing the trapping region's upper boundary at the surface. This led to the field of helioseismology and produced new information about the solar internal rotation rate, composition, neutrino flux, convection zone depth, and internal changes associated with the solar activity cycle. It has also generated progress in remote sensing techniques (acoustic holography, ring diagrams, inversions), observing facilities (GONG Network, MDI instrument on SOHO), and insights into the coupling of turbulent convection and sound. Presumably, other stars also contain sound and their interiors could thus be probed, but so far no conclusive observations of these vibrations have been obtained. [Work supported by NSF and NASA.]

2:05

2pPAa3. Modeling acousto-gravity waves in the atmosphere of Jupiter. Joseph F. Lingevitch, Michael D. Collins (Naval Res. Lab., Washington, DC 20375), and Catherine Stamoulis (Dept. of Ocean Eng., MIT, Cambridge, MA 02139)

The modeling of wave propagation in the atmosphere of Jupiter became a topic of great interest after the discovery of the fragments of Comet Shoemaker-Levy 9. Circular fronts were observed expanding from several of the impact sites [Hammel *et al.*, Science **267**, 1288–1296 (1995)], but the nature of these features has not yet been explained in terms of a wave model. This is due in part to a lack of data, which has been enhanced by direct measurements from the Galileo probe [Seiff *et al.*, J. Geophys. Res. **103**, 22 857–22 889 (1998); Atkinson *et al.*, *ibid.* **103**, 22911–22928 (1998)]. There have also been some recent improvements in propagation modeling using parabolic equation techniques, which can handle acousto-gravity waves [Lingevitch *et al.*, J. Acoust. Soc. Am. **105**, 3049–3056 (1999)] and the advective effects of winds [Collins *et al.*, J. Acoust. Soc. Am. **97**, 2147–2158 (1995)]. These techniques and their application using Galileo data will be presented. [Work supported by ONR.]

2:35

2pPAa4. Global infrasonic monitoring of large meteoroids. Douglas O. ReVelle (Los Alamos Natl. Lab., P.O. Box 1663, MS J577, Los Alamos, NM 87545)

The hypersonic meteoroid/bolide–atmosphere interaction can generate numerous phenomena, including strong shocks and infrasonic waves. For a nonfragmenting bolide, the line source, blast wave radius, Ro, is the product of the Mach number and the body diameter. For bolides reaching continuum flow, Ro can range from ~1 m to many kilometers. To be detected, Ro must exceed ~10 m [at a minimum source energy, Es, of ~4.10(7) J]. Beyond 10 Ro, a frequency equal to Cs/(2.81 Ro) will be generated which will change during propagation due to effects of nonlinearity, absorption, and dispersion. Infrasonic arrays are now being routinely used to detect and locate such large near-Earth objects. Data from the CTBT IMS monitoring system will help refine estimates of the large body influx rate. These arrays have baselines of ~1 to 2 km using microbarometers with response from ~10 to 0.01 Hz. For ranges from 100 to 14 000 km, such bolides have been observed previously with amplitudes from 0.02 to 16 Pa, with periods from 0.50 to 300 s, with total durations from 0.10 to 25 min and with computed acoustic efficiencies varying from 0.01% to ~10%. The influx rate will be discussed as well as several recent bolides detected by Los Alamos arrays and by US DoD satellites.

Contributed Papers

3:05

2pPAa5. Absorption of sound in the Martian atmosphere. Henry E. Bass and James P. Chambers (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, University, MS 38677, chambers@olemiss.edu)

Future missions to the planet Mars might include microphones to listen for sounds in the tenuous Martian atmosphere. The chemical composition of the atmosphere is well established by previous missions and ground based observations. The dominant constituent is CO_2 with a minor amount of N_2 and argon and smaller amounts of H_2O. The factors important to compute the absorption of sound in a gas are reasonably well known, the most uncertain being the relaxation time of CO_2 at the low temperatures encountered (200 K–300 K). By extrapolating higher temperature measurements of relaxation times, analytical expressions have been developed for the absorption due to viscosity and thermal conduction (classical absorption), rotational relaxation, and vibrational relaxation. Calculations are presented for a surface level pressure of 6 millibars (600 Pa). The absorption at mid-audio frequencies (500 Hz) is 0.03 (200 K)–0.1 (300 K) Np m^{-1}. This is about 100–500 times greater than for the earth's atmosphere (depending upon relative humidity).

3:20

2pPAa6. The other ocean: probing Europa's interior with natural ambient noise sources. Aaron M. Thode, Michele Zanolin, Sunwoong Lee, Purnima Ratilal, Josh Wilson, and Nicholas C. Makris (MIT, 77 Massachusetts Ave., Cambridge, MA 02139)

Jupiter's moon, Europa, possesses a surface composed primarily of water ice. Gravity and magnetic data collected by the NASA Galileo Orbiter over the past five years have provided increasing evidence that an ocean exists underneath this layer, and that the combined ice/water layer thickness is 80–170 km, although the ice shell thickness remains unknown. The surface is covered by numerous fractures and ridges, believed to have formed in response to tidal deformations generated by Europa's slightly eccentric 85-hour orbit around Jupiter. A recently published model [Hoppa *et al.*, Science **285**, 1899–1902] argues that certain cycloidal fractures must form during a single revolution, and propagate horizontally at an average speed of 3 km/h. Considerations from ice mechanics suggest that these propagating fractures would generate significant acoustic energy in the frequency range 0.5–4 Hz, where low attenuation in the ice/water environment is expected. In this presentation an acoustic sound speed profile for Europa is constructed, and standard ocean acoustic techniques are used to demonstrate how an array of geophones on the icy surface could simultaneously localize discrete events and invert for the ice-layer thickness.

Session 2pPAb

Physical Acoustics: Materials Characterization

Stanley A. Cheyne, Chair
Department of Physics and Astronomy, Hampden-Sydney College, Box 158, Hampden-Sydney, Virginia 23943

Contributed Papers

3:45

2pPAb1. Correlation of resonance spectra excited by acoustic waves on elongated elastic objects with elastic bulk parameters and object identification. H. Uberall and M. F. Werby (NRL Code 7180, Stennis Space Center, MS 39529 and Dept. of Phys., The Catholic Univ. of America, Washington, DC 20064)

Acoustic resonance spectra are calculated for scattering of acoustic signals from elongated objects composed of six elastic materials with two aspect ratios of 3 and 6. The incident field is along the axis of symmetry and broadside. A comparison is made of the resonance locations of the six materials (brass, nickel, aluminum, steel, manganese, and tungsten carbide) representing a broad spectrum of elastic materials. For the principal resonances the ratios are seen to correspond exactly to the Rayleigh phase velocities on an evacuated half-space, or alternatively to the shear bulk velocity and are a function of the Poisson ratios of the material. Further, the resonance widths are related inversely to the density of the material and the shear velocity (the mechanical impedance of the shear wave). Time-domain calculations are also carried out and the resonance widths and travel times may be identified with the material properties of the target. Thus, the material properties of such objects including elongation may be distinguished for submerged objects, and this is a useful tool for inverse issues.

4:00

2pPAb2. Gel formation monitoring by acoustic spectroscopy. Loïc Martinez, Stéphane Serfaty, Brahim Senouci (Lab. d'Electronique Appliquée, Univ. de Cergy, 5 mail Gay-Lussac, F 95 031 Neuville sur Oise Cedex, France, loic.martinez@iupge.u-cergy.fr), Pascal Giesmar (Lab. de Physico-Chimie des Materiaux Organiques, Univ. de Cergy Pontoise, F 95 031 Neuville sur Oise Cedex, France), and Marcel Gindre (Lab. d'Imagerie Paramétrique, UMR CNRS 7623, F 75006 Paris, France)

An acoustic technique in the audible range has been developed to characterize the sol–gel process. Resonances appear at the sol to gel transition of a sol–gel matrix when submitted to an acoustic wave. The range of the associated resonance frequencies leads to a very low propagation speed of sound (about 20 m/s). The resonance frequencies versus time curves, corresponding to the harmonic propagation modes, converge to a unique intersection point with the time axis corresponding to the gelation time t_g. The temporal evolution of the resonance frequencies features the formation of the network. Actually, the evolution of the matrix is independent of the initial conditions (precursor concentration, hydrolysis rate). Depicting the ''reduced frequency'' $f_i/f_i(\infty)$ [$f_i(\infty)$ is the long-term resonance frequency for the harmonic mode i] versus the ''reduced time'' t/t_g for various Si concentrations and hydrolysis rates results in a unique curve, revealing the insensitivity of the matrix formation process to the input parameters. A propagation model of an acoustic wave in a cylindrical cavity is proposed, to compute the resonance frequencies as a function of the longitudinal to transverse propagation velocity ratio C_L/C_T. The comparison with the experimental data is conclusive.

4:15

2pPAb3. The influence of a strong acoustic cw wave on a weak signal in rock. Irina Soustova, Veniamin Nazarov, Alexander Sutin, and Andrey Radostin (Inst. of Appl. Phys. RAN, Nizhny Novgorod, Russia)

We are conducting studies of the effect of a strong pump wave on a probe wave in rock. Specifically, we are measuring resonance frequencies and quality factors (applying linear resonant ultrasound spectroscopy) for a relatively weak probe wave in the presence of a strong, cw pump wave. Experiments conducted in sandstone demonstrate a shift of the resonance frequency and increasing of Q factor in the presence of the low frequency pump wave. The model describing this phenomenon is based on including small hysteretic loops in the stress–strain dependence for the probe wave. The hysteretic loop parameters depend on the loop position on the stress–strain dependence for the pump wave. [Work was partially supported by International Science Technical Center, Grant No. 1369.]

4:30

2pPAb4. Resonant ultrasound spectroscopy studies of clathrate thermoelectrics. Michael McGuire, Alem Teklu, Veerle Keppens (Natl. Ctr. for Physical Acoust. and Dept. of Phys., The Univ. of Mississippi, Oxford, MS 38677), David Mandrus, and Brian Sales (Solid State Div., Oak Ridge Natl. Lab., Oak Ridge, TN 37831)

Resonant ultrasound spectroscopy (RUS) measurements have been carried out for 2 cubic clathrate materials, $Eu_8Ga_{16}Ge_{30}$ and $Ba_8Ga_{16}Ge_{30}$. In these materials, the Eu and Ba ions rattle in oversized atomic cages. They have attracted attention as promising thermoelectric materials, having thermal conductivities with a glasslike order of magnitude while maintaining crystalline electronic properties. RUS has proven to be useful for the study of similar cage-like materials [Keppens *et al.*, Nature **395**, 876–878 (1988)], identifying 2 local modes in the filled skutterudite $La_{0.75}Fe_3CoSb_{12}$. The RUS measurements we present here were carried out as a function of temperature (2-300K) on single crystals, and allow the determination of the 3 elastic moduli. The results are compared to ultrasonic attenuation measurements on $Sr_8Ga_{16}Ge_{30}$ and RUS measurements on filled and unfilled skutterudites. [Work supported in part by ONR. Oak Ridge National Laboratory is managed by UT-Battelle, LLC, for the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.]

4:45

2pPAb5. Diffuse backscattered ultrasound for material characterization. John Burkhardt (Mech. Eng. Dept., U.S. Naval Acad., Annapolis, MD)

An ultrasonic technique is proposed that allows for the characterization of scattering materials by the Legendre moments of their differential scattering cross section per unit volume. Estimates of Legendre moments are determined from an ultrasonic backscatter theory that models acoustic propagation in the medium as a time-dependent diffusion process. Monte Carlo simulations are presented which support the proposed technique. Initial experimental progress will be reported.

Session 2pPP

Psychological and Physiological Acoustics: Physiology; Binaural Processing (Poster Session)

William P. Shofner, Chair
Parmly Hearing Institute, Loyola University of Chicago, 6525 North Sheridan Road, Chicago, Illinois 60626

Contributed Papers

All posters will be on display from 1:30 p.m. to 5:00 p.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 1:30 p.m. to 3:15 p.m. and contributors of even-numbered papers will be at their posters from 3:15 p.m. to 5:00 p.m.

2pPP1. Fluid volume displacement at the stapes footplate and round window due to air and bone conduction stimulation. Stefan Stenfelt (Dept. of Signals and Systems, Chalmers Univ. of Technol., Goteborg, Sweden), Naohito Hato, and Richard L. Goode (Stanford Univ. Medical Ctr., Stanford, CA)

The fluid volume displacement of the two major in- and outlets of the cochlea, the stapes footplate and the round window, are normally considered equal but with opposite phase. However, other channels, such as the cochlear and vestibular aqueducts, may affect the fluid flow in the cochlea. To test this hypothesis, the volume displacement at the stapes footplate and the round window was measured with a laser Doppler vibrometer with air conduction stimulation in seven fresh human temporal bones and with bone conduction stimulation in eight temporal bones. The volume displacement was computed by measuring at five positions on the footplate and by scanning the motion of the round-window membrane at approximately thirty positions. With air conduction stimulation, the average volume displacement of the two windows was within 5 dB of each other and the phase difference was close to 180 deg within the frequency range 0.1–10 kHz. With bone conduction stimulation and below 1.5 kHz the volume displacement of the round window was about 10 dB greater than at the footplate with a phase difference of 150–200 deg. Above 1.5 kHz this difference rolled off with 6 dB/octave and 100 deg/octave.

2pPP2. Oxidative DNA damage associated with intense noise exposure in a rat. Luann E. van Campen, William J. Murphy (Hearing Loss Prevention Section, NIOSH, 4676 Columbia Pkwy., M.S. C-27, Cincinnati, OH 45226-1998), and Mark A. Toraason (NIOSH, Hearing Loss Prevention Section, Cincinnati, OH 45226-1998)

Increasing evidence suggests that noise-induced hearing loss may be prevented with antioxidant therapy. In order to appreciate treatment constraints, a biochemical marker(s) of reactive oxygen species (ROS)-induced damage is necessary. Without a marker, the timing of damage and biologically effective exposure(s) cannot be understood fully. This study examined the time course of ROS damage following noise exposure resulting in permanent threshold shift in a rat. Cochlea, brain, liver, serum and urine were analyzed. Oxidative DNA damage was assessed by measuring 8-hydroxy-2′-deoxyguanosine (8OHdG) by high performance liquid chromatography with electrochemical detection (HPLC-EC). Lipid peroxidation was measured via the thiobarbituric acid reactive substance's (TBARS's) colormetric assay for detection of aldehydes (e.g., malondialdehyde). Auditory brainstem response and distortion product otoacoustic emissions thresholds showed progressive elevation up to 3–8-h post-exposure, then notable recovery at 72 h, and some worsening at 672 h. 8OHdG/dG was significantly elevated in cochlea at 8-h post-exposure, and in the brain and liver at 72 h. TBARS were significantly elevated in serum at 72 h. This is the first evidence that oxidative DNA damage is present in cochlea following intense noise, and that the timing of damage corresponds to the timing of functional impairment.

2pPP3. Temporary threshold shift in ABRs and DPOAEs following noise exposure in Long–Evans rats. William J. Murphy and Luann E. van Campen (NIOSH, Hearing Loss Prevention Sect., 4676 Columbia Pkwy., Cincinnati, OH 45226-1998)

In conjunction with a study of reactive oxygen species induced damage in Long–Evans rats [van Campen *et al.*, ARO abstracts (2001)], the progressions of temporary shifts in auditory brainstem response thresholds and distortion product otoacoustic emission input/output functions were analyzed. Thirty-five rats were exposed to 2 h of 120-dB SPL band-limited noise (7.5–15 kHz). ABR and DPOAE measurements were collected at 1-, 3-, 8-, 72-, and 672-h post-exposure for seven rats per condition. One sham-exposed rat was included in each exposure group. Thresholds for ABR elicited by tone bursts (2, 4, 8, 16, 32 kHz) exhibited the greatest threshold shifts for the 8-, 16- and 32-kHz stimuli. Recovery following noise exposure was greatest for the 72-h group with a lesser recovery for the 672-h group. The DPOAE thresholds (level of the $f2$ primary stimulus where LDPOAE was 5 dB above the noise floor) exhibited no significant difference for the 72- and 672-h groups. However, the input/output functions for the 72 group exhibited more DPOAE energy than the 672-h group.

2pPP4. Measurement and interpretation of stimulus frequency otoacoustic emissions: Input/output functions and transient analyses. Douglas H. Keefe, Kim S. Schairer, Fei Zhao, and Jeffrey L. Simmons (Boys Town Natl. Res. Hospital, 555 N. 30th St., Omaha, NE 68131, keefe@boystown.org)

A stimulus frequency otoacoustic emission (SFOAE) is a cochlear-based signal measured in the ear canal in response to sine-tone stimulation. SFOAE responses are potentially related to other auditory-system responses to sine-tone stimulation, e.g., basilar membrane input/output (I/O) functions, and behavioral studies of nonlinearity. In preliminary experiments, system distortion has been measured in an ear simulator using SFOAE and DPOAE paradigms, and different probes. SFOAE I/O functions have been measured in subjects using the double-evoked method [D. H. Keefe, J. Acoust. Soc. Am. **103**, 3489–3498 (1998)] in two approaches: equal-level stimuli in each loudspeaker, or a fixed high level in one loudspeaker and varying level in the other. The latter approach is particularly helpful, because it controls for mutual suppression effects in SFOAE responses. Based on the use of transient-gated stimuli, time-frequency representations of the OAE responses enable measurement of cochlear delays and detection of possible multiple internal reflections. Re-

sults will be described in terms of the reflection emission theory of OAEs [Zweig and Shera, J. Acoust. Soc. Am. **98**, 2018–2047 (1995)] extrapolated to moderate stimulus levels, in which SFOAE nonlinearities are directly related to basilar-membrane nonlinearities. [Work supported by NIH (R01 DC003784, T32 DC00013).]

2pPP5. Transient otoacoustic emissions and frequency selectivity. Julie Hazelbaker and Lawrence Feth (Dept. of Speech and Hearing Sci., The Ohio State Univ., Columbus, OH 43210)

The relationship between transient evoked otoacoustic emissions (TEOAE) and frequency selectivity in the auditory system was examined. Three subjects with normal hearing sensitivity and TEOAE spectra with prominent peaks and valleys participated. Psychophysical tuning curves were measured at two pairs of frequencies: one where a robust emission occurs, and the other for an emission with a signal-to-noise ratio at least 6 dB SPL lower than the first. A forward masking paradigm with a 2AFC, 3 up – 1 down adaptive procedure was used to construct three point tuning curves. Quiet threshold for a given probe frequency was established first. The tip of the tuning curve was then obtained by determining the masker level for $P(C)=79.1$ with a probe at 10 dB SL. Finally, the Q_10 frequencies were acquired by adapting the frequency of the masker, which was set 10 dB above the tuning curve tip [Spetner and Olsho, 1990]. Preliminary results indicate that the Q_10 values for the tuning curves from the TEOAE peak frequencies are consistently greater than those of the tuning curves from valley frequencies. These results suggest that a relationship exists between frequency selectivity in the auditory system and the relative minima and maxima of TEOAE patterns. [Research supported by a grant from The Ohio State University College of Social and Behavioral Science.]

2pPP6. Wideband reflectance measurement of the contralateral acoustic reflex threshold for broadband noise and 1000-Hz tones. M. Patrick Feeney and Lindsay Marryott (Speech and Hearing Sci., The Ohio State Univ., 110 Pressey Hall, 1070 Carmack Rd., Columbus, OH 43210)

Recent work has suggested that wideband reflectance measures may provide a sensitive estimate of the acoustic reflex threshold [M. P. Feeney and D. H. Keefe, J. Speech, Lang. Hear. Res. **42**, 1029–1041 (1999)]. This study examined contralateral reflex thresholds for broadband noise and 1000-Hz tones in 21 young adults. The acoustic reflex was detected with the reflectance system by assessing the correlation between pairs of wideband reflectance shifts for a given activator level. Reflex threshold estimates obtained with this method were compared with those obtained with a clinical admittance system. Reflex thresholds for the noise activator were around 10 dB lower for reflectance ($M=77.4$-dB SPL, Zwislocki coupler) than for the clinical measure [$M=87.8$-dB SPL, $t\ (20)=6.2$, $p<0.001$]. For the 1000-Hz activator, reflectance reflex thresholds averaged 5 dB lower ($M=95.3$-dB SPL) than for the clinical system [$M=100.2$-dB SPL, $t\ (20)=8.2$, $p<0.001$]. Moreover, the tone-noise reflex threshold difference was significantly larger for reflectance ($M=17.9$-dB SPL) than for the clinical measure [$M=12.4$-dB SPL, $t\ (20)=6.2$, $p<0.001$]. These results support the use of wideband reflectance as a sensitive measure of the acoustic reflex threshold. [Supported by NIH-R03DC04129.]

2pPP7. An auditory-nerve model with an instantaneous frequency glide and its application to modeling psychophysical results. Qing Tan and Laurel Carney (Dept. of Biomed. Eng., Hearing Res. Ctr., Boston Univ., Boston, MA 02215)

A computational model was developed to simulate the responses of auditory nerve (AN) fibers. The model consists of a time-varying band-pass filter with a nonlinear feed-forward control path, which changes the bandwidth and gain of the signal path. This model produces realistic response features to various stimuli, including pure tones, two-tone combinations, wide-band noise and clicks. The parameters of the band-pass filter were estimated by fitting the model revcor functions to revcor functions from experimental data of cat. The Marquardt method was used to minimize the difference between the model revcor function and the cat revcor function at various characteristic frequencies. Instantaneous frequency (IF) glides in the reverse correlation function of the models response to broadband noise were achieved by simple configuration of the locations of the poles and zeros in the band-pass filter. The locations of the poles are continuously varied by the control signal to change the gain and bandwidth of the signal path, without affecting the IF profile, which is level independent. Other important properties, such as nonlinear compression, two-tone suppression and reasonable Q10 values are also included. Applications of this model for studying AN responses to speechlike signals will be discussed.

2pPP8. Encoding of sinusoidal and natural speech tokens in the auditory nerve. Jeremy Loebach and Robert Wickesberg (Dept. of Psych., Univ. of Illinois, 603 E. Daniel St., Champaign, IL 61820)

Sinewave speech is a synthetic variation of natural speech that replaces natural formants with time varying sinusoidal waves. This synthesis preserves the general formant motion of the natural utterance while removing many spectral details; yet sinewave speech tokens can convey a linguistic message [Remez *et al.*, Science **212**, 947–950 (1981)]. This study examined the extent that phonetic information in sinusoidal speech is preserved at the level of the auditory periphery. We compared the responses of individual auditory nerve fibers in ketamine anesthetized chinchillas to the natural and sinusoidal tokens of the word /bal/. Comparisons of the first 100 milliseconds of the responses revealed high correlations over the first 20 milliseconds of the responses representing the initial consonant ($r=0.82$) despite low correlations between the stimuli themselves ($r=0.18$). Correlation coefficients decreased in the vowel portion of the responses ($r=0.30$). Spectral reduction in the sinusoidal token may produce the low correlation between the neural representations of the vowel. The preservation of temporal cues may explain the high correlation between the neural representations of the consonant despite spectral differences.

2pPP9. Consistency of the auditory nerve response to normal and whispered speech. Stephanie Justin and Robert Wickesberg (Dept. of Psych., Univ. of Illinois, 603 E. Daniel St., Champaign, IL 61820)

Multiple presentations of an acoustic stimulus are often used to study the encoding of sounds that in the natural environment are not repeated. This study examined how consistent the response of the auditory nerve is with respect to the number of repetitions. Naturally voiced and whispered forms of /ta/ were presented 50 times at 60 and 90 dB to ketamine anesthetized chinchillas. For each stimulus, global average peri-stimulus time (GAPST) histograms for the first 20 trials were compared to those for the last 20 trials. Correlation coefficients for the normal syllable were 0.87 and 0.72 during the consonant and vowel, respectively. In the whispered condition, correlation coefficients were 0.95 during the consonant, but only 0.44 during the vowel. Similar correlations were obtained with comparisons of GAPSTs for either the first or last 20 trials to those for all 50 trials. Correlations obtained from the GAPSTs were higher than those of individual responses for both speech conditions and at all intensities. Individual auditory nerve fiber correlations during the speech stimuli were variable. These results demonstrate that the use of many repetitions to achieve consistency in the response may not be necessary when one studies the ensemble response of the auditory nerve.

2pPP10. Stimulation rate influences frequency tuning characteristics of inferior colliculus neurons. Jeremy Smalling, Alexander Galazyuk, and Albert Feng (Univ. of Illinois, Urbana, IL 61801)

Sounds in real world environments such as animal calls and human speech are complex and often occur in rapid succession. In light of this, it is important to gain an understanding of how the rate of acoustic stimulation influences the units' basic response properties. Previous studies of frequency tuning were generally based on investigating neuronal responses to tonal stimuli presented in isolation, and the results derived therefrom were assumed to accurately portray the cell's response range. However, the firing history of an auditory neuron has been shown to shape its response to subsequent sounds. The goal of the present study was to investigate the frequency tuning characteristics of neurons in the inferior colliculus (IC) of the little brown bat to tone pulses presented at various rates. Eighty-five percent of the IC neurons studied showed rate-dependent changes in their frequency selectivity. Half of these neurons exhibited narrowing of their frequency response range at higher rates. These results indicate that the response properties of central auditory neurons at low stimulation rates do not necessarily reflect the units' response properties to sounds presented at high and behaviorally relevant rates. The possible cellular mechanisms and behavioral importance of this effect are discussed.

2pPP11. Noninvasive recording of fetal auditory evoked responses using a new 151 channel sensor array (SARA). Hari Eswaran, James Wilson (Dept. of Appl. Sci., Univ. of Arkansas at Little Rock, Little Rock, AR 72204), Hubert Preissl, Pamela Murphy, and Curtis Lowery (Univ. of Arkansas for Medical Sci., Little Rock, AR 72205)

The goal was to successfully record fetal auditory evoked responses (AERs) using the newly developed noninvasive fetal magnetoencephalographic (MEG) system called SARA. 131 AER measurements were performed on 31 fetuses starting at 27 weeks. The SARA system, equipped with an array of 151 SQUID (superconducting quantum interference device) sensors spread over an area of 1300 cm^2, is curved to fit the shape of the pregnant abdomen. The auditory stimulus was—1 kHz tone; duration—100 ms; rate—0.5 tones/s.; intensity (external)—100-dB SPL in air. The data were collected at a sampling rate of 312.5 Hz. The AERs were averaged over a period of 700-ms window after removing interfering maternal and fetal heart signals. Control recordings were performed on the same fetus with the same parameters but without any sound stimulation. The peak of the evoked responses was around 250 ms with amplitudes of about 20 fT. No evident evoked response was seen in the control recordings on the same fetus. In summary, it is feasible to noninvasively acquire MEG auditory evoked responses from the fetus using SARA. AER can be a useful tool for the neurological assessment of fetal well-being throughout gestation. [Work funded by NIH.]

2pPP12. A coincidence-detection model for level discrimination of tones in roving-level noise. Xuedong Zhang and Laurel H. Carney (Hearing Res. Ctr. and Dept. of Biomed. Eng., Boston Univ., Boston, MA 02215)

This study describes a monaural cross-frequency coincidence-detection mechanism to model level discrimination performance for a tone in the presence of noise in a roving-level paradigm. Model coincidence detectors receive population input from model auditory-nerve (AN) fibers from the same ear with different characteristic frequencies (CFs). The response of each model coincidence detector is sensitive to both the rate and the phase of its AN fiber inputs. A set of model coincidence detectors was constructed that received input from a population of model AN fibers. To detect intensity increments of tones in wideband noise, the most sensitive coincidence detector was one that received the inputs from two AN fibers with tone responses that were opposite in phase (phase opponency); this mechanism was robust in the presence of roving-level maskers. For narrow-band noise, detectors that were sensitive to changes in rate also contributed information to the detection task; this rate-based mechanism was affected by a roving-level paradigm. The difference in the effect of the rove on model performance with different masker bandwidths was similar to that observed for human subjects [Kidd *et al.*, J. Acoust. Soc. Am. **86**, 1310 (1989)].

2pPP13. A low-frequency paradox in binaural pitch. William M. Hartmann (Phys. and Astron., Michigan State Univ., East Lansing, MI 48824) and Peter Xinya Zhang (Michigan State Univ., East Lansing, MI 48824)

Binaural pitches, such as the Huggins pitch and binaural coherence edge pitch (BICEP), can be approached in terms of a model network of binaural neurons that are tuned in frequency and in interaural delay [e.g., Raatgever and Bilsen, J. Acoust. Soc. Am. **80**, 429–441 (1986)]. The basic rule of binaural combination resembles addition. Binaural pitches appear as peaks or edge features in the frequency-delay plane. The characteristic frequency of a feature determines the pitch, and the internal delay of the feature determines the lateralization of the pitch image. Because delay lines become sparser with increasing delay, the model predicts that binaural pitches of very low frequency should be more difficult to hear the more they are lateralized. However, our Huggins pitch and BICEP discrimination and detection experiments show that the reverse is true. These experiments favor a model in which the basic rule of binaural combination resembles subtraction. However, the predictions of the subtractive model disagree with the observed lateralization. Furthermore, the low-frequency discrimination and detection experiments do not agree with a model in which additive and subtractive binaural combinations are alternative options. This is the low-frequency paradox. [Work supported by NIDCD.]

2pPP14. The binaural processing of double vowels: Experimental measurements and computational models. Michael A. Akeroyd (Dept. of Neurosci., Univ. of Connecticut Health Ctr., Farmington, CT 06030 and Lab. of Exp. Psych., Univ. of Sussex, Brighton, UK)

The degree to which the binaural processing of a ''target'' vowel is affected by the simultaneous presentation of a ''distractor'' vowel (of different ITD to the target) was studied using three pairs of synthetic vowels. The stimuli were an ''er'' (of 100-Hz $F0$) paired together with either an ''ai,'' ''er,'' or ''oo'' (all of 125-Hz $F0$). In Experiment 1 ITD thresholds for the target vowel were measured. The results showed that thresholds were larger in double-vowel pairs than in control, target-alone situations. Moreover, the size of the effect was dependent both on whichever vowel of the pair was the target and on the relative level of the two vowels. Computational modeling of the binaural correlograms of the stimuli showed that these results do not require any explicit segregation of frequency channels for the two vowels. In Experiment 2 the perceived lateralization of the target vowel (when given an ITD of 0 ms) was measured. The results showed that this lateralization was ''pulled'' slightly toward the distractor vowel. Further computational modeling showed that some segregation of the frequency channels in the binaural correlogram, perhaps based on the pitch information in the corresponding autocorrelogram, is required to account for this result.

2pPP15. MLDs for signals placed in masker envelope minima and maxima. Emily Buss, Joseph HallIII, and John Grose (Dept. of Otolaryngol./Head and Neck Surgery, Univ. of North Carolina School of Medicine, 610 Burnett-Womack Bldg., CB# 7070, Chapel Hill, NC 27599)

Previous data on the masking level difference (MLD) have suggested that NoSpi detection for a long-duration signal is dominated by signal energy occurring in masker envelope minima [Hall *et al.*, J. Acoust. Soc. Am. **103**, 2573–2577 (1998); Grose and Hall, *ibid.* **103**, 2590–2594 (1998)]. This finding was expanded upon using a brief (30-ms) 500-Hz tonal signal that coincided with either the envelope maximum or envelope minimum of a 50-Hz wide bandpass Gaussian noise masker centered at

500 Hz. On each interval a masker band was digitally generated in the frequency domain, the envelope maximum (or minimum) was identified, and the sample was rotated until that maximum (or minimum) fell in the temporal center of the sample. Data were collected at a range of masker levels (25, 40, and 55 dB spectrum level) in order to assess the possible contributions of compression. Conditions employing frozen noise samples were included to assess the possible role of masker uncertainty. Results to date suggest substantially larger MLDs for signals coinciding with masker envelope minima as contrasted with envelope maxima. This result is primarily due to decreased NoSpi thresholds for signals coinciding with mimima. [Work supported by NIH NIDCD.]

2pPP16. Effect of reverberation on the masking of narrow-band signals. Richard Freyman, Uma Balakrishnan (Dept. of Commun. Disord., Univ. of Massachusetts, Rm. 3 Arnold House, Amherst, MA 01003), and Patrick Zurek (Sensimetrics Corp., Somerville, MA 02144)

This study explores the factors that influence the masking of narrow-band signals in reverberant sound fields. Rectangular rooms were simulated using the image method with a rigid sphere modeling the head. The signals were third-octave noise bands presented in a continuous broadband noise masker. For a given room simulation, the signals and masker were convolved with impulse responses obtained in the room at multiple azimuth angles and distances, for subsequent presentation via headphones. Adaptive forced-choice procedures were used with the resulting signals to find monaural and binaural thresholds in listeners with normal hearing. The benefit of separating the masker and signal spatially in the simulation was reduced by reverberation, as expected, although the effects were complex. Among the predictions confirmed by these experiments is that reverberation actually improves thresholds when the masker is closer to the listener than the signal, apparently due to both an increase in monaural signal power a decrease in interaural signal correlation. With masked thresholds obtained over the frequency range of the articulation index, the results of these experiments will allow quantitative predictions of speech intelligibility as a function of room charcteristics and speech and noise source locations. [Work supported by NIH.]

2pPP17. Short-term adaptation to novel combinations of acoustic spatial cues. Timothy Streeter, Barbara G. Shinn-Cunningham, and Andrew Brughera (Hearing Res. Ctr., Dept of Cognit. and Neural Systems, Boston Univ., 677 Beacon St., Boston, MA 02215, timstr@cns.bu.edu)

Previous work demonstrates that a listener's interpretation of spatial cues can change due to long- or short-term training. Results suggest that short-term training does not alter how spatial locations are computed, but how spatial percepts are interpreted. This series of experiments examines how subjects adapt to different rearrangements of auditory cues. In Experiment I, subjects were trained to identify the spatial cues normally associated with exocentric location x as coming from $2x$ (a pure linear remapping of space). In Experiment II, ITDs were remapped in the same way, but were paired with normal IID and spectral cues (i.e., the ITD normally associated with azimuth $2x$ was paired with IID and spectral cues associated with location x). Results examine how spatial resolution and bias are affected by unusual mappings between spatial cues and exocentric location as well as by combining spatial cues that normally do not correspond to one source location. These experiments test the hypothesis that listeners rapidly and subconsciously recalibrate a ''gain'' describing how to map an internal representation of auditory space to exocentric space, but cannot rapidly alter how the source positions in this internal representation are computed. [Work supported by a grant from the Whitaker Foundation.]

2pPP18. The independence of adaptation to interaural time and level changes. Michael F. Neelon, Carrie M. Benedon, and Rick L. Jenison (Dept. of Psych., Univ. of Wisconsin, Madison, WI 53706)

Changes in interaural time and interlevel differences (ITDs and ILDs) for a sound moving in the free field normally imply movement in the same direction. Research has shown that listeners can be adapted to and show aftereffects for each of these cues separately [W. H. Ehrenstein, Perception **23**, 1249–1255 (1994)]. Psychophysical evidence of ILD to ITD conversion, however, raises the question whether adaptation is independent between these two channels. The present experiment explores their possible independence in order to better specify where in the auditory pathway their adaptation occurs. Listeners will be adapted to sounds with dynamically varying ITD and ILD cues alone, in concert or in conflict with each other (the latter conditions implying either the same or opposite direction of movement). ITD and ILD adaptation will be measured separately via test sounds changing in either of these cues alone. If aftereffects for one cue are unaffected by the implied movement direction of the other, then adaptation most likely lies early in the auditory pathway (e.g., the olivary bodies). However, if aftereffects show an influence of one cue on the other, some later physiological site of convergence may be involved in adaptation (e.g., the inferior colliculus).

2pPP19. The role of the precedence effect in sound source lateralization. Daniel E. Shub (Speech and Hearing Sci. Prog., Harvard–MIT Div. of Health, Sci. and Technol.), Robert H. Gilkey (Wright State Univ.), and H. Steven Colburn (Boston Univ.)

This work examines sound source lateralization of ongoing broadband noise targets in the presence of a single ongoing broadband noise jammer, which was either a simple simulated reflection of the target or was uncorrelated with the target. Identification thresholds of ± 300-μs ITD targets and lateralization thresholds were determined. Identification thresholds of the reflection jammers were 3.2 dB lower ($P = 0.027$) than the identification thresholds of the jammers that were uncorrelated with the targets. Lateralization thresholds of the reflection jammers were 2.85 dB lower ($P = 0.040$) than the lateralization thresholds of the uncorrelated jammers. There was a slight trend, in agreement with the current understanding of the precedence effect, for both the identification and lateralization thresholds of the simple reflection jammers to be dependent on jammer ITD. The 0-μs ITD reflection jammer produced a threshold that was 1.9 dB lower ($P = 0.104$) than the threshold of the 643-μs ITD reflection jammer. Additionally comparisons between the lateralization thresholds of normal-hearing listeners and two cross-correlation models were made. The models obtained lateralization thresholds as low as -13 dB, up to 10 dB better than normal-hearing performance. [Work supported by NIH Grant No. R01 DC00100.]

2pPP20. The minimum audible facing angle. John G. Neuhoff, Mary-Alice Rodstrom, and Tanaya Vaidya (Dept. of Psych., The College of Wooster, Wooster, OH 44691, jneuhoff@wooster.edu)

Many sound sources have a directional characteristic such that they project sound from one planar orientation at a time (e.g., talkers and loudspeakers). Yet, almost all localization studies employ a unidirectionally facing source aimed at the listener. Although there are good reasons for holding the facing direction of a loudspeaker constant relative to the listener when investigating localization, it is curious that the acoustic facing direction itself has received little or no attention. Here, blindfolded listeners heard pulsed noise from a loudspeaker facing a reference point of either 0 deg (directly facing the listener) or 30 deg. The loudspeaker was then rotated clockwise or counterclockwise to one of six facing angles (5, 10, 15, 20, 25, or 30 deg), and the stimulus was replayed. The listeners indicated perceived direction of rotation. The minimum threshold for detection, or in this case what is termed the minimum audible facing angle (mAFA), was 15 deg for the 0 deg reference point and 21 deg for the 30

deg reference point. Within the limitations of the current experimental conditions, listeners showed appreciable sensitivity to the facing angle of a unidirectionally facing sound source. Our results also show evidence for a maximum auditory facing angle (MAFA).

2pPP21. Effect of auditory cuing on azimuthal localization accuracy. Norbert Kopčo (Hearing Res. Ctr., Dept. of Cognit. and Neural Systems, Boston Univ., 677 Beacon St., Boston, MA 02215, kopco@cns.bu.edu), Albert Ler, and Barbara G. Shinn-Cunningham (Boston Univ., Boston, MA 02215)

Auditory localization in the horizontal plane was measured following the presentation of a cue in order to explore whether attentional focus could improve localization accuracy. Subjects pointed to the heard location of a broadband target source that was presented (at a delay of either 50 or 300 ms) after a cue source. In half of the blocks of trials, the cue source came from the same (left/right) hemifield as the target on most (75%) of the trials, and thus (on average) provided the listener with information about the target location. In the other half of the blocks of trials, the cue source location was equally likely to come from either the same or the opposite hemifield and provided no information to the subject regarding target position. The presence of a cue biased localization performance in both conditions rather than improving accuracy when the cue provided information about the target laterality, even for a delay of 300 ms between the cue and target. Results suggest that auditory cuing, which has been shown to decrease response times, degrades localization accuracy. [Work supported by a grant from the Air Force Office of Scientific Research.]

2pPP22. Localization and speech-identification ability of hearing-impaired listeners using phase-preserving amplification. Ward Drennan, Stuart Gatehouse, Patrick Howell (MRC Inst. of Hearing Res., Glasgow Royal Infirmary, Queen Elizabeth Bldg., Glasgow G31 2ER, UK), Dianne VanTasell, and Steven Lund (Starkey Labs., Eden Prairie, MN 55344)

Hearing-impaired listeners experience increased difficulties recognizing speech and localizing sounds in adverse environments. This study investigated the benefits of signal processing in binaural hearing aids designed to preserve cues that accompany spatial location. The ability of listeners to localize click-trains in noise was tested, along with their ability *both* to localize *and* to identify words in noise (a dual task). Listeners experienced two types of bilateral hearing aid fittings: (1) a custom fitting that provided appropriate gain while also matching the phase measured near the tympanic membranes without the hearing aids, and (2) a conventional fitting (using the same hearing aid device) that provided the same gain with noncustom, linear phase. Testing occurred for each fitting immediately and following 3-months listening experience using a within-listener, within-device, randomized, single blind, crossover design. A rating-scale questionnaire was administered to assess perceived speech-hearing and spatial abilities. In the dual task, the listeners exhibited superior localization ability for the phase-preserving fitting initially and after 3-months experience. This advantage did not occur for the click-train localization task. Listeners rated their spatial abilities higher with the phase-preserving fitting, although little improvement was observed or reported for speech hearing.

2pPP23. Localization of sound by binaural cochlear implant users. John P. Preece, Richard S. Tyler, Jay T. Rubinstein, Bruce J. Gantz (Dept. of Otolaryngol., Univ. of Iowa, Iowa City, IA 52242), and Richard J. M. van Hoesel (CRC for Cochlear Implant and Hearing Aid Innovation, E. Melbourne 3002, Australia)

We examined the localization ability in five adult patients who were implanted bilaterally with the CI24M implant from Cochlear Corporation. These patients demonstrated a difference in either length of time deaf before implantation, preimplant thresholds, or both. Patients were tested in an anechoic room. Signals were four 200-ms bursts of broadband noise separated by 55 ms of silence. Stimuli were randomly presented from one of eight loudspeakers arrayed in an arc at ear level in front of the patient. The speakers were separated by 15 deg azimuth. The patient was seated 1.5 m from the speakers and responded orally with a speaker number. The level of individual stimuli was varied randomly over an 8-dB range with an average level of presentation of 65-dB SPL measured at the approximate location of the center of the patient's head. Patients were tested with each ear separately and with both ears together. The results show a very good ability in all five patients to localize sounds with two cochlear implants. The monaural abilities varied considerably across patients, and often between ears for each patient, but were always worse than the binaural abilities. [Work supported by NIDC and CRC.]

2pPP24. Auditory motion aftereffects with varying interaural phase difference. Takayuki Kawashima and Takao Sato (Dept. of Psych., Univ. of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan, ll97031@mail.ecc.u-tokyo.ac.jp)

It has been known that the auditory motion aftereffect (MAE) is spatially specific [see, for example, D. W. Grantham, Acoustica **84**, 337–347 (1998), exp. 3]. However, it still is not very clear which cues for sound localization are responsible for this spatial specificity of the MAE, since several possible cues, such as the spectral profile or interaural level difference, covaried in most past studies. In this study, MAE and its spatial specificity were investigated with sound motion created by using only interaural phase difference (IPD) as a first step to identify the responsible cues. We used the probe method with the method of constant stimuli. Either with or without adaptation to a moving tone, subjects were asked to judge the direction of probe tone motion (0.7 ms duration, either to the left or the right). The slope and the position of the psychometric functions were affected by the direction of the adapter's motion direction, but only when IPD ranges (spatial ranges of motion) of the adapter and probe tones overlapped each other. These results suggest that the change in IPD is at least one of the cues which produce the spatial specificity.

2pPP25. Minimum dynamic lateralization for multiple moving sources. Michael F. Neelon and Rick L. Jenison (Dept. of Psych., Univ. of Wisconsin, Madison, WI 53706)

It is still unresolved whether auditory motion is perceived via specialized motion detectors or inferred from static samples of changes in spatial position. This question has been investigated by comparing minimum audible movement angles for single sources moving discretely or continuously [Perrott and Marlborough, J. Acoust. Soc. Am. **85**, 1773–1775 (1989)]. But using one source does not force the listener to determine movement by only one of the aforementioned processes. To better measure sensitivity to pure auditory dynamics, the following study simulates multiple overlapping sources moving in the same direction across random sections of the auditory hemifield. Experimental stimuli are composed of different portions of the trajectories of circling resonant sources, which are individually created by dynamically varying interaural time and level differences. The multiple, variable endpoints in the composite stimulus should inhibit the listener from relying solely on such cues to determine movement direction. Pilot studies using from 1–5 concurrent sources show dynamic lateralization thresholds are lowest for one moving source. This implies movement is best perceived when stimulus endpoints can be sampled. However, thresholds across multiple sources do not significantly differ regardless of number, which may represent a measure of sensitivity to pure auditory dynamics.

2pPP26. Finite-difference computation of head-related transfer functions for human hearing. Tian Xiao and Qinghuo Liu (Dept. of Elec. Eng., Duke Univ., Durham, NC 27708-0291)

Modeling the head-related transfer function (HRTF) is a key to many applications in spatial audio. To understand and predict the effects of head geometry and the surrounding environment on the HRTF, a three-dimensional finite-difference model has been developed for the head-related impulse response. A perfectly matched layer (PML) is used to absorb outgoing waves at the truncated boundary of an unbounded medium. An external source is utilized to decrease the computational domain size. This numerical model has been validated by analytical solutions for a spherical head model. The FD code is then used as a computational tool to predict the HRTF for various scenarios. In particular, a simplified spherical head model is compared to a realistic head model up to about 7 kHz. The HRTF is also computed for a realistic head model in the presence of a wall. It is demonstrated that this FD model can be a useful tool for spatial audio applications.

TUESDAY AFTERNOON, 5 JUNE 2001 SALON V, 1:15 TO 5:00 P.M.

Session 2pSA

Structural Acoustics and Vibration: Scattering, Plates and Shells

Sean F. Wu, Chair
Department of Mechanical Engineering, Wayne State University, 505 Anthony Wayne Drive, Detroit, Michigan 48202

Contributed Papers

1:15

2pSA1. Interaction of Lamb wave with a welded rib. Bruno Morvan, Jean Duclos, Hugues Duflo, and Alain Tinel (Laboratoire d'Acoustique Ultrasonore et d'Electronique (LAUE), UMR 6068 CNRS, Pl. R. Schuman, BP 4006, 76 610 Le Havre, France)

This work deals with Lamb wave conversion on a rib stiffened plate. Some previous work has studied the scattering of plane waves at a *T* junction at low frequency (FE$<$200 kHz mm) for which the wavelengths are higher than the plate thickness. In this study, the conversion phenomenon is investigated in the case of an *A*1 incident Lamb wave (FE=4 MHz mm). The experimentation is done with a stainless steel *T* structure, and the plate and rib are 2 mm thick. The normal surface displacement is related theoretically to the power flow. Then, the energy of the converted waves is obtained from measurement of the plate normal displacements by means of a laser interferometer. Next, we compute a finite-element simulation on two various meshes: the plates are connected perpendicular to each other either with or without a weld. The comparison between experimental and numerical methods shows the importance of the weld shape on the wave transmission to the rib. The weld funnels the incident wave and improves the energy transmission on the rib. Moreover, the experimental and MEF results are in a good agreement.

1:30

2pSA2. Homogenization of a ribbed cylindrical shell over the full frequency range for efficient calculation of structural-acoustic behavior. Linda Franzoni and Donald Bliss (Mech. Eng. and Mater. Sci., Duke Univ., Durham, NC 27708-0300)

A homogenization method for complex structures, applicable for all frequencies, is described for a cylindrical shell with periodically spaced ribs. This approach has application to naval and aerospace structures. The homogenization method utilizes a local–global decomposition facilitated by adding and subtracting canceling smooth forces. The smooth global problem has an infinite-order structural operator, and periodic discontinuities are replaced by equivalent distributed suspension terms. The global problem can be solved very efficiently since all rapidly varying scales have been removed. The local problem, which provides transfer function information for the global problem, is solved separately and independently, except for amplitude information from the global problem. Once formulated, the self-contained global problem is solved first, and the local solution can be reconstructed afterwards. In the present work, the ribs are modeled as annular plates periodically attached to the cylindrical shell. The coupled shell equations are re-cast using the method of local–global homogenization, and are solved efficiently using a combination of symbolic manipulation and numerical methods. Comparisons withother solution methods are presented, and the extension to fluid loading is discussed. [Work sponsored by ONR.]

1:45

2pSA3. Improvement of the solution convergence rate for an elastic shell driven by a vibratory constraint using analytical-numerical matching. Christopher Park, Linda Franzoni, and Donald Bliss (Mech. Eng. and Mater. Sci., Duke Univ., Durham, NC 27708-0300)

A three-dimensional elastic shell is driven by the motion of an attached support and the solution is constructed using the method of analytical-numerical matching (ANM) with finite-element analysis and the structural-acoustic scattering code SONAX. First, the area around the structural constraint is resolved in great detail as a separate local three-dimensional FEA solid model. Second, a Love–Timoshenko shell-type model of the local region is solved analytically for a set of smooth forces that will be applied to the full three-dimensional shell in place of the constraint. Third, the model of a full three-dimensional shell with end caps is analyzed in SONAX, with the smooth forces applied on a mode-by-mode basis. The complete solution to the original problem is the superposition of the solutions of each of the subproblems described. Note that the high modal content of the original problem is contained in the high-resolution separate local problem (small domain), leaving only low modal content in the global (large domain) problem that is solved using SONAX. This sample problem demonstrates the ability of ANM to enhance the capabilities of SONAX and to improve the convergence rate (and overall computation time) for problems of this type. [Work sponsored by ONR.]

2:00

2pSA4. Circumferential-wave phase velocities for empty, fluid-immersed spherical metal shells. Herbert Überall (Phys. Dept., Catholic Univ. of America, Washington, DC 20064), A. Claude Ahyi, P. K. Raju (Auburn Univ., Auburn, AL 36849-5341), I. K. Bjørnø, and L. Bjørnø (Tech. Univ. of Denmark, DK-2800 Lyngby, Denmark)

Our earlier studies regarding acoustic scattering resonances and the dispersive phase velocities of the surface waves that generate them, have demonstrated the effectiveness of obtaining phase velocity dispersion curves from the known acoustic resonance frequencies, and their accuracy. This possibility is offered by the picture of phase matching after each complete circumnavigation of these waves, which leads to close agreement of resonance results with those calculated from 3-D-elasticity theory

whenever the latter are available. The present investigation is based on the mentioned resonance frequency/elasticity-theory connection, and we obtain comparative dispersion-curve results for water-loaded, evacuated spherical shells of various metals. In particular, the characteristic upturn of the dispersion curves of low-order shell-borne circumferential waves (A or A_0 waves) which takes place on spherical shells when the frequency tends towards low values, is demonstrated here for all the metals under consideration.

2:15–2:30 Break

2:30

2pSA5. Visualization of the energy flow for a guided forward wave in and around a fluid loaded elastic cylindrical shell. Cleon E. Dean (Phys. Dept., P.O. Box 8031, Georgia Southern Univ., Statesboro, GA 30460-8031, cdean@gasou.edu) and James P. Braselton (Georgia Southern Univ., P.O. Box 8093, Statesboro, GA 30460-8093)

Previous work [Cleon E. Dean and James P. Braselton, J. Acoust. Soc. Am. **107**, 2921 (2000)] showed the energy flow for a fluid loaded elastic cylindrical shell at the resonance frequency. The results were equivocal since there are two counterpropagating guided waves. The current work uses a clever analysis due to Kaduchak and Marston [Gregory Kaduchak and Philip L. Marston, J. Acoust. Soc. Am. **98**, 3501–3507 (2000)] to separate the two waves and show the energy flux associated with just one forward propagating surface guided wave.

2:45

2pSA6. Scattering and radiation from a fluid-loaded coated cylindrical shell. J. M. Cuschieri and C. Treffot (Ctr. for Acoust. and Vib., Dept. of Ocean Eng., Florida Atlantic Univ., Boca Raton, FL 33431)

A common question when dealing with fluid-loaded coated structures is how well a normally reacting coating layer models the radiation and scattering from a fluid-loaded coated cylindrical. Modeling with a normally reacting layer simplifies the model which therefore allows more complex conditions to be analyzed. To explore this question analytic solutions are obtained for the scattering from a fluid-loaded cylindrical shell with a compliant coating using both multilayer shell theory [L. Flax and W. G. Neubauer, J. Acoust. Soc. Am. **61**, 307 (1977)], and a normally reacting coating solution [J. M. Cuschieri and D. Feit, J. Acoust. Soc. Am. **107**, 3196 (2000)]. The similarities and differences in the scattering results that are obtained using either of these two approaches are presented for different coating thicknesses and characteristics.

3:00

2pSA7. Cepstrum of the scattering response of a cylinder. Angie Sarkissian and Louis R. Dragonette (Naval Res. Lab., Washington, DC 20375-5350, angie@aqua.nrl.navy.mil)

The cepstrum of the scattered field of a ribbed, hemispherically end-capped cylindrical shell is computed for various incident and scattered angles, for $2 \leqslant ka \leqslant 30$, where k is the wave number and a is the radius of the cylinder. Highlights observed in the cepstrum domain are related to various scattering mechanisms. The cepstrum representation is compared to the frequency domain and the time domain response of the cylinder. Benefits and disadvantages of the cepstrum representation are discussed. [Work supported by ONR.]

3:15

2pSA8. Acoustic scattering from a fluid-loaded membrane with periodic discontinuities using local–global homogenization (LGH). Donald Bliss, Linda Franzoni, and Pavel Danilov (Mech. Eng. and Mater. Sci., Duke Univ., Durham, NC 27708-0300)

A homogenization method for complex structures, applicable over the entire frequency range, is being developed. The method utilizes a local–global decomposition facilitated by adding and subtracting canceling smooth forces. The low-wave-number global problem has an infinite-order structural operator, and periodic discontinuities are replaced by an equivalent distributed suspension. The rapidly varying local problem, which provides transfer function information for the global problem, is solved separately and independently, except for amplitude information from the global problem. Once formulated for a specific structure, the self-contained global problem is solved first, and the local solution can be reconstructed afterwards. The LGH reformulation allows the global problem to be solved at much lower resolution than the length of flexural waves on the original structure. The application of this approach to scattering from a fluid-loaded membrane is described. The effects of fluid radiation are transferred entirely to the smooth global problem, whereas evanescent fluid modes are contained within the local problems. Sample calculations are presented comparing the method with classical solutions. [Work supported by ONR.]

3:30–3:45 Break

3:45

2pSA9. Large coated plates under random excitation: Spatial-domain insights into radiation efficiency. R. Martinez and J. M. Garrelick (Cambridge Acoustics Assoc./Anteon Corp., 84 Sherman St., 3rd Level, Cambridge, MA 02140, rmartinez@caa.atinc.com)

A formal spatial-domain analysis of a large plate's cross-spectral response to TBL-like excitations yields the following two results, unified and in closed form: (1) a part that contains a moderate area effect (from the acoustic field's point of view, despite the plate being below coincidence by choice) and a stronger edge effect; and (2) a second part with a weak edge effect and a potentially strong area effect. The strong-edge/moderate-area part contains as coefficient the inverse of the plate's effective loss factor. The strong-area part does not, enabling it to dominate the former with rising frequency particularly over that range in the frequency spectrum encompassing the typically broad (lossy) thickness resonances of the plate's coating. Use of this unified cross-spectral response function to compute the plate's acoustic field confirms the expected rise in radiation efficiency at higher loss factors and frequencies, as brought about by the relative growth of the strong-area effect under both of those conditions. [Work supported by NSWC/Carderock Division.]

4:00

2pSA10. Reconstruction of transient sound radiation from impulsively accelerated objects. Manjit Singh and Sean Wu (Dept. of Mech. Eng., Wayne State Univ., 5050 Anthony Wayne Dr., Detroit, MI 48202)

This paper presents reconstruction of transient acoustic radiation from impulsively accelerated objects using the HELS method [Wu, J. Acoust. Soc. Am. **107**, 2511–2522 (2000)]. The radiated acoustic pressure is expanded in terms of the spherical wave functions and spherical harmonics in the frequency domain. The coefficients associated with these expansion functions are determined by matching the assumed solution to the measured acoustic pressures. The errors incurred in this process are minimized by the least squares. Once the frequency-domain solution is obtained, the transient acoustic pressure signal is reconstructed by taking an inverse Fourier transformation via the reside theorem. The objects considered include an explosively expanding sphere, impulsively accelerated rigid sphere, and impulsively accelerated baffled sphere. The acoustic pressure signals thus obtained are compared with the analytic solutions. It is shown that this methodology can be extended to a nonspherical object subject to an arbitrarily time-dependent excitation. The resulting transient acoustic pressure can be reconstructed by a convolution integral of the impulse response function to the time history of the measured acoustic pressure signals. [Work supported by NSF.]

4:15

2pSA11. The onset of the flexural mode in fluid loaded elastic shells revisited, and its value in object identification in both the frequency and time domain. H. Uberall (Dept. of Phys., The Catholic Univ. of America, Washington, DC 20064) and M. F. Werby (Stennis Space Center, MS 39529)

The onset of the excitation of flexural resonances for fluid loaded evacuated elastic shells produces a striking event. This issue has been interpreted theoretically in 1988 by means of a partial wave decomposition which showed a very narrow-peaked subsonic wave and a broad-peaked weak wave that begins at the speed of sound of the entraining fluid and increases. The narrow peaks are identified with subsonic water-borne waves that resonate in the fluid along the surface of an elastic object, and the broad peaks correspond to the inception of flexural waves. They exist over a small interval in wave number at the point when the flexural wave begins to couple with the fluid. Here we examine the pulse solution. With increasing frequency the partial waves change phase (a 90 degree phase change) and the overlapping flexural waves transition from a partially coherent constructive signal to one that is partially destructive. This leads to an envelope or hump effect, also called mid-frequency enhancement; it is a function of shell thickness as well as material property. We demonstrate how this effect may be employed to identify a submerged elastic shell in either the pulse or frequency solution.

4:30

2pSA12. The importance of the first and second symmetric Lamb modes in target identification of submerged elastic shells. M. F. Werby (NRL Code 7180, Stennis Space Center, MS 39529 and Dept. of Phys., Catholic Univ. of America, Washington, DC 20064) and H. Uberall (Dept. of Phys., Catholic Univ. of America, Washington, DC 20064)

Resonances excited on any shell of constant thickness are one large resonance at a frequency inversely proportional to the shell thickness. Hence for thin shells it occurs at very high frequency. This effect occurs at the inception of the S1 resonance, which may be shown to be an amplitude-modulated wave at inception. Its critical frequency may be determined by the condition: ka=3.14 VLa/(2 daVW) or ka=3.14 VTa/(daVW) where VW, VT, VL, a, da, and k are the speed of sound in the ambient fluid, transverse and longitudinal velocities in the elastic material, largest dimension of the object, object thickness relative to a and the wave number in the fluid. If 2VT>VL the first condition defines the S1 critical frequency and the second that of the S2. The converse is true otherwise. The S2 resonance is not striking but may be identified as such. Thus, ratios of the two resonances lead to ratios of the bulk velocities and other considerations can isolate the shell thickness. This offers for any shell of constant thickness a means to determine the presence of certain submerged objects. We discuss the reasons for this and illustrate results in both time and frequency domains.

4:45

2pSA13. Eigenfrequencies of fluid waves entrained in elastic shells. H. Uberall (Dept. of Phys., The Catholic Univ. of America, Washington, DC 20064) and M. F. Werby (Stennis Space Center, MS 39529)

Acoustic signals that scatter from an elastic shell may excite elastic shell resonances. If the shell is evacuated then these as well as the pseudo-Stoneley resonances are the only ones allowed. However, if there is a fluid inside then the eigenfrequencies of the enclosed fluid may be excited. Further, the presence of the fluid may alter the elastic body resonances. When the impedance of the enclosed fluid is not much smaller than the mechanical impedance of the elastic material then the scattered signal is greatly influenced by the enclosed fluid, but only at the allowed eigenfrequencies. Since the fluid is enclosed this leads to a discrete spectrum and an eigenvalue problem which is quite manageable. In this work we outline a method for determining the eigenfrequencies as well as their nature and use their values to isolate the actual body resonances due to the elastic material. We also illustrate the influence of the presence of the enclosed fluid on the elastic resonance frequencies.

TUESDAY AFTERNOON, 5 JUNE 2001 STATE BALLROOM, 1:00 TO 5:00 P.M.

Session 2pSC

Speech Communication: Dynamics of Speech Production and Perception

Pierre L. Divenyi, Chair
Speech and Hearing Research, VA Medical Center, 150 Muir Road, Martinez, California 94553

Chair's Introduction—1:00

Invited Papers

1:05

2pSC1. Reconciling static and dynamic aspects of the speech process. Björn Lindblom (Dept of Linguist., Univ. of Stockholm, S-10961 Stockholm, Sweden, lindblom@ling.su.se) and Randy L. Diehl (Univ. of Texas, Austin, TX 78712)

In speech, movement and spectral change are pervasive whereas steady-states are rare. Yet descriptive frameworks (e.g., IPA) focus mainly on static properties. Similarly, in quantitative models, dynamics is generally attributed to response characteristics while the underlying control consists of static units, e.g., spatial ''*targets*,'' ''*equilibrium points*,'' or ''*stable auditory goals*.'' Accordingly, the input to the production of a diphthong would not be some dynamic aspect of the event, but an ordered string of two articulatory states. This approach creates a paradox when placed in the context of findings on visual and auditory perception that show that perceptual systems are more sensitive to changing stimulus arrays than to purely static ones. Clearly nervous systems have evolved to detect change. We are led to ask: If perception likes change, why assume production control in terms of static targets? We maintain that the two perspectives can be reconciled by according a significant role to positional targets in movement control and, at the same time, fully acknowledging the importance of dynamics in perception. In support of this position we present some recent computational work on simulating vowel systems by selecting optimal sets of vowels seen as objects of spectral change. [Work supported by NIH.]

1:35

2pSC2. Dynamic units in speech production: Evidence from speech production errors. Louis Goldstein (Dept. of Linguist., Yale Univ. and Haskins Labs., 270 Crown St., New Haven, CT 06511, louis.goldstein@yale.edu) and Dani Byrd (Univ. of Southern California, Los Angeles, CA 90089-1693)

Measurements of vocal tract activity during speech production show many distinct parts, all exhibiting different patterns of near-continuous motion. This description stands in sharp contrast to the description of speech as composed of a small number of discrete phonological units. In recent years the tools of dynamical systems have been brought to bear on the problem of relating these descriptions. The continuous vocal tract activity can be theoretically decomposed into discrete units of action, or gestures, each of which is a dynamical control regime for vocal tract articulators. These gestures can also serve as primitive phonological elements. Evidence for gestures as units of the speech production process has recently been obtained in experiments that measure the kinematics of speech errors produced when a simple phrase (e.g., ''cop top'') is repeated. In errors, one of the gestures comprising the phrase may be duplicated in an anomalous position in the utterance. Since the anomalous gesture is often reduced in magnitude, such errors could not be detected without articulatory movement data. The nature of the errors and their asymmetries are consistent with the hypothesis that they represent phase transitions of a system of coupled dynamic units. [Work supported by NIH.]

2:05

2pSC3. Production, coproduction and perception of speech gestures. René Carré (Ecole Nationale Supérieure des Télécommunications, Département Signal, 46 rue Barrault, 75634 Paris cedex 13, France, carre@sig.enst.fr)

An acoustic model was used to derive distinctive speech gestures that result in dynamic deformations of the area function of an acoustic tube. These deformations are efficient, i.e., they provide maximum acoustic contrasts. It can be demonstrated that the gestures generated by the model are consistent with those in natural speech production. Using sequential and/or parallel combination of gestures, V_1V_2 and V_1CV_2 utterances were generated and presented to listeners to assess the effect of a particular coproduction scheme on phonemic identification. Perception of some gesture combinations remained remarkably invariant despite considerable variations of coproduction parameters: In a V_1CV_2 token, for example, the onset of the consonant closure /re/ the V_1V_2 transition can change to a large degree without altering the percept. In contrast, tokens such as [aya], obtained with two coproduced gestures (one tongue and one labial), require strict intergesture timing when [y] is just reached, in order for the French labial /y/ to be perceived. A comparison of the two classes of coproduction schemes and their perceptual consequences suggests possible ways of extracting, from formant trajectories (in the $F1$-$F2$ plane or the $F1$-$F2$-$F3$ solid) represented in the time domain, acoustic cues that may underlie a number of coproduced gestures.

2:35–2:50 Break

2:50

2pSC4. The synergy between speech production and perception. Shihab A. Shamma (Dept. of Elec. Eng., Inst. for Systems Res., Univ. of Maryland, College Park, MD 20742, sas@eng.umd.edu)

Speech intelligibility is relatively unaffected by certain spectro-temporal deformations of the signal spectrogram. These include spectral and dynamic translations, stretching or contracting dilations, and shearing of the spectrogram along its temporal or logarithmic frequency axis. We shall argue that this stability results from a synergy between the dynamics and characteristics of the auditory cortex (the receiver) and of the vocal tract (the source). Specifically, on the perception side, neurophysiological evidence suggests that the acoustic spectrogram is represented in the primary auditory cortex along multiple spectro-temporal scales, and that the above deformations correspond to simple translations of the representation of the dynamic acoustic spectrum along these axes. On the production side, these spectro-temporal distortions can be caused by common variations among talkers in speaking rates, slurred articulatory dynamics, vocal tract length, cross-sectional profile, and losses. Using a simplified sinusoid model of the vocal tract, it is possible to relate ''articulatory'' parameters (e.g., the extent and location of a vocal tract constriction) directly to the formants, and hence to the cortical representation of the spectrum. A similar remarkable synergy exists between the dynamics of perception (e.g., via the perceived spectro-temporal modulation transfer functions) and the speed of articulation and hence the syllabic rates of speech. These results suggest an intimate link between perception and production, a view which offers new insights into the organization of speech phonemes (such as vowels) in terms of different vocal tract configurations.

3:20

2pSC5. Vowel identity change: Trajectory length, transition velocity, or effort? Pierre L. Divenyi (Speech and Hearing Res., V.A. Medical Ctr., Martinez, CA 94553, pdivenyi@ebire.org)

To what degree does the vocal tract shape need to be changed for a listener to perceive a vowel change? A 100-ms vowel V_1 was followed by a transition toward vowel V_2. V_1 and V_2 were two vowels chosen from the triplet /a/, /i/, and /u/. Subjects had to indicate their confidence level that the transition, in fact, reached V_2. Data were converted to response-operating-characteristic (ROC) curves from which the transition velocity and transition duration were calculated for a fixed criterion of V_2 percept. Results indicated that, regardless of the actual V_1V_2 trajectory, the criterion was reached when the product of transition duration and transition velocity (=the trajectory length) was approximately constant. Since there is no logical explanation that could account for an invariance of trajectory length across all V_1V_2 pairs, and since the criterion was reached with transition velocities that are not invariant, it appears that a constant percept is obtained when the transition encounters an elastic, opposing force that is inversely proportional to the acceleration at the start of the transition. This explanation, based on inferred articulatory effort, suggests the existence of a link between production and perception dynamics. [Work supported by NIH and the VA Medical Research.]

3:50

2pSC6. What are the essential cues for understanding spoken language? Steven Greenberg (Intl. Computer Sci. Inst., 1947 Center St., Berkeley, CA 94704, steveng@icsi.berkeley.edu)

Classical models of speech recognition (by both human and machine) assume that a detailed, short-term analysis of the acoustic signal is essential for accurately decoding spoken language. Several lines of evidence call this assumption into question: (1) intelligibility is relatively unimpaired when the frequency spectrum is distorted under a wide range of conditions (including cross-spectral asynchrony, reverberation, waveform time reversal and selective deletion of 80% of the spectrum), (2) the acoustic properties of spontaneous speech rarely conform to canonical patterns associated with specific phonetic segments, and (3) automatic-speech-recognition phonetic classifiers often require ca. 250 ms of acoustic context (spanning several segments) to function reliably. This pattern of evidence suggests that the essential cues for understanding spoken language are largely dynamic in nature, derived from (1) the complex modulation spectrum (incorporating both amplitude and phase) below 20 Hz, (2) segmentation of the speech signal into syllabic intervals between 50 and 400 ms, and (3) a multi-time-scale, coarse-grained analysis of phonetic constituents into features based on voicing, manner and place of articulation. [Work supported by the U.S. Department of Defense and NSF.]

4:20–5:00

Panel Discussion

TUESDAY AFTERNOON, 5 JUNE 2001 MONROE ROOM, 1:00 TO 4:35 P.M.

Session 2pSP

Signal Processing in Acoustics and Underwater Acoustics: Bayesian Signal Processing Approach in Acoustics

James V. Candy, Cochair
Lawrence Livermore National Laboratory, University of California, L-156, P.O. Box 808, Livermore, California 94551

Ning Xiang, Cochair
National Center for Physical Acoustics, University of Mississippi, P.O. Box 1848, University, Mississippi 38677

Chair's Introduction—1:00

Invited Papers

1:05

2pSP1. Bayesian signal processing in acoustics: detection, estimation and tracking. Leon H. Sibul (Appl. Res. Lab., Penn State Univ., University Park, PA 16802, lhs2@psu.edu)

A tutorial introductory lecture on Bayesian signal processing for detection, estimation, and tracking is presented. After a brief historical overview, Bayes rule and risk are defined and used for development of detectors that minimize Bayes risk. Detectors that minimize Bayes risk are called Bayes detectors. Bayes risk in statistical signal processing is the expected cost of making a wrong decision. The decision process of deciding between to mutually exclusive and exhaustive alternatives (i.e., an echo and noise are present versus noise only is present) is a binary hypothesis test or detector. We show how Bayes detectors are related to the maximum likelihood test and Neyman–Pearson and minimax criteria. Multiple hypotheses tests are also reviewed. Estimate of random parameters that minimize the risk are called Bayes estimates and the resulting risk, the Bayes risk. Minimum mean square error (MMSE), MAP (mode of the posteriori density) and other estimates can be derived using appropriate cost functions. Sequential application of Bayes rule can be used to derive Wiener and Kalman filters. Some of the basic difficulties and issues of Bayesian signal processing will be discussed. [Supported by ONR, Code 333, Les Jacobi, Program Officer.]

1:55

2pSP2. Incorporating uncertainty in ocean acoustics for optimum signal detection. Loren W. Nolte, Stacy L. Tantum, and Liewei Sha (Dept. of Elec. and Computer Eng., Duke Univ., Durham, NC 27708)

An overview of detection, classification and localization of signals in an uncertain ocean environment is presented from a Bayesian perspective. The Bayesian approach allows one to both model and incorporate, in an optimal way, the inherent uncertainty that often exists in the knowledge of ocean acoustic environmental parameters. The relationship between environmental parameter estimation and signal detection is illustrated, and it is shown how environmental parameter estimation can be viewed as an inherent part of the optimum Bayesian detector. Furthermore, by using the Bayesian approach, one can design optimum detection algorithms that are

robust with respect to precise knowledge of environmental parameters. Using the receiver operating characteristic (ROC), one can illustrate the trade-off among detection performance, environmental uncertainty, and signal-to-noise ratio. Several examples are taken from an uncertain shallow water environment to illustrate these trade-offs. Finally, it is shown how depth classification can be viewed as an optimal detection problem. [Work supported by ONR.]

2:15

2pSP3. Towed array processing as a Bayesian problem. Edmund J. Sullivan (OASIS, 5 Militia Dr., Lexington, MA 02421, sullivan@oasislex.com)

The towed array range and bearing estimation problem is cast in the form of a Bayesian estimation problem. It is shown that by casting the problem in the form of a MAP estimator, the Kalman estimator naturally follows. There are two advantages to this. First, the problem becomes recursive, resulting in an adaptive processor. Second, once the problem is in the Kalman form, it becomes possible to include sophisticated models of the signals and noise in a natural way. Thus, performance can be enhanced, since the models essentially provide *a priori* information to the processor. In this work an algorithm, which explicitly contains the forward motion of the array, has been developed. It is capable of performing bearing and range (wavefront curvature) estimation with a short array. Here, short means that the ratio of the physical aperture of the array to the range of interest is small. An example using simulated data is given where an array with a length of 45 meters is capable of determining the range and bearing of a narrow band source. Results are shown for several array speeds, signal-to-noise ratios bearings and ranges.

2:35–2:50 Break

2:50

2pSP4. Bayesian inference in architectural acoustics: Estimation of decay times in coupled spaces. Paul M. Goggans (Dept. of Elec. Eng., Univ. of Mississippi, University, MS 38677) and Ning Xiang (Univ. of Mississippi, University, MS 38677)

Sound energy in coupled spaces decays at multiple rates under certain conditions. Identifying decay times in these coupled spaces often demands considerable effort. Generally, different portions of multi-rate decay functions obtained from measured data cannot always be distinctly recognized. A model-based parameter estimation approach within the Bayesian framework facilitates the evaluation of multiple decay times through an extension of the decay model established in a previous work [N. Xiang, J. Acoust. Soc. Am. **98**, 2112–2121 (1995)]. Unlike gradient-based approaches, no careful estimation of initial values is required. Therefore, a robust algorithmic estimation of multiple decay times from experimentally measured decay functions shows advantages over the existing nonlinear regression approach.

3:10

2pSP5. Separation and localization of acoustic signals. Kevin H. Knuth (Ctr. for Adv. Brain Imaging, The Nathan Kline Inst., 140 Old Orangeburg Rd., Orangeburg, NY 10962)

The cocktail party problem is considered where there are several sound sources present in a free-field environment and an observer, who has the ability to record only linear mixtures of these sounds, is interested in obtaining estimates of the original source signals. Previous application of Bayesian inference to the general problem of blind source separation has demonstrated that it is possible to include additional prior information such as the transmission of the signals to the detectors and the possible locations of the sources to improve the source estimates. The derivation of prior probabilities describing such information will be discussed as will the construction of useful models and application of a general Bayesian methodology allowing one to deal with noninstantaneous inverse square transmission, noisy environments and any number of sources and detectors. Methods for dealing with environments in enclosure will also be described. [Work supported in part by NARSAD Young Investigator Award.]

3:30

2pSP6. Bayesian estimation framework for source separation with mixture densities. Ali Mohammad-Djafari and Hichem Snoussi (L2S, Supelec, Plateau de Moulon, 91192 Gif-sur-Yvette, France)

One of the main signal processing problems in acoustics applications is the sources separation. This problem is inherently an ill-posed problem. The Bayesian inference framework is a coherent way to solve such problems by modeling sources and canals and by combining prior information coming from these probabilistic modeling and information included in the data. In this contribution, after a brief presentation of general source separation problems and the Bayesian inference framework, we present new algorithms to source separation for the case of noisy instantaneous linear mixture, within the Bayesian estimation framework. The prior source distribution is modeled by a mixture of Gaussians and the mixing matrix elements distributions by a Gaussian. We model the mixture of Gaussians hierarchically by means of hidden variables representing the labels of the mixture. Then, we consider the joint *a posteriori* distribution of sources, mixing matrix elements, labels of the mixture, and other parameters of the mixture with appropriate prior probability laws to eliminate degeneracy of the likelihood function of variance parameters and we propose iterative algorithms to estimate jointly sources, mixing matrix, and hyperparameters: Joint MAP (maximum *a posteriori*) algorithms.

3:50

2pSP7. Acoustic inversion via linearization and Bayesian multipath identification. Xiaoqun Ma and Zoi-Heleni Michalopoulou (Dept. of Mathematical Sci., New Jersey Inst. of Technol., University Heights, Newark, NJ 07102)

A linearization approach to acoustic inversion is proposed employing distinct ray paths (direct arrival, first surface bounce, and first bottom bounce) for source localization and bathymetry and sound speed estimation. The ray path arrivals are selected from broadband, shallow water, synthetic data using a Bayesian time delay estimation scheme calculating posterior probability density functions of the delays in an efficient way. A linear system is then formed relating unknown parameters and arrival time data. The regularization method is used for the solution of the linear system [S. E. Dosso *et al.*, J. Acoust. Soc. Am. **104**, 846–859 (1998)] with excellent results. Also a simple least-squares approach for the solution of the system is implemented; results of the two approaches are compared. Finally, the linearization multipath based technique is successfully applied to real acoustic broadband data for source and receiver localization, and bathymetry and sound speed estimation. [Work supported by ONR.]

4:05

2pSP8. A Bayesian framework for perceptually motivated audio signal enhancement. Patrick J. Wolfe and Simon J. Godsill (Signal Processing Group, Eng. Dept., Cambridge Univ., Trumpington St., Cambridge CB2 1PZ, UK, p.wolfe@ieee.org)

Spurred by the success of perceptual models in audio coding applications, researchers have begun to address audio signal enhancement in a similar manner. Here a statistical model-based approach is presented that uses cost functions based on auditory perception. As noise reduction inevitably occurs at the expense of signal resolution, why not take advantage of human perception in order to optimize this trade-off? By mathematically incorporating the notion of perceived signal quality, Bayesian risk theory does just that. The Bayesian paradigm is shown to provide an ideal framework within which to formalize such an approach, as it is rigorous, powerful, and generalizable. At the same time it allows the incorporation not only of psychoacoustic optimality criteria, but also of additional nonperceptual prior information such as that concerning audio signal behavior in time and frequency. Importantly, it also permits sequential estimation for real-time noise reduction. Audio examples and the corresponding noise reduction software may be found at http://www-sigproc.eng.cam.ac.uk/~pjw47. [Material by the first author is based upon work supported by a U.S. National Science Foundation graduate research fellowship.]

4:20

2pSP9. Acoustic benchmark analysis of GRASP optimum sonar search plans. Donald R. DelBalzo, Melvin A. Wagstaff, Mona A. Collins, James H. Leclere, and Kevin P. Hempsteter (Naval Res. Lab, Stennis Space Center, MS 39529-5004)

An application of a genetic algorithm and Monte Carlo simulation with Bayesian detection statistics is used to optimize sonar search tracks in nonhomogeneous environments. The optimization metric is maximum cumulative detection probability for a specific sonar (passive or active) against a target with specified characteristics (acoustic and tactical) during a fixed time period. This new search planning capability is named GRASP for genetic range-dependent algorithm for search planning. The sensitivity of GRASP solutions to various ocean environments is examined under benchmark conditions, i.e., fairly simple synthetic environments and a simple target model. The results indicate that the genetic algorithm produces intuitive tracks that correlate highly with acoustic signal excess predictions, as expected.

TUESDAY AFTERNOON, 5 JUNE 2001 ADAMS ROOM, 1:25 TO 5:00 P.M.

Session 2pUW

Underwater Acoustics and Acoustical Oceanography: North Pacific Acoustic Laboratory Experiment

Peter F. Worcester, Chair

Scripps Institute of Oceanography, University of California at San Diego, 9500 Gilman Drive, La Jolla, California 92093-0225

Chair's Introduction—1:25

Contributed Papers

1:30

2pUW1. The North Pacific Acoustic Laboratory (NPAL) experiment. **The** NPAL Group (J. A. Colosi, B. D. Cornuelle, B. D. Dushaw, M. A. Dzieciuch, B. M. Howe, J. A. Mercer, R. C. Spindel, P. F. Worcester) (SIO-UCSD, La Jolla, CA 92093; APL-UW, Seattle, WA 98105; and WHOI, Woods Hole, MA 02543, pworcester@ucsd.edu)

The North Pacific Acoustic Laboratory program augmented the existing ATOC acoustic network with a sparse, two-dimensional receiving array installed west of Sur Ridge, California, from July 1998 through June 1999, to receive transmissions from the 75-Hz ATOC source north of Kauai. The NPAL array consisted of four 20-element vertical arrays, each with a 700-m aperture, and one 40-element vertical array with a 1400-m aperture. The arrays were deployed transverse to the 3900-km acoustic path from the Kauai source and had a total horizontal aperture of 3600 m. Data collected with the billboard array and the U.S. Navy SOSUS receivers are being used (i) to study the temporal, vertical, and horizontal coherence of long-range, low-frequency resolved rays and modes, (ii) to study 3-D propagation effects, (iii) to examine directional ambient noise properties, and (iv) to improve basin-scale ocean nowcasts via assimilation of acoustic data and other data types into models. In addition to acoustic data, environmental data along the path from the Kauai source to the billboard array were acquired by two oceanographic sub-surface moorings and by two XBT/CTD/ADCP transects along the path. The experiment will be described and some preliminary results presented.

1:45

2pUW2. Extracting acoustic observables from the NPAL billboard array data. Matthew Dzieciuch (Scripps Inst. of Oceanogr., Univ. of California, San Diego, CA, mad@ucsd.edu) 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, P. F. Worcester) (SIO-UCSD, La Jolla, CA 92093; APL-UW, Seattle, WA 98105; and WHOI, Woods Hole, MA 02543)

Low-frequency (75-Hz) acoustic signals were repeatedly transmitted over a 1 year period and sampled vertically (with up to a 1400-m aperture) and horizontally (with a 3600-m cross-range aperture) by a distant billboard array (3900-km range) as described by the NPAL Group. The data are complicated by the fact that the sound interacts with the bottom near both the source and receiver. Vertical beamforming is used to filter the bottom interacting energy, and thus allow analysis of the fundamental acoustic properties. Subband and subarray processing is used to produce estimates of arrival times and angles or resolved ray arrivals. Time-series of acoustic travel-times and arrival angles are then developed.

2:00

2pUW3. Analysis of mode coherence and intensity at megameter ranges. Kathleen E. Wage (George Mason University, MS 1G5, 4400 University Dr., Fairfax, VA 22030) 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, P. F. Worcester)

The low-order modes constitute some of the most energetic arrivals at long ranges. Understanding fluctuations of these mode arrivals is crucial to their use as observables in matched field processing and tomography. Both simulated and experimental data indicate that at megameter ranges, the low modes have complex arrival patterns due to internal-wave-induced coupling. Analysis of broadband receptions at 3515 km from the ATOC experiment has shown that mode coherence times are on the order of 6 minutes and that centroid statistics provide useful measures of arrival time trends over the course of several months [Wage *et al.*, IEEE Sensor Array and Multichannel Signal Processing Workshop Proceedings, pp. 102–106, 2000]. The North Pacific Acoustic Laboratory (NPAL) experiment presents an opportunity for further research on broadband mode arrivals at megameter ranges. This study examines temporal coherence, intensity variations, and other mode statistics using data from the 40-element NPAL vertical line array. Experimental results are compared with PE simulations of propagation through internal waves of varying strengths, and the impact of the up-slope propagation near the receivers on the mode statistics is discussed.

2:15

2pUW4. Observing horizontal wave fronts from the NPAL billboard array data. Matthew Dzieciuch (Scripps Inst. of Oceanogr., Univ. of California, San Diego, CA, mad@ucsd.edu) 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, P. F. Worcester) (SIO-UCSD, La Jolla, CA 92093; APL-UW, Seattle, WA 98105; and WHOI, Woods Hole, MA 02543)

One of the main objectives of the NPAL experiment is to investigate the horizontal refraction and coherence of the acoustic wave fronts at long range. Given time series of acoustic arrival times and angles of resolved ray arrival arrivals, a detailed look at the acoustic wave fronts is possible. First and second order statistics (density functions and coherences) of the wave fronts are investigated. The wave fronts are shown to vary with time, frequency, depth and across the horizontal aperture.

2:30

2pUW5. Horizontal refraction of acoustic signals retrieved from NPAL billboard array data. Alexander G. Voronovich, Vladimir E. Ostashev (NOAA/Environ. Technol. Lab., 325 Broadway, Boulder, CO 80305-3328), 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, P. F. Worcester)

A comprehensive, long-range sound propagation experiment was carried out with the use of the billboard acoustic array of the North Pacific Acoustic Laboratory (NPAL) in 1999. The antenna consisting of five vertical line arrays was deployed near a California coast and received broadband acoustic signals transmitted from Hawaii over a distance of about 4000 km. Acoustic signals propagating over such a long distance might exhibit noticeable horizontal refraction. The paper will present results of processing the NPAL data to infer horizontal refraction angle (HRA) as a function of time. Different methods were used for determining HRA. The first approach employed cross correlation of the acoustic signals at different VLAs. Time delay corresponding to maximum of cross correlation is related to HRA assuming the angle is approximately the same for all rays (or modes). The second method used modal representation of the arriving broadband signals. The dependency of the amplitudes of acoustic modes on mode number, frequency, and arrival angle was determined independently within narrow frequency bins, and then the results were averaged over whole frequency range. This method allowed, in particular, to evaluate angular width of the arrived signal, which appeared to be of the order of a few milliradians.

2:45

2pUW6. The terminal problem: Model-data comparisons of basin scale transmissions including effects of shallow bathymetry near source and receiver. Kevin D. Heaney (ORINCON Corp., 4350 N. Fairfax Dr., Arlington, VA 22203) and the NPAL Group (P. F. Worcester, J. A. Colosi, B. D. Cornuelle, B. D. Dushaw, M. A. Dzieciuch, B. M. Howe, J. A. Mercer, R. C. Spindel)

Acoustic transmissions on basin scale ranges are being used to determine depth-dependent temperature variability. With travel time being the primary observable, stationary sources and nearly stationary receivers are experimental requirements. This has led to the use of bottom-mounted sources and receivers to reduce travel time variability. The NPAL (North Pacific Acoustics Laboratory) experiment has transmitted broadband acoustic pulses from two bottom-mounted sources near the sound channel axis. Recordings have been taken on the NPAL billboard array, a linear series of five vertical line arrays moored in 1800 m of water near Monterey, CA. Additional recordings have been taken from the SOSUS system throughout the Pacific basin. The effects of the near source and near receiver scattering are examined. In particular, near source scattering leads to excess high-angle energy entering deep water with a travel time delay of nearly 1 s due to the low group speeds of high-angle rays/modes in shallow water. We also compare the energetics of the arriving rays that have bounced once on the rising seafloor near the NPAL receivers. Comparisons of models and data for bottom interacting acoustics lead us to the perennial issue of geoacoustic parameters.

3:00–3:15 Break

3:15

2pUW7. On ''ray-like'' arrivals in the deep-ocean shadow zone of megameter range acoustic transmissions. Brian Dushaw 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, Peter F. Worcester) (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105-6698)

Receptions of long-range acoustic transmissions by deep hydrophone arrays in the Pacific and Atlantic often have ''ray-like'' arrivals that occur in the shadow zone of the predicted time front. These ''ray-like'' arrivals can frequently be identified with the cusps of the predicted time front, but

the receivers are up to 750 m below the depth of the cusps. Preliminary calculations show that the observed acoustic energy is not accounted for by errors in the sound speed, leakage of acoustic energy below the cusps as predicted by the full wave equation, or scattering due to internal waves. Data obtained during experiments in the Atlantic and Pacific will be reviewed. Experiments that have been conducted with receivers of vertical line arrays have not had receivers deep enough to observe this phenomena. The effect is seen when bottom-mounted or midwater acoustic sources are used. These data present a number of problems: If the ray paths are wandering all over the water column, why are predictions of ray travel times usually accurate? How does the energy loss associated with these data increase the attenuation of very long-range acoustic transmissions? Without knowing the forward problem, how can these data be used to determine oceanographic information?

3:30

2pUW8. Bottom grazing acoustic arrivals: Explanation of anomalous arriving deep multipaths. Kevin D. Heaney and Henry Cox (Orincon Corp., 4350 N. Fairfax Dr., Arlington, VA 22203)

During the 1996 Acoustic Engineering Test (AET) of the ATOC project, receptions were recorded on deep SOSUS arrays in the central Pacific. With the given SNR, several pairs of doublet arrivals are visible. These doublets have the form one would expect from an axial source, but to a receiver much shallower in the water ($\sim$500–1000 m). The presence of this energy, well below the turning point of the corresponding rays has been termed the ''acoustic shadowzone phenomena'' [Dushaw *et al.*, IEEE J. Ocean. Eng. **24**, 202–214 (1999)]. Range independent normal mode modeling for an axial source reveals bottom grazing acoustic energy at all ranges. The arrival envelope (pairs of doublets) is independent of range. These arrivals are present in ray theory calculations when bottom grazing ray arrivals are included. With a critical angle of 18 deg, there are many subcritical bottom interacting rays with very little attenuation. This phenomenon has been understood in shallow-water acoustic problems where the reflection coefficient is linear with angle. The time spread of the pulse (number of rays) grows linear with range. The attenuation of the pulse decays linearly with range. The result is an envelope that is independent of range.

3:45

2pUW9. Scattering into the shadow zone. Walter Munk (Scripps Inst. of Oceanogr., Univ. of California–San Diego, La Jolla, CA 92093)

We consider the scattering of sound into the shadow zone in long-range transmission by processes other than internal waves. Microfrontal activity measured by Dan Rudnick in the north Pacific appears to be a significant factor.

4:00

2pUW10. Does semiclassical ray theory work for megameter propagation? Frank S. Henyey (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105)

In the presence of internal waves, a very large number of rays connect any source to a distant receiver. The ray intensities vary by orders of magnitude, causing the scintillation index predicted by semiclassical ray theory to be very large. In this talk, possible diffractive effects at frequencies at or below 100 Hz are considered. One effect is caustics. A region where Airy functions must be used can be identified; if this region contains additional caustics, Airy functions do not suffice. This caustic region is identified and characterized from numerical ray-tracing simulations. The rays that are high intensity occur in narrow beams. The diffraction of these beams is estimated from the same numerical simulations. [Work supported by ONR.]

4:15

2pUW11. Scintillations in long-range acoustic transmissions. John Colosi (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543), Frederick Tappert (Rosenstiel School of Marine and Atmospheric Sci., Miami, FL 33149), and Matthew Dzieciuch (Scripps Inst. of Oceanogr., La Jolla, CA 92093)

In the Acoustic Thermometry of Ocean Climate (ATOC) program's Acoustic Engineering Test (AET), broadband 75-Hz center frequency transmissions were recorded on a 700-m-long vertical array, 3252 km distant. Previously reported results from the AET using 12.7-min. averaged data by Colosi *et al.* [J. Acoust. Soc. Am. **105**, 3202–3218 (1999), hereafter referred to as Colosi99] revealed surprisingly weak acoustic scattering for early arriving identifiable wave-fronts; these results have been confirmed using 1.8-min averaged data. It is shown that scintillation index (SI) is a strong function of position along the pulse with smallest values occurring at the peak and larger values occurring at the tails. Intensity PDFs of identifiable wave-fronts are reanalyzed in terms of both peak intensity and integrated pulse energy (IE), but both PDFs are very closely log-normal. Regarding multipathing along the wave-fronts it is found that on average there are 1.7 peaks per wave-front segment per hydrophone. The combined observation of weak scattering and multipathing is a novel result. Colosi99 analyzed the finale in terms of peak scintillations and found a near log-normal intensity PDF. Reprocessing the full field without limiting data to intensity peaks and accounting for mean intensity nonstationarity yields an exponential intensity PDF.

4:30

2pUW12. Noise field statistics and coherence on the NPAL array. Arthur B. Baggeroer (Dept. of Ocean Eng., MIT, Cambridge, MA 02139, abb@arctic.mit.edu), Edward K. Scheer (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543), 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, P. F. Worcester)

While the NPAL array was primarily deployed to examine the spatial coherence of the Hawaii source, it is also a rich data set for ambient noise studies. Shipping noise, earthquakes and biologics all have been identified in the NPAL data. Moreover, ambient noise coherence is the primary issue in maximizing the SNR output of a sonar system. The first and second order statistics of data from the NPAL ''noise only'' segments have been analyzed with the following results: (i) There is a wide spread in the peak levels, most likely due to the proximity to shipping lanes. The maximum peak level in the recording band is 117 dB. (ii) Full broadband coherences tend to be low because of the presence of many ships. (iii) If one examines frequency bands of 1–2 Hz, then lines of individual ships can be identified and associated and they are very coherent across NPAL aperture. (iv) Vertical beamforming indicates relatively highly directional spectra at low grazing angles and ''noise notch'' for the spectra at higher frequencies. Horizontal beamforming has been difficult to implement due to element positioning errors and the large array transit time.

4:45

2pUW13. A comparison of ocean ambient sound levels after 30 years for a coastal site off California. Rex K. Andrew, Bruce M. Howe, James A. Mercer (Appl. Phys. Lab., 1013 NE 40th St., Seattle, WA 98195-6698, randrew@apl.washington.edu), and Matthew A. Dzieciuch (Univ. of California at San Diego, La Jolla, CA 92093-0225), 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, P. F. Worcester)

As part of the North Pacific Acoustic Laboratory project, ambient sound data from 1994 to the present has been collected. Long-term averages of these data from a receiver on the continental slope west of Point Sur, CA, are compared to earlier measurements made at the same site over 1963–1965 by Wenz [Wenz, J. Underwater Acoust. **19** (1969)]. The levels

Wenz reported fall below our 10% quantile from 5 Hz to 50 Hz, rise to the 50% quantile (i.e., the median) at 100 Hz, and again fall below the 10% quantile by 250 Hz. Wenz removed highly variable ''transient'' data before calculating his averages. We mimicked his processing with the NPAL data and obtained a result which is virtually indistinguishable from the median, which is approximately 1 dB below the (dB) mean of each one-third octave band. Hence, our median levels are directly comparable to Wenz' results, and this comparison shows that the 1994–2000 levels exceed the 1963–1965 levels by 9 dB or less below 100 Hz and again at 250 Hz, but are roughly similar at 100 Hz. [Work sponsored by ONR.]

Session 3aAAa

Architectural Acoustics: Student Design Competition in Architectural Acoustics

Robert C. Coffeen, Cochair
Architectural Engineering, University of Kansas, Marvin Hall, Lawrence, Kansas 66045

Lily M. Wang, Cochair
University of Nebraska, 200B PKI, Omaha, Nebraska 68182-0681

Robin S. Glosemeyer, Cochair
Jaffe Holden Acoustics, Inc., 114A Washington Street, Norwalk, Connecticut 06854

The Technical Committee on Architectural Acoustics of the Acoustical Society of America and the National Council of Acoustical Consultants are sponsoring a student design competition which will be judged at the meeting.

The purpose of this design competition is to encourage students enrolled in Architecture, Architectural Engineering, and other University curriculums that involve building design to express their knowledge of achitectural acoustics and building noise control in the schematic design of portions of a building where acoustical considerations are of primary importance.

The submitted designs displayed in this session, will be judged by a panel of professional architects and acoustical consultants. An award of $1,000 will be made to the submitter(s) of the design judged ''first honors.'' Four awards of $500 each will be made to the submitters of four entries judged ''commendation.''

Session 3aAAb

Architectural Acoustics: Architectural Acoustic Modeling and Imaging

Neil A. Shaw, Chair
Menlo Scientific Acoustics, Inc., P.O. Box 1610, Topanga, California 90290-1610

Contributed Papers

9:00

3aAAb1. Computer modeling and prediction in the design of coupled volumes for a 1000-seat concert hall at Goshen College, Indiana. Byron W. Harrison, Gary Madaras (The Talaske Group, Inc., 105 N. Oak Park Ave., Oak Park, IL 60301, byron@talaske.com), and Robert D. Celmer (Univ. of Hartford, West Hartford, CT 06117)

The effects of specific architectural parameters (volume, shape, location, aperture size, and location) on the acoustic characteristics of variable-acoustic coupled volumes were analyzed using the computer modeling application, CATT-Acoustic. The purpose of the study was to optimize design and determine the effectiveness of coupled volumes for use in the Goshen College Concert Hall, Goshen, Indiana. The computer modeling technique allowed specific variables to be isolated and parameters to be analyzed. Prototype models were evaluated by a visual inspection of decay curves, comparisons of T15, T30, and T60, and auralizations. The goal was to preserve the early decay rate of the main hall while achieving a slower decay rate after 750 ms. The analysis confirmed the use of computer modeling and prediction in studying the acoustic performance of coupled volume designs. Chamber volumes were found to be the most influential factor in realizing the double-sloped decay. Shape, location, and aperture configuration had noticeable, but lesser, effects on acoustic performance than did chamber volume. To realize efficient acoustic variability for the small-volume hall researched, the chambers required volumes that compared to large percentages of the main enclosure volume (50–75 percent).

9:15

3aAAb2. A model for acoustical evaluation of rectangular concert halls. Nurgun Tamer Bayazit (Faculty of Architecture, Dr. Istanbul Tech. Univ.)

The studies on room acoustics have shown the importance of hall geometry on the acoustical performance perceived subjectively by the audience. The aim of this study is to examine the effects of the geometrical parameters on the objective acoustical parameters and to develop an analytical model which will help designers in the early stage of design. The acoustical parameters of the model halls designed, selected as rectangular forms with varying dimensional ratios, were calculated with the aid of the ODEON room acoustical simulation program. In order to obtain realistic results with the proposed model, the geometrical parameters of model halls were chosen so as to allow comparison with those of existing halls and the set of data used for the calculation of acoustical parameters were chosen to reflect real conditions in the best possible way. Calculated acoustical parameters were statistically analyzed to derive an analytical

model based on the relationship between the hall geometry and the acoustical parameters. The geometrical parameters are fed into the model as independent variables, and the acoustical parameters are calculated on the basis of them.

9:30

3aAAb3. Applications of physical and mathematical modeling in concert hall design. Paul Calamia and Martha Larson (Kirkegaard Assoc., 801 W. Adams St., Chicago, IL 60607)

A variety of testing methods were employed throughout the early and mid stages of design for an eighteen-hundred-seat flexible-use concert hall. Full-scale laboratory testing, model-scale measurements, and mathematical modeling were undertaken to provide guidance on particular design decisions. Mathematical techniques were used to guide the full-scale laboratory testing, and full-scale testing results were applied in the model-scale measurements. Auralization techniques were implemented to study specific effects.

9:45

3aAAb4. Calculation of diffuse sound reflections in predicting room acoustics using the image method. Hee-Won Lee, Il-Doo Ko (Seoul Natl. Univ. of Technol., Seoul, South Korea), Se-Jin Doo (Dong-Ah Broadcasting College, Ansung, Kyonggi, South Korea), Yang-Ki Oh (Mokpo Natl. Univ., Mokpo, Chonnam 534-729, South Korea), Dae-Up Jeong (Chonbuk Natl. Univ., Jeonju, Chonbuk 561-756, South Korea), Chul-Min Choi, and Lai-Hoon Kim (Seoul Natl. Univ., Seoul, South Korea)

Numerous investigations have demonstrated that diffuse reflection is one of the most important factors in predicting room acoustics by computer simulation. Recent studies have suggested several computational algorithms in order to account for diffuse reflections in the ray-tracing or beam-tracing method. In this study, a computational algorithm for the calculation of diffuse sound reflections in the image method is suggested and a computer simulation system is developed based on the suggested algorithm. The methodology adopted in our computer simulation system is similar to the extended radiosity method, which was developed for computer graphics. Various descriptions of room acoustics, including spatial distribution of image sound sources, impulse responses and other commonly used room acoustical measures, can be obtained from the MLS based monaural room acoustics measurement system. The measured results in a midsized rectangular classroom with/without chairs on the floor were compared with the predicted results using the computer simulation in which diffuse reflection coefficients of the floor were varied. [Work supported by Korean Research Foundation Grant KRF-1999-1-310-004-3.]

10:00

3aAAb5. Acoustic simulation and auralization in architectural design practices. Quinsan Ciao (Cao) (Dept. of Architecture, Virginia Tech., Blacksburg, VA 24061)

This paper is dedicated to predictive relationships between space design and acoustic performance in architectural practice through auralization. With such a relationship, design alternatives can be evaluated subjectively, as well as objectively, without being constructed. In traditional practice, designers rely on vague experiences and rough estimates to predict acoustic effects while designing architectural spaces. Acoustic analysis and predication are seen as a luxury remedy and only affordable in large-scale theaters and concert halls. The recent available personal-computer-based auralization technologies brought the possibility of applying the latest science and art of architectural acoustics into design practices. Case studies presented in this paper illustrate that the auralization technology makes it possible for potential occupants as well as designers to evaluate the acoustic performance of a design by hearing it directly before a decision to build is made. They also illustrate that the auralization is a powerful tool for the general public to uncover everyday acoustic problems in common building types that have been constantly harming their well being and would otherwise be undetected. Furthermore, they demonstrate that auralization is an effective means to evaluate noise effect on subjective perception where the noise effect is signal context dependent.

10:15

3aAAb6. Interactive and aliasing-free acoustic modeling of reflections and diffractions in architectural environments. Thomas Funkhouser (Princeton Univ./Bell Labs., Murray Hill, NJ 07974), Nicolas Tsingos, Ingrid Carlbom, Gary Elko, Gopal Pingali (Bell Labs., Murray Hill, NJ 07974), Mohan Sondhi, and Jim West (Bell Labs.)

A primary challenge in geometrical acoustic modeling is computation of reverberation paths from sound sources fast enough for real-time auralization. This paper describes an aliasing-free beam tracing algorithm based on precomputed spatial subdivision and beam tree data structures that enables real-time acoustic modeling and auralization for sound sources in interactive applications. The proposed method traces convex polyhedral beams from the location of each sound source and receiver through a precomputed spatial subdivision data structure, constructing a beam tree representing the regions of space reachable by potential sequences of transmissions, diffractions, and specular reflections at surfaces of a 3D polygonal model. By computing beam trees asynchronously (off-line), our system can generate reverberation paths between sources and receivers at interactive rates and spatialize audio signals in real-time as sources and receivers move under interactive user control. Unlike previous geometrical acoustic modeling work, our beam tracing method: (1) supports evaluation of early reverberated paths at interactive rates, (2) scales well for large, densely occluded architectural environments, and (3) computes paths of diffraction without aliasing using the uniform theory of diffraction. This system is being used to develop interactive applications in which a user experiences a virtual environment immersively via simultaneous auralization and visualization.

Session 3aAB

Animal Bioacoustics: Animal Vocalizations

Jeanette Thomas, Chair
Laboratory of Sensory Biology, Western Illinois University, 3561 60th Street, Moline, Illinois 61265

Contributed Papers

8:30

3aAB1. Spatial and temporal variation of sperm whale (*Physeter macrocephalus*) codas in the northern Gulf of Mexico. Trent C. Apple (Marine Acoust. Lab., Texas A&M Univ. at Galveston, 5007 Ave. U, Galveston, TX 77551)

Sperm whale (*Physeter macrocephalus*) codas, short rhythmic patterns of clicks, were recorded annually from a towed linear array throughout the northern Gulf of Mexico during 2-week cruises spanning October 1991–August 1997. Codas were sampled and classified according to their temporal pattern and the number of clicks they contained. Regular codas, with equally spaced intervals between clicks, were analyzed along with irregular codas, with double intervals between the last two clicks. In this sample, short codas (less than five clicks) were more common than long codas (greater than six clicks). Preliminary analysis ($n=11$), using Student's t-test, suggests that irregular codas may not vary in the northern Gulf of Mexico. When comparing interclick intervals, irregular coda types did not differ significantly across multiple years, and no spatial variation was observed. This pattern of similarity in coda types, if upheld with additional analyses, may suggest that coda repertoires remain stable over periods of years or that a resident population of animals may inhabit this area. Further analysis, however, is required to substantiate this supposition. [Work supported by Texas A&M University at Galveston.]

8:45

3aAB2. Analysis of frequency structure of sperm whale clicks as a function of dive depth and animal orientation. Aaron M. Thode (MIT, 77 Massachusetts Ave., Cambridge, MA 02139), David K. Mellinger (Cooperative Inst. for Marine Resources Studies, Oregon State Univ., Newport, OR 97365), Sarah Stienessen (Alaskan Fisheries Sci. Ctr. (AFSC), Seattle, WA 98115), Anthony Martinez (Natl. Marine Fisheries Service, Key Biscayne, FL 33149), and Keith Mullin (Natl. Marine Fisheries Service, Pascagoula, MS 39568)

During overnight tracking of a pod of sperm whales in the Gulf of Mexico in July 2000, the NOAA ship GORDON GUNTER recorded their characteristic ''click'' sounds on a five-element towed hydrophone array. Multiple reflections from the surface and ocean bottom were also recorded. Analysis of the arrival times and bearings of the reflections allowed the computation of three dimensional dive profiles for several animals. By assuming the orientation of the animal was aligned with its velocity, the relative orientation of the animals relative to the array could also be estimated. A visual examination of the frequency content of the received clicks versus dive time suggested that the double resonances in the 1.2 and 2.2 kHz band increased 20%–30% during depth changes of 1000 m. In this presentation the possible relationship between click structure and whale depth and orientation is rigorously analyzed, and the observed relationships are compared with predictions from various sound production and resonator models, in an attempt to gain insight into the sound production mechanism. [Work supported by Minerals Management Service, National Marine Fisheries Service, and ONR.]

9:00

3aAB3. An ''acoustic niche'' for Antarctic killer whale and leopard seal sounds. Julia A. Mossbridge (Audiol. and Hearing Sci. Prog., 2299 N. Campus Dr., Northwestern Univ., Evanston, IL 60208-3550, j-mossbridge@northwestern.edu) and Jeanette A. Thomas (Lab. of Sensory Biol., Western Illinois Univ., Moline, IL 61265)

It is not known whether species that share the same acoustic environment develop an ''acoustic niche'' to reduce competition. To address this issue, spectrograms were used to analyze underwater killer whale (*Orcinus orca*) and leopard seal (*Hydrurga leptonyx*) sounds from recordings made in December and January 1977/1978 in Antarctica. Forty-seven leopard seal sounds were found in December recordings; no leopard seal sounds were found in January recordings. FM points were defined as the frequency within a sound at which a detectable change in the frequency slope occurred. The FM-point distribution for December killer whale sounds showed a gap between 2500 and 3750 Hz, which is near the frequency range of the most common leopard seal sound (2777–3802 Hz). No gap was found in January's FM-point distribution. The FM-point distributions differed significantly between months, while the number of FM points per sound and the duration of killer whale sounds did not differ between months. December killer whale sounds did contain frequencies between 2500 and 3750 Hz, but frequency modulations were uncommon in this frequency range. It seems that killer whales may have used frequency modulation to adapt their acoustic niche when leopard seal sounds were present. [Work supported by NSF.]

9:15

3aAB4. The underwater whistle repertoire of wild false killer whales (*Pseudorca crassidens*). A. E. Nester (Lab. of Sensory Biol., Western Illinois Univ., 23 Kensington Circle, Apt. 208, Wheaton, IL 60187), J. A. Thomas (Lab. of Sensory Biol., Western Illinois Univ., Moline, IL 61265), and A. Acevedo-Guiterrez (Texas A&M Univ. of Galveston, Galveston, TX 77551)

There is limited information on the underwater repertoire of false killer whales (*Pseudorca crassidens*). Underwater recordings of feeding and traveling whales were made near Isla de Coco, Costa Rica in January and February 1994. This study classified the whistle repertoire, analyzed acoustic characteristics, examined differences associated with feeding and traveling whales, and examined whistles in varying noise environments. Whistles were categorized by contour into eight types; 197 were during feeding and 172 during traveling. Whistles were 1220 to 17 891 Hz, had harmonics with intervals between 1797 and 13 125 Hz, and were 0.03 to 2.07 s long. Some differences in whistles from traveling and feeding whales were found. Cluster analysis, principle component analysis, and stepwise discriminant function were used to classify whistle types. Ambient noise was measured at 1 and 2 kHz in recordings of seven different ambient environments (over a 25-dB range). Quieter environments were associated with feeding whales. When ambient noise was high, whales shifted the frequency of whistles higher and increased the duration and total number of whistles in a series. Results provide evidence for whales adapting frequency and time structure of their whistles in different noise environments.

9:30

3aAB5. Whistle repertoire of Pacific white-sided dolphins (*Lagenorhynchus obliquidens*) at the John G. Shedd Aquarium.. Julie L. Whitten and Jeanette A. Thomas (Lab. of Sensory Biol., Western Illinois Univ. Regional Ctr., 3561 60th St., Moline, IL 61265)

Whistles of Pacific white-sided dolphins (*Lagenorhynchus obliquidens*) at the John G. Shedd Aquarium were studied to determine: (1) whistles associated with a particular individual, (2) if whistles of isolated dolphins differed from whistles from dolphins in a group, (3) time and/or frequency variables important in classification, and (4) whistle types. Individual dolphin whistles were recorded in an isolated pool using a hydrophone and a Telex recorder. Whistles were recorded from dolphins in a group setting using the same hydrophone, but with a Marantz cassette recorder. Whistles were digitized using Canary software. Frequency and time measurements were taken from power spectra and spectrograms. ANOVA, principle component analysis (PCA), discriminate analysis (DFA), and cluster analysis were performed. Whistles were classified visually into six types. There were no significant differences between whistles from isolated or social dolphins. PCA showed duration of the series, number of whistle in the series, maximum frequency, minimum frequency, and peak frequency were important in classification. DFA showed dominant bandwidth, duration of first whistle, harmonic structure, maximum frequency, and peak frequency were important in categorizing types. Using the variables in PCA, cluster analysis separated the whistles into different types than those classified visually. This study provides no evidence of individualized signature whistles.

9:45

3aAB6. Separation of the acoustic ''sound space'' by nearshore fishes in the Great Barrier Reef, Australia. Robert D. McCauley (Ctr. for Marine Sci. Technol., Curtin Univ., Kent St. Bentley, Perth, Western Australia) and Douglas H. Cato (Defence Sci. and Technol. Organisation, Pyrmont NSW, Australia)

Long-term sea noise recordings were made off Cowley Beach. Four fish sources dominated ambient noise. Each displayed unique patterns in calling: character, habits, timing, and spatial extent. Two sources, a knocking and drumming, were up to 2 s long, always occurred as distinct calls, had spectral peaks over 350–450 Hz and had respective mean squared pressure (msp) source levels of 132 and 149 dB *re* 1 μPa @ 1 m. Call rates to 120/10 min occurred during the Austral summer, the highest calling period for all sources. The other sources, a ''trumpet'' and ''pop,'' occurred in choruses and, respectively, were 100–220 ms and <20 ms long; had spectral peaks over 800–1500 Hz and 400–700 Hz; and had source levels of 150 and 141 dB *re* 1 μPa @ 1 m (msp). Trumpet choruses were up to 2 km across and detectable to 8 km from chorus center in water <10 m deep. Popping choruses were dispersed over larger regions. It was possible that in nearshore waters all sources occurred within a 24-h period. Sources avoided competition for the acoustic ''space,'' partly by offsetting spectral peaks and geographic separation, but primarily by separating the time of chorus or maximum call-rate.

10:00

3aAB7. Analysis of surface vocalizations from mother and pup Weddell seals (*Leptonychotes weddellii*). Beth A. Noe and Jeanette Thomas (Western Illinois Univ., 1 University Circle, Macomb, IL 61455)

The antarctic Weddell seal (*Leptonychotes weddellii*) is a vociferous species with an extensive underwater repertoire, but little is documented about their airborne sounds. Weddell seals return to traditional pupping colonies, are approached easily, and recordings of individual seals are available from a tagging study conducted since the early 1970's. Therefore, this study on airborne vocalizations of Weddell seal mothers and their pups provides a unique opportunity for detailed analysis of calls. During 1976, 1977, 1979, and 1996, surface vocalizations of mother and pup Weddell seals were recorded. Analysis variables included beginning, ending, maximum, and minimum frequency, fundamental harmonic interval, components per call, intercall interval, and duration. This study determined that surface calls of Weddell seal pups changed as the pup grew older. Although pups within an age bracket showed little variability in call structure, adult vocalizations varied among individuals. Relatedness of mothers and pups could not be concluded solely by the call variables examined. Further studies need to be conducted to determine the extent of the change that pup vocalizations undergo as they mature and to identify the acoustic variables influencing mother/pup vocalizations. [Acknowledgments: Dr. Jeanette Thomas, Dr. Douglas Quin, Bart Jones.]

10:15

3aAB8. Patterns of echolocation in a group of captive bottlenose dolphins in Roatan, Honduras. Heather L. Heinemeyer and J. A. Thomas (Western Illinois Univ. Regional Ctr., 3561 60th St., Moline, IL 61265)

Patterns of echolocation in a group of bottlenose dolphins (Tursiops truncates) at the Institute for Marine Sciences in Roatan, Honduras were examined. While much is known about echolocation, the role of echolocation in the context of dolphin social or communicatory behavior has been, for the most part, speculative (Xitco, 1996). Two females and two males were paired and presented a variety of novel targets in the water. Approximately equal numbers of each pair combination were conducted over 22 sessions consisting of a 5-min no-target trial followed by a 5-min target trial. Data were obtained via an underwater video camera and by an underwater observer. Number of click trains, silent passes, passes by both dolphins, passes by a clicking dolphin, and times a dolphin approach and click on the target first were analyzed. Clearly, dolphins paid attention to the targets. The number of trains and passes significantly increased during target trials. Females clicked more often during target trials and were first to approach and click at the target. Target type had no effect on echolocation behavior. This study should be repeated at other facilities and with other dolphins to determine whether gender differences are real or due to small sample sizes.

10:30

3aAB9. Blue whale calling off the Western Australian coast. Robert D. McCauley (Ctr. for Marine Sci. Technol., Curtin Univ., Kent St. Bentley, Perth, Western Australia), Curt Jenner (Ctr. for Whale Res. (WA), Inc., Fremantle, WA 6160, Australia), John L. Bannister (Western Australian Museum, Perth, Australia), and Douglas H. Cato (Defence Sci. and Technol. Organisation, Pyrmont NSW, Australia)

Throughout 2000, aerial surveys, boat-based studies and acoustical censusing was carried out to search for blue whales in the Rottnest trench, west of Perth. Historical records indicated a Western Australian population of blue whales, while a preliminary boat survey in 1994 suggested a Rottnest trench aggregation. This was confirmed in 2000, with 17 blue whales sighted. Of 5000 acoustic records almost all contained blue whale calling, some having up to nine distinct callers. Although of a different format, the calls had a similar character to those described from other populations. A call was composed of three signals, each respectively 44–45 s, 20–23 s and 20–25 s in length, separated by 5–10 s and 23 s with the sequence repeated every 78 s. Each component was dominated by constant or slowly up-sweeping amplitude-modulated tones over 18–26 Hz, with harmonics evident up to 100 Hz and a secondary pulsed source of frequency 65–66 Hz present. Components had different source levels. Several other baleen whale calls were also common. The low frequency (<100 Hz) sea noise spectra made over 33.5 days off Rottnest was dominated by whale calling.

10:45

3aAB10. An improved contour extractor for bottlenose dolphin whistles. Ryuji Suzuki (Speech and Hearing Sci. Prog., Harvard–MIT Div. of Health, Science and Technol., 77 Massachusetts Ave., E25-518, Cambridge, MA 02139, rsuzuki@mit.edu) and John R. Buck (Univ. of Massachusetts–Dartmouth, North Dartmouth, MA 02747)

The time-frequency contours of bottlenose dolphin (*Tursiops truncatus*) whistles are commonly extracted features for the classification and comparison of these signals. We have previously developed a method to extract the whistles' fundamental frequency contours [J. Acoust. Soc. Am. **108**, 2635 (2000)] by sequentially tracking the contour. These contours are often extracted from recorded rather than real-time signals, which allows us to read ahead or access the ''future'' signal. This talk extends our previous work to propose a technique to estimate the fundamental frequency at each time from the ''future'' values of the recorded signal in addition to the past and present signals. This technique is similar to the Kalman smoother except that the states are probability densities, and it uses a probabilistic state transition model with a Bayesian projection to produce the maximum *a posteriori* probability estimate of the frequency contour based on all recorded data, not just the present and past signal. Experimental evaluation indicates that this method greatly improves the reliability of contour extraction from noisy signals, with considerable resistance to strong harmonics, some dropouts and echolocation clicks, when compared to our previous technique. A huge improvement is achieved from the technique proposed by Buck and Tyack [J. Acoust. Soc. Am. **94**, 2497–2506 (1993)].

WEDNESDAY MORNING, 6 JUNE 2001 STATE BALLROOM, 7:30 A.M. TO 12:00 NOON

Session 3aAO

Acoustical Oceanography and Underwater Acoustics: Inverse Methods for Sub-Bottom Surveys I

Altan Turgut, Cochair
Naval Research Laboratory, Acoustics Division, Code 7127, 4555 Overlook Avenue, SW, Washington, DC 20375

David N. Lambert, Cochair
Naval Research Laboratory, Code 7431, Stennis Space Center, Mississippi 39529

Chair's Introduction—7:30

Invited Papers

7:35

3aAO1. Modeling of sound propagation in sandy ocean sediments. Nicholas P. Chotiros (Appl. Res. Labs., Univ. of Texas, Austin, TX 78713-8029, chotiros@arlut.utexas.edu)

A physically accurate model of sound reflection and penetration into ocean sediments is the objective. Sandy sediments are of particular interest because current models are not compatible with acoustic measurements. The discrepancies of the fluid and elastic solid approximations are clearly demonstrable. The Biot theory has the potential to satisfy the objective. Stolls formulation requires 13 input parameters, which may be divided into three groups according to the accuracy with which they are known. The first group consists of tabulated physical constants; the second is less precisely known, and the third is not measurable. An inversion procedure was devised to estimate the immeasurable group from simple acoustic measurements — reflection loss, compressional and shear wave speeds, and attenuations. The imprecisely known group was handled in a probabilistic manner. The inversion results for water-saturated sand, using published laboratory and at-sea measurements, show a definite incompatibility between model and measurements. To find a solution, a couple of hypotheses were considered: (1) the inclusion of some proportion of the pore fluid within the frame and (2) the relaxation of the uniform, elastic frame assumption. The latter proved to be a viable solution. [Work supported by ONR, Ocean Acoustics.]

8:00

3aAO2. WARBLE: A high resolution approach to subbottom geoacoustic inversion. Charles W. Holland (SACLANT Undersea Res. Ctr., La Spezia, Italy) and John Osler (Defense Res. Estab. Atlantic, Halifax, Canada)

Subbottom geoacoustic information supports a broad range of activities serving military, scientific, and commercial needs. Since direct sampling (i.e., coring) of the seabed is costly and time-consuming, numerous acoustic inverse methods have been developed. One of the most serious weaknesses, which is suffered by all the inversion methods, is a uniqueness problem, wherein multiple solutions fit the data equally well. An associated problem is that some methods have no means to determine the number of sediment layers in the solution space. Our WARBLE (wide-angle reflection and bottom loss experiment) method addresses these problems by conducting the inversion in both the space-time and the space-frequency domains in order to maximize the amount of independent data. The space-time data provide the number of important layers as well as the layer thicknesses, and interval sound speeds. The

space-frequency domain data provide sound velocity, density, and attenuation for each layer. High-resolution results can be obtained with this method, resulting in layer thickness of order several tens of centimeters. Penetration depends upon sediment type and water depth but is typically 50–150 m. The inversion method will be illustrated with several shallow-water measurement/inversion results, including comparisons with core data.

8:25

3aAO3. Quantitative characterization of the uppermost seabed stratum: Some results from the ISACS project. Andrea Caiti (Interuniv. Ctr. Integrated Systems for the Marine Environment, c/o DSEA, Univ. Pisa, v. Diotisalvi 2, 56100 Pisa, Italy, caiti@dsea.unipi.it), Jens M. Hovem, and Bjarte Berntsen (Univ. of Trondheim, Trondheim, Norway)

ISACS is a research initiative funded by the European Union in the years 1996–1999. The objective was to prove the feasibility of seafloor characterization by exploitation and integration of data from commercially available sonar equipment. The activities have focused on analysis of side-scan, multibeam and parametric sonar data, and have resulted in processing methods for classification, visualization, and quantitative characterization of the uppermost seabed stratum. Among these, two approaches have been implemented as stand-alone software packages for identification of bottom roughness, impedance and volume inhomogeneities by inversion of parametric sonar data. The first package, SIROB, relies on a model-based approach in which the forward model is the prediction code BORIS. Bottom parameters are identified by minimizing the data-prediction discrepancy in a generalized time-frequency domain. The second package, FARIM, uses the coherent part of the backscattered time signal. The frequency shift in the data is used for finding the roughness, and roughness estimate is used to correct the impedance estimate found from the returned echo. FARIM is simpler and faster than SIROB, but it determines a subset of the parameters identified by SIROB. Results obtained on field data show good agreement among both methods and available ground truth.

8:50

3aAO4. Techniques for estimating sediment properties from chirp sonar data. Steven G. Schock (Dept. of Ocean Eng., Florida Atlantic Univ., Boca Raton, FL 33431, schock@oe.fau.edu)

The chirp sonar is a towed wideband FM reflection profiler developed to collect normal incidence reflection data suitable for sediment property inversions; the sonar also generates detailed reflection profiles of the seabed. The chirp sonar system transmits FM pulses with high time-bandwidth products to attain high pulse energy; consequently, the acoustic data have a high signal-to-ambient-noise ratio after correlation processing. The vertical resolution of the images is about 1 cm when transducer bandwidth is approximately 40 kHz. Data processing methods have been developed to automatically map the locations of sediment layer interfaces, and to estimate compressional wave attenuation, acoustic impedance, and compressional wave velocity of sediment layers using normal incidence FM reflection data. The height of sand ripples and the phase dispersion of compressional waves can also be measured from the acoustic imagery. Field data collected by other investigators during the High Frequency Acoustics DRI experiments off Fort Walton Beach were used to verify the acoustic property estimates. Comparisons show that chirp sonar estimates of acoustic properties agree with direct measurements of the properties based on *in situ* acoustic probe data and sediment core data collected by investigators from Naval Research Laboratory, Stennis Space Center, and APL, Washington State.

Contributed Papers

9:15

3aAO5. Matched-field geoacoustic inversion in a range-varying waveguide. Chris A. Gillard,[a] David J. Thomson, Gordon R. Ebbeson, and Garry J. Heard (Defence Res. Establishment Atlantic, P.O. Box 1012, Dartmouth, NS B2Y 3Z7, Canada)

In recent years, geoacoustic inversion methods based on matched-field processing concepts have been devised for inferring the sub-bottom properties of shallow-water waveguides using measurements of acoustic signals received on hydrophone arrays. The development of these matched-field inversion (MFI) procedures is driven by two goals: (1) to obtain an accurate geophysical description of the morphology of the sub-bottom structure, and (2) to characterize the acoustic response of the sub-bottom so that its effect on sound propagation can be predicted. Knowledge of (2) can improve the effective ranges for using sound in underwater communication and/or source localization applications. Of practical interest is the question: How crude can our estimates of the sub-bottom structure be and still allow successful acoustic source localization using model-based signal processing methods? In this paper, a hybrid simplex simulated annealing MFI code [M. R. Fallat and S. E. Dosso, J. Acoust. Soc. Am. **105**, 3219–3230 (1999)] is combined with a high-order parabolic equation (PE) model and applied to geoacoustic inversion in a shallow-water waveguide whose sub-bottom properties vary with range. In particular, we investigate simple sub-bottom characterizations for some test cases considered recently at the Inversion Techniques Workshop held in Gulfport, MS.

[a] Presently on exchange from the Defence Science and Technology Organisation, Australia.

9:30

3aAO6. Range-dependent geoacoustic inversion. Anna-Liesa Lapinski, Stan Dosso, and Ross Chapman (School of Earth and Ocean Sci., Univ. of Victoria, Victoria, BC V8W 3P6, Canada)

This paper considers acoustic inversion for a minimum-structure geoacoustic model in a range-dependent environment. The general survey configuration consists of recording acoustic fields at a vertical array of sensors due to a number of sources distributed in range along a track. The unknown model consists of the sound speeds and thicknesses of sub-bottom layers, which vary with range in an arbitrary manner, and geometric parameters of the experiment. The goal is to determine a range-dependent geoacoustic model with the least structure that is consistent with the resolving power of the data. An under-parametrized approach is developed, which consists of solving the inverse problem a number of times, each time increasing the range variability allowed in the model. The optimal parametrization is subsequently obtained by examining the data mismatch and a norm of the model structure as a function of the number of parameters. The individual inversions are performed using an adaptive hybrid inversion algorithm applied to all data simultaneously.

9:45

3aAO7. Range-dependent seabed characterization by inversion of acoustic data from a towed receiver array. Martin Siderius, Peter L. Nielsen (SACLANT Undersea Res. Ctr., Viale San Bartolomeo 400, 19138 La Spezia, Italy), and Peter Gerstoft (Marine Physical Lab., Scripps Inst. of Oceanogr., Univ. of California, San Diego, CA)

The MAPEX2000 experiments were conducted by the SACLANT Undersea Research Centre in the Mediterranean Sea in March 2000 to determine the performance of acoustic inversion for seabed properties using a towed, horizontal receiver array. Towed systems are advantageous because they are easily deployed from a ship and the moving platform offers the possibility of estimating spatially variable (range-dependent) seabed properties. Previous research has successfully applied matched field processing (MFP) geoacoustic inversion techniques to measured acoustic data, however, in nearly all cases the inverted data were collected on moored, vertical receiver arrays. In the MAPEX2000 experiments acoustic data were collected on both towed and moored vertical receiver arrays. Results will be presented showing that seabed properties (e.g., seabed sound speed, sediment layer thickness and attenuation constant) can be extracted using MFP inversion of acoustic measurements from a towed array of receivers. These seabed properties agree with those inverted using data received simultaneously on a vertical array. These findings imply that a practical technique can be developed to map range-dependent seabed properties over large areas using a towed acoustic system. An example of such a range-dependent inversion is given for two sites from the MAPEX2000 experiments.

10:00–10:15 Break

10:15

3aAO8. Parameter coupling in broadband geoacoustic inversion. John S. Perkins, Michael D. Collins, Dalcio K. Dacol, Laurie T. Fialkowski, and Joseph F. Lingevitch (Naval Res. Lab., Washington, DC 20375)

For many geoacoustic inverse problems involving a single frequency data set, the search parameters are often strongly coupled to one another. This may mean that individual parameters are difficult to determine, but combinations of parameters can be resolved. In terms of the optimization process, the parameter landscape that is being searched has multidimensional valleys that hinder location of a global minimum. A coordinate rotation technique has been developed [J. Acoust. Soc. Am. **98**, 1637–1644 (1995)] to speed the search for optimal parameters. This technique also gives valuable information about the parameter hierarchy and parameter couplings. One approach to decoupling the parameters and obtaining realistic physical parameter values is to process broadband data or multifrequency data. A new fast forward modeling technique [J. Acoust. Soc. Am. **106**, 1727–1731 (1999)] makes it possible to consider broadband data for geoacoustic inversion. The coordinate rotation technique is combined with fast forward modeling to study parameter couplings for broadband and multifrequency problems. [Work supported by ONR.]

10:30

3aAO9. Geoacoustic inversion hybrid model. Mark S. Haire (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029) and David P. Knobles (Appl. Res. Labs., Univ. of Texas, Austin, TX 78713)

Simulated annealing (SA) or genetic algorithms are accepted optimization methods for geoacoustic inversion to obtain seabed parameters from measured acoustic data. These techniques permit optimizations where numerous local minima are present, but are computationally intensive. Gradient methods, such as the Levenberg–Marquardt (LM) approach, are extremely efficient at finding local minima, but typically are unable to locate the global minimum. To improve the computation speed and efficiency of SA, a ''hybrid'' algorithm combining SA and LM has been developed. The hybrid approach applies a LM algorithm after each temperature cycle of SA if an improved cost function value has been obtained during the cycle. The hybrid technique takes advantage of the fact that SA often examines a point that is in the proximity of the global minimum when the temperature is high. Factors of 3–8 increases in speed over SA alone are obtained with simulated data. Results will be presented from the application of the hybrid model to experimental data collected during the Area Characterization Test-I in the Gulf of Mexico. [Work supported by ONR.]

10:45

3aAO10. Seismo-acoustic inversion using artificial neural networks optimized with a genetic algorithm—direct inversion. Timothy Ruppel (Naval Res. Lab., Stennis Space Center, MS 39529, tim.ruppel@nrlssc.navy.mil)

Traditional inversion methods search the parameter space for a minimum in the difference between the observed acoustic field and that predicted by a forward acoustic model. While this method is well established and has exhibited excellent results, it requires that many time-consuming forward model runs be performed for each prediction. This paper will describe an inversion technique (direct inversion) which uses genetic algorithms to train a neural network before any observations are made. In essence, an inverse acoustic model is generated whose inputs are the acoustic field values and whose outputs are the environmental parameters of interest. This allows for rapid inversion calculations at the time of observation if sufficient time has been allowed for training. Some results inverting simulated and experimental acoustic signals will be given. [Work supported by ONR.]

11:00

3aAO11. Geoacoustic inversion using surface ship sources and HLA beam data. Robert A. Koch (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029), David P. Knobles, Tracianne B. Neilsen, and Mark S. Haire (Appl. Res. Labs., Univ. of Texas., Austin, TX 78713)

The focus of this research is the synthesis of spatial and temporal processing to perform simultaneous inversion for environmental parameters and to localize moving broadband sources in shallow water environments. The processed data are a sample of broadband cross spectra from either individual phones or subaperture beams for one or more FFT sequences. An inversion using simulated annealing to estimate both environmental and source parameters is performed using a cost function that measures the correlation between the sample of data cross spectra and corresponding modeled cross spectra. Thus, the data are processed coherently in frequency and time. For each iteration of the simulated annealing search, the cost function is computed on a grid of initial source positions for a specified source speed and course and for specified seabed geoacoustic parameters. Horizontal line array data collected in February 1998 in the Gulf of Mexico using SWAMI are processed, and the results for the geoacoustic parameters from the inversion using the exercise ship as the source are compared with previous results obtained from light bulb implosions. [Work supported by ONR.]

11:15

3aAO12. Chirp sonar inversion of geoacoustic properties of finely layered seabed. Altan Turgut and Stephen N. Wolf (Naval Res. Lab., Acoust. Div., Washington, DC 20375)

Chirp sonar reflection amplitude and phase data can be used for real-time inversion of sediment properties such as density, porosity, and sound-speed profiles by assuming homogeneous layering within the seabed [Turgut and Wolf, J. Acoust. Soc. Am. (1996)]. In case of fine layering, inversion results might be erroneous if the chirp pulse does not have enough bandwidth to resolve the layers. An alternative method, based on adjoint-state inversion techniques, is used to estimate sediment impedance fluctuations at wavelength scales. Both methods are tested using real and simulated reflection amplitude data. Simulated reflection amplitude data

were obtained by using a poro-viscoacoustic sediment model based on Biot's theory. Accuracy of the inversion can be improved when both methods are used in a complimentary manner. Limits of these two inversion techniques are also discussed and several recommendations are made to improve the data collection methods. [Work supported by ONR.]

11:30

3aAO13. Chirp sonar inversion results from SWAT East China Sea and New Jersey Shelf experiments. Altan Turgut (Naval Res. Lab., Acoust. Div., Washington, DC 20375), Dawn L. Lavoie, Douglas N. Lambert, Donald J. Walter (Naval Res. Lab., Stennis Space Center, MS 39529), and Kazuhiko Ohta (Fifth Res. Ctr., TRDI, Japan)

Subbottom surveys have been conducted during recent shallow water acoustic technology (SWAT) experiments to invert bottom geoacoustic properties at the East China Sea and New Jersey Shelf sites. Sediment properties such as density, porosity, and sound-speed profiles are inverted by using reflection amplitude and phase data obtained from a hull-mounted 2- to 12-kHz chirp sonar, a deep-towed 2- to 12-kHz chirp sonar, and a 30-kHz acoustic sediment classifier system. The attenuation coefficient is estimated using the frequency shift method which seems to be relatively insensitive to reflection and transmission effects, source-receiver beam patterns, and instrument responses. All three systems provided high-quality reflection data and inversion results are in agreement with those of sediment core measurements. [Work supported by ONR and JDA.]

11:45

3aAO14. Geoacoustic parameters inversion by an annealing-genetic algorithm. Ying-Tsong Lin, Yuan-Ming Chuang, and Chifang Chen (Dept. of Naval Architecture and Ocean Eng., Natl. Taiwan Univ., 73 Chou-Shan Rd., Taipei, Taiwan, ROC)

A matched-filed processing (MFP) with the annealing genetic algorithm is used to invert the geoacoustic parameters by matching acoustic field data with numerical replica field. While considering the MFP as an optimization problem, it is unavoidable that a huge amount of replica field will be needed for matching with the measured field data. Furthermore, due to complicated geoacoustic parameters, the inversion becomes difficult to deal with. The simulated annealing algorithm has the property of better local search ability, while the genetic algorithm might jump over local optimal and have the better ability in global search. By merging simulated annealing algorithm into genetic algorithm, an annealing genetic algorithm is developed for efficiency and accuracy. Several scenarios of ocean model are used as test cases. The geoacoustic parameters are accurately estimated. [Work supported by National Science Council of Republic of China.]

WEDNESDAY MORNING, 6 JUNE 2001 PDR 17, 8:00 A.M. TO 12:00 NOON

Session 3aBB

Biomedical Ultrasound/Bioresponse to Vibration and Signal Processing in Acoustics: Beamforming and Adaptive Aberration Correction

T. Douglas Mast, Chair
Applied Research Laboratory, Pennsylvania State University, University Park, Pennsylvania 16802

Chair's Introduction—8:00

Invited Papers

8:05

3aBB1. Acoustics in medical ultrasound beamformation. Kai E. Thomenius (GE Corporate R&D, KW-C300A, One Research Circle, Niskayuna, NY 12309, thomeniu@crd.ge.com)

Today's medical ultrasound scanners continue to push the envelope with efforts to improve imaging performance. Increasingly this demands more sophisticated application of acoustics in the design of the beamformers for these scanners. In this paper, limitations of present designs are reviewed by the use of acoustic field simulations. These design limitations include spatial sampling of the transducer aperture and the contributions of two-dimensional apertures and the impact of variations in the speed of sound in the various tissues. The role of beamformation in defining the likely scatterer populations that will be included in any beam will be discussed. The introduction of harmonic imaging into medical ultrasound has changed the landscape significantly especially in fields such as echocardiography. This has generated the need to understand the harmonic generation process of the propagating acoustic wave. Comparisons of the acoustic beams due to the fundamental frequency spectrum and the harmonic spectra are discussed along with means by which the scanners take advantage of them. From this base one can begin to identify areas of further improvement and speculate on their likely impact on the diagnostic process.

8:30

3aBB2. Synthetic array imaging for biomedical ultrasonics. Matthew O'Donnell (Biomed. Eng. Dept., Univ. of Michigan, Ann Arbor, MI 48109-2125)

Real-time array imaging with ultrasonics has been one of the primary tools of medical diagnostics for over two decades. Scanning systems based on phased array principles require a completely independent, programmable electronic channel for each element in the active aperture. Such complexity is acceptable for clinical systems propagating ultrasound across the body surface at frequencies ranging from 1 to 10 MHz. However, for high-frequency systems integrated into therapeutic devices entering the body through catheters or needle probes, system complexity must be minimized. Indeed, a full phased array approach is nearly impossible for most of these systems. An alternative method is to use synthetic imaging principles, greatly simplifying system and array complexity. There

are many practical problems in implementing synthetic imaging principles for real-time medical scanners. Theses issues, along with possible solutions, will be presented in this talk. In addition, a fully confocal, real-time intravascular synthetic array imaging system operating at ultrasound frequencies ranging from 10 to 30 MHz will be presented. Finally, methods for fully synthetic imaging systems operating at frequencies in excess of 50 MHz will be described, and prospects for implanting such systems in the human body will be discussed.

8:55

3aBB3. Ultrasonic imaging with aberration correction. Robert C. Waag (Ultrasound Res. Lab., Univ. of Rochester, Rochester, NY 14642, waag@ece.rochester.edu)

Measurement and simulation of ultrasonic wave front distortion produced by propagation through abdominal wall, breast, and chest wall show that the distortion varies with tissue type, scattering contributes significantly to distortion, and arrival time and energy level fluctuations can be large compared to precision in current ultrasonic imaging instrument beamformers. Investigation of focus degradation and compensation by measuring the transmit focus with a hydrophone and by reconstructing the time history of a virtual point source in each case without compensation and with various compensation techniques indicate that, while time-shift compensation in the aperture can improve the focus significantly, techniques capable of compensating amplitude and shape changes in waveforms can improve the focus even more. Pulse-echo studies of focusing that use estimates of aberration from random scattering show that both low transmit f-numbers and transmit compensation are important for accurate estimation of aberration. B-scan images formed through a water path, through tissue-mimicking aberration, and through the aberration using aberration correction demonstrate that images can be substantially improved by adaptive beamformation that employs estimation and correction of aberration on both transmit and receive.

9:20

3aBB4. Adaptive ultrasonic imaging using SONOLINE Elegra. D-L. Donald Liu, Pat Sutcliffe, John Lazenby, Bruce McDermott, Pat Von Behren, and Jin Kim (Siemens Medical Systems Ultrasound Group, 22010 SE 51st St., Issaquah, WA 98027)

Adaptive correction of the effects of propagation through inhomogeneous tissue is critical to the improvement of current ultrasonic imaging systems. Currently, estimation and correction of time-delay errors is more feasible than other more sophisticated approaches. Data acquisition, time-delay estimation and compensation have been implemented on the SONOLINE Elegra system using the system CPU, the Crescendo image processor, and the existing front-end electronics. Experimental results with this implementation will be reported. The effects of compensating the transmit beam is studied using the waveform similarity factor and single transmit imaging. On an RMI404 phantom plus a 1-D aberration layer with a rms time fluctuation of 40 ns and correlation length of 5 mm, the waveform similarity factor of randomly scattered waveforms improved from 0.362 to 0.477 by iteration. Correspondingly, the -20 dB lateral resolution improved from 1.62 to 0.77 mm, and the image contrast improved by 8.5 dB (speckle region is 6 dB brighter while echo-free region is 2.5 dB darker). Experiments with a 2-D aberration layer and with a special phase aberration phantom showed less image improvements. Preliminary body scans with adaptive imaging showed improved image contrast and details in some cases but the results are mixed and influenced by such factors as isoplanatic patch size and complex scattering structures. [Study partially supported by NIH R29 CA81688.]

9:45–10:00 Break

Contributed Papers

10:00

3aBB5. Effect of aperture size on compensation of arrival time distortion. James C. Lacefield and Robert C. Waag (Dept. of Elec. and Computer Eng., Univ. of Rochester, Rochester, NY 14627, lacefiel@ece.rochester.edu)

The effects of aperture size on focus distortion and time-shift compensation were investigated using a two-dimensional array system and a tissue-mimicking distributed aberration phantom that produces 65 ns root-mean-squared arrival time fluctuation. Pulse-echo data were acquired at 3 MHz using six apertures with areas ranging from 12.6×12.6 to 47.4×47.4 mm^2, each focused at 55 mm. Transmit beams produced with and without compensation for arrival time fluctuations estimated from random scattering data were measured using a hydrophone. Focus degradations were characterized using 25% deviation levels, i.e., the level below the peak at which the aberrator path beam becomes 25% broader than the water path beam. Without focus compensation, the 25% deviation level was above -6 dB in five of eight trials performed with the smallest aperture and seven of eight trials with the largest aperture. With compensation, the 25% deviation level dropped below -6 dB in five trials with the smallest aperture and six trials with the largest. The results demonstrate that larger apertures are more likely to experience a decrease in point resolution caused by aberration. Point resolution is restored by time-shift compensation more readily than contrast resolution, so that method is more useful with larger apertures.

10:15

3aBB6. Expansion of the isoplanatic patch for correction of ultrasonic aberration by assignment of corrections based on similarity of propagation paths. Wayne C. Pilkington and Robert C. Waag (Ultrasound Res. Lab., Univ. of Rochester, Rochester, NY 14642, pilkingt@ece.rochester.edu)

Aberration correction for ultrasonic b-scan images can be effectively implemented with less computation by employing cross correction, i.e., using a single set of aberration corrections computed for a reference aperture position to correct the transmit and receive focus at other points within an image region defined as the isoplanatic patch. An uncomplicated implementation of this approach compensates the signal at each location in the aperture with the aberration correction from the corresponding location in the reference aperture regardless of the aperture position. Alternatively, the signal at each location in the aperture may be compensated using the correction for the reference aperture signal with a propagation path most similar to the propagation path of the signal being corrected. This latter approach requires a different map of the reference aberration corrections for each aperture position, but the map can be implemented by simple shift operations. Measurements using a novel 2-D array system demonstrate that the assignment of corrections based on similarity of propagation paths can significantly improve the effectiveness of cross correction and enlarges the effective size of the isoplanatic patch.

10:30

3aBB7. Aberration-corrected time-domain ultrasound diffraction tomography. T. Douglas Mast (Appl. Res. Lab., Penn State Univ., University Park, PA 16802)

The inverse scattering problem of reconstructing a spatially dependent sound speed variation from far-field time-domain acoustic scattering measurements is considered. Such reconstructions are quantitative images with applications including ultrasonic mammography. Although the linearized time-domain inverse scattering problem is shown to have no general solution for finite signal bandwidth, an approximate solution to the linearized problem is constructed using a simple delay-and-sum method analogous to ''gold standard'' ultrasonic beamforming. The form of this solution suggests that the full nonlinear inverse scattering problem can be approximated by applying appropriate angle- and space-dependent time shifts to the time-domain scattering data; this analogy leads to a general approach to aberration correction. Two related methods for aberration correction are presented: one in which delays are computed from estimates of the medium using an efficient straight-ray approximation, and one in which delays are applied directly to a time-dependent linearized reconstruction. Numerical results indicate that these correction methods achieve substantial quality improvments for imaging of large scatterers. [Work supported by the Breast Cancer Research Program of the U.S. Army Medical Research and Materiel Command.]

10:45

3aBB8. Theoretical considerations for the use of microbubbles as point targets for phase aberration correction. Dimitris Psychoudakis (Elec. Eng. and Computer Sci., Univ. of Michigan, Ann Arbor, MI 48109-2122), J. Brian Fowlkes, John L. Volakis, Oliver D. Kripfgans, and Paul L. Carson (Univ. of Michigan, Ann Arbor, MI 48109-0553)

Bubbles can be produced by vaporization of perfluorocarbon droplets of a few μm diameter. These bubbles can reach up to 100 μm in diameter and their backscatter is calculated to be more than 10 dB above that of several organ tissues. At these sizes and for diagnostic frequencies (2–8 MHz), bubbles can be approximated by the nonrigid sphere scattering solution employed here. This presentation concerns the bubble size and its implications on the backscatter amplitude and the phase error introduced in diagnostic ultrasound when assuming that the bubble acts as a point target for phase aberration correction. The phase error is the difference between the phase at each location along the receiving aperture relative to that at the aperture center, compared with the same relative phase for a perfect point target. Evaluations were made of the phase error with respect to a range of transducer f-numbers (0.5–2.0) for a specific bubble size (30 μm radius) and at certain frequencies (2–8 MHz). For example, at 5 MHz the phase error introduced by the point target assumption is maximally 5 deg, while the phase error of breast tissue scattering is around 160 deg. [Work supported by PHS Grant No. R01HL54201 from the National Heart, Lung, and Blood Institute.]

11:00

3aBB9. Comparison between time reversal and spatio-temporal inverse filter application to focusing through a human skull. Mickael Tanter, Jean-Francois Aubry, Jean-Louis Thomas, and Mathias Fink (Laboratoire Ondes et Acoustique, ESPCI, Paris VII Univ., Paris, France)

Ultrasonic imaging systems capabilities are strongly dependent on the focusing quality of the ultrasonic beam. In the case of brain imaging, the skull strongly degrades the ultrasonic focusing pattern by introducing strong phase and amplitude aberrations of the wave-front. In previous work, this degradation of the beam focus had been partially corrected by coupling the time reversal focusing process to an amplitude compensation of the emission signals. In that case, the optimal focus was reproduced down to -20 dB, but the sidelobe level remained at about -25 dB. This elegant technique will be compared to another focusing technique recently developed in our laboratory, called spatio-temporal inverse filtering. Thanks to this method, based on the inversion of the propagation operator at each frequency within the bandwidth of our transducers, experimental focusing through the skull is now comparable to optimal focusing in a homogeneous medium. Those two methods not only differ theoretically, but also suffer differently from all the experimental defects, such as the limited bandwidth of the transducers or the limited aperture of the arrays. A comparson of the results obtained with both techniques in water and through a human skull clearly highlights the advantages and the drawbacks of each method.

11:15

3aBB10. Integrated matrix arrays. Ken Erikson, Jason Stockwell, and Robert McPhie (BAE Systems, P.O. Box 868, Nashua, NH 03061-0868)

Improved image quality requires the use of matrix ($n \times m$) arrays with a thousand or more elements. As element numbers increase and their dimensions grow smaller, limitations to present fabrication technologies arise. Cost, ergonomics, produceability, and reliability are important issues. Signal loss due to the capacitance of interconnecting coax cables becomes a fundamental problem. Connecting an integrated circuit directly to the array elements alleviates all these problems. Each unit cell of such a custom transmitter/receiver integrated circuit (TRIC) may have high voltage switches for transmitting; a preamplifier which minimizes signal loss due to capacitance in coax cables and a multiplexer to send the array signals over fewer wires. Additional signal processing and beam forming may also be included. Issues with currently available arrays are reviewed. The new technology for direct connection of arrays to IC's is described. The paper concludes with speculation about future possibilities of this approach.

11:30

3aBB11. Measurements of the spatial coherence of the fundamental and second-harmonic beams for a clinical imaging system. Russell J. Fedewa, Kirk D. Wallace, Mark R. Holland (Lab. for Ultrason., Dept. of Phys., Washington Univ. of St. Louis, One Brookings Dr., St. Louis, MO 63130), James R. Jago, Gary C. Ng, Matthew R. Rielly, Brent S. Robinson (ATL Ultrasound, Bothell, WA 98041-3003), and James G. Miller (Lab. for Ultrason., Washington Univ. of St. Louis, St. Louis, MO 63130)

Spatial coherence of backscattered signals underlies correlation-based phase aberration corrections. The van Cittert–Zernike theorem relates frequency-independent spatial coherence to the autocorrelation of the transmit apodization. Previous studies suggest that the mainlobe of the nonlinearly generated harmonic beam is wider and exhibits lower sidelobes than a beam linearly generated at the harmonic frequency. The objective of this study was to measure the spatial coherence associated with fundamental and nonlinearly generated harmonic beams. Using data experimentally acquired from a clinical scanner (ATL HDI5000), two independent methods were employed to measure the spatial coherence. One approach measured the spatial coherence of backscatter from a tissue-mimicking phantom using rf signals from individual elements of a linear array. In the second approach, the effective apodization was determined by a linear angular spectrum backpropagation of hydrophone-sampled data from a transverse plane in the focal zone. The results show that the effective apodization of the nonlinearly generated harmonic beam is more aggressive than the actual transmit apodization. The spatial coherence associated with the second-harmonic beam differs from the spatial coherence of the fundamental beam, but is predicted by the effective apodization. [Supported in part by NIHHL40302 & ATL.]

11:45

3aBB12. Efficient computation of field of 2-D array with limited diffraction array beams. Jian-yu Lu and Jiqi Cheng (Ultrasound Lab, Dept. of Bioengineering, The Univ. of Toledo, Toledo, OH 43606, jilu@eng.utoledo.edu)

Two-dimensional (2-D) arrays are useful for improving quality of three-dimensional (3-D) medical imaging in ultrasound. Beams produced with a 2-D array are usually simulated with the Rayleigh–Sommerfeld diffraction formula (RSDF). In general, the RSDF requires a 2-D integra-

tion for each field point in space and thus is very time consuming. Fresnel approximation may reduce the 2-D integration to 1-D but will not yield satisfactory results near the transducer surface or for field at a large angle from the beam axis. In this work, limited diffraction array beams are used to synthesize beams produced with a 2-D array [J.-Y. Lu, Int. J. Imaging Syst. Technol. **8**, 126–136 (1997)]. In this method, the 2-D integration is replaced with a 2-D summation leading to a much faster computation. The method is accurate even if the field to be evaluated is very close to the surface of a transducer. Results of Bessel beams, *X* waves, and focused Gaussian beams will be shown and compared with those obtained with the RSDF and the experiments. [This work was supported in part by Grant No. HL60301 from the National Institutes of Health of the U.S.A.]

WEDNESDAY MORNING, 6 JUNE 2001 PARLOR B, 8:00 TO 10:05 A.M.

Session 3aEA

Engineering Acoustics: Acoustic Devices and Systems

Elizabeth A. McLaughlin, Chair
Naval Undersea Warfare Center, Code 2132, 1176 Howell Street, Newport, Rhode Island 02841-1708

Chair's Introduction—8:00

Contributed Papers

8:05

3aEA1. A PVDF long time-constant force sensor. Aaron M. Foulk and W. Jack Hughes (Appl. Res. Lab., Penn State Univ., P.O. Box 30, State College, PA 16804, wjh2@psu.edu)

A low-profile large-area underwater force sensor was desired to measure quasistatic forces on a rigid structure. The design for a PVDF sensor (a low-profile piezoelectric) intended for use in measuring sustained (10-s) impact forces with minimal error is presented. The final sensor design integrates a PVDF bimorph piezoelectric element and long time constant charge amplifier electronics into a 1-in. diam by 3/8-in. waterproof package. This work addresses design issues including electronics design to maximize the useful time constant, piezoelectric material selection, thermal design and simulation to minimize pyroelectric effects and noise minimization efforts. [This work was supported by NSWC-CD and ONR.]

8:20

3aEA2. Modulated piezoresistive sensor for airborne infrasound. Thomas B. Gabrielson (Appl. Res. Lab., The Penn State Univ., P.O. Box 30, State College, PA 16804)

This paper will describe the design, construction, and testing of a piezoresistive sensor for airborne acoustics in the frequency range 0.005–100 Hz. Intended as a candidate for many-element infrasound arrays where element cost is critical, the sensor was designed around an inexpensive silicon-membrane chip with an implanted piezoresistive bridge. The prototype has a measured self-noise below one millipascal per root hertz at 1 Hz with a $1/f$ power spectrum to 0.005 Hz, the designed low-frequency rolloff. The responsivity is nominally 20 mV/Pa with a linear signal range to 150 Pa. When the bridge is operated with high-frequency drive and synchronous detection, the self-noise is within a factor of 2 of the intrinisic Johnson noise of the resistive bridge. This self-noise level is well below ambient levels in the microbarometric region from 0.2 to 0.7 Hz. Basic sensor-element limitations will be discussed along with the electronics and package design. In addition, the techniques for calibration and self-noise evaluation will be presented. [Work sponsored by the U.S. Army Space and Missile Defense Command.]

8:35

3aEA3. The determination of the depth to the liquid-gas interface in a producing well. Elmer L. Hixson, Augasto L. Podio, and Fernando Garcia-Osuno (Univ. of Texas, Austin, TX 78712)

Many oil wells use electrically driven pumps at the bottom of the well to pump oil to the surface. If the oil-gas interface falls below the pump, gas is pumped, the pump overheats and usually fails. A passive acoustic system is described here that uses the annulus between the production tubing and the cemented steel casing as a closed end gas filled tube. Pump noise excites many longitudinal resonant modes. These modes occur at *n*-half wavelengths of the tube length. The depth is then related to the sound speed in the gas and the difference frequency between modes as measured by a microphone at the top of the well and a sprectum analyzer. When the depth falls below a critical value as indicated by a minimum frequency, the pump is turned off.

8:50

3aEA4. A new hand-held meter for measuring the occlusion effect produced by, and leakage around, earmolds. Mead C. Killion and Jack Goldberg (Etymotic Research, 61 Martin Ln., Elk Grove Village, IL 60007, abonso@aol.com)

Three of the most common complaints by hearing aid wearers are (a) difficulty hearing in noise, (b) the occlusion effect (hollow voice sound), and (c) feedback. The latter two can be largely controlled by earmold construction. What has been lacking is a simple way to measure earmold performance with regard to those two properties. A new hand-held occlusion effect meter using a 0.4-mm-thick crescent-shaped probe tube will be described.

9:05–9:20 Break

9:20

3aEA5. Research into the comfortable volume control curve of TV in Korean style apartments. Ki Duk Kim (Quality and Reliability Lab., Daewoo Electron. Co., Ltd., 412-2, Chongchun2-Dong, Pupyung-Ku, Inchon, Korea, soundexpert@korea.com), Chul Whan Kim, Hoon Ki Choi, and Dong Su Ryu (Daewoo Electron. Co., Ltd., Pupyung-Ku, Inchon, Korea)

We occasionally experience anger because of unsuitable TV volume for comfortable and clear listening. We do not concentrate on TV while annoyed that we are not tuned in at the desired sound volume. Therefore, we made the preference volume width in watching TV with consideration

of age, sex, time periods, variety of broadcasting programs, and watching space structure. We determined the variable sound volume control step based on jnd (just noticeable difference). On the basis of the data obtained, this paper is the research into the preferential sound while watching TV and the optimum TV volume curve.

9:35

3aEA6. Audio analysis method of a car stereo. Ki Duk Kim, Chul Whan Kim, Hoon Ki Choi, Dong Su Ryu (Quality and Reliability Lab., Daewoo Electron. Co., Ltd., 412-2, Chongchun2-Dong, Pupyung-Ku, Inchon, Korea, soundexpert@korea.com), Chul Whan Kim, Hoon Ki Choi, and Dong Su Ryu (Daewoo Electron. Co., Ltd., Pupyung-Ku, Inchon, Korea)

Presently, a car stereo system plays an important part in choosing a car, as people spend more time in their cars and pursue comfortable space, including audio. Thus, the sound quality analysis of a car becomes more conspicuous. We made an analysis of acoustic characteristics with 4-channel loudspeakers of three car types and carried out a subjective listening test with a test reference CD based on sound factors using a scale of seven. This paper proposes how the sound reproduction characteristics are changed according to position of loudspeakers, interior space of the cars, and the characteristics of sound reproduction equipment, in particular, a delay time and frequency response. Moreover, we examined the relation between measurement data and human listening sense.

9:50

3aEA7. Evaluation of relative sensitivity of hydrophone with varying hydrostatic pressure employing high pressure tube assembly. Janardan Singh (Underwater Acoust. (Ultrason. Group) Natl. Physical Lab., Dr. K. S. Krishnan Rd., New Delhi 110012, India)

The high pressure tube assembly employing active impedance termination method has been found most suited to calibrate deep submergence underwater transducers of low frequencies which cannot be calibrated using conventional anechoic tanks because of fabrication engineering problems as well as multiplying factors of cost. A high pressure tube assembly (12.5 cm internal diameter and 200 cm length) has been employed for the measurements of relative sensitivity of hydrophones with varying hydrostatic pressure. The suitable reference projectors were fabricated which were placed at the end of tube assembly as termination transducers. Four hydrophones as probe hydrophone and Bruel & Kjaer hydrophone (B&K-8103) for the relative sensitivity measurement have been suitably rigged in the transducer carriage. Free field condition is created with active impedance termination method which is achieved by precise adjustment of amplitudes as well as phases of termination transducers. A method of achieving the free field condition has been evolved which minimizes hit and trial procedure. The variations of the relative sensitivity of a hydrophone (B&K-8103) with hydrostatic pressure up to pressure of 1000 psi have been measured and the values were found close to the values specified (up to 400 meters of ocean depth) by Bruel and Kjaer.

WEDNESDAY MORNING, 6 JUNE 2001 CRYSTAL ROOM, 8:30 TO 11:40 A.M.

Session 3aMU

Musical Acoustics: New Synthesis Techniques I

James W. Beauchamp, Chair

School of Music, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, 2136 Music Building, 1114 West Nevada, Urbana, Illinois 61801

Chair's Introduction—8:30

Invited Papers

8:35

3aMU1. Real-time additive synthesis with envelope parameter stremas: an alternative to sampling. Lippold Haken, Kelly Fitz, and Paul Christensen (Cerl Sound Group, c/o Lippold Haken, ECE Dept., 358 Everitt, 1406 W. Green St., Urbana, IL 61801, l-haken@cerlsoundgroup.org)

A sampling synthesizer uses a set of recordings for sound synthesis. A shortcoming of the sampling technique is the inability to produce complex spectral changes associated with the performer's actions during a note. Additive synthesis, on the other hand, represents each sound as a collection of sine wave components, or partials. A wide variety of modifications are possible with additive synthesis, including frequency shifting, time dilation, cross synthesis, and sound morphing. Our real-time additive synthesis uses a set of recordings, like a traditional sampling synthesizer. Unlike a traditional sampling synthesizer, our additive synthesis models timbres as collections of bandwidth-enhanced partials (sine waves with noise) with time-varying parameters. The timbres are encoded in envelope parameter streams that can be manipulated in real time. These envelope parameter streams, which provide amplitude, frequency, phase, and bandwidth information for each partial, are the counterpart of sample streams in sampling synthesis. The talk will include demonstrations using the continuum fingerboard.

9:00

3aMU2. Note-by-note synthesis is not enough. Roger B. Dannenberg (School of Computer Sci., Carnegie Mellon Univ., Pittsburgh, PA 15213)

Traditional analysis and synthesis of musical tones are based on the idea that music can be deconstructed into note units that can be individually synthesized and recombined to create music. This ''divide and conquer'' strategy is natural for scientists and engineers, but it does not always work well in practice. In particular, brass instrument tones in a musical context are quite different from brass tones produced in isolation. Synthesis methods that are capable of reproducing individual tones fail to sound realistic when these tones are sequenced to produce a musical phrase. The combined spectral interpolation synthesis (CSIS) method uses a fairly

simple spectrally based model in which the dynamically changing spectrum is a function of fundamental frequency and rms amplitude. The spectrum function is determined by measuring actual acoustic performances. The synthesis model must be controlled by appropriately evolving frequency and amplitude controls. A rule-based system computes these controls from a machine-readable version of a musical score, which includes slurs and other articulation markings. The rules take musical context and phrasing into consideration. The combination of musical context and a good synthesis model results in greatly improved synthesis.

9:25

3aMU3. A hybrid additive-wavetable synthesis model for the horn. Andrew Horner and Lydia Ayers (CS Dept., HKUST, Clear Water Bay, Kowloon, Hong Kong)

Is it possible to synthesize the French horn well enough to pass for the real thing? Most commercial synthesizers and soundcards are limited to producing timbres that are ''horn-like'' or ''clarinet-like.'' It is much more difficult for a model to be flexible enough to allow different types of articulations, phrasings, and tone colors. This work presents a hybrid additive-wavetable synthesis model for the horn using genetic algorithm optimization. The parameters for the model include fundamental frequency, overall amplitude, vibrato amount, attack and release times, brightness, and articulation shape. The model gets brighter as the amplitude increases, and the brightness decreases as the amplitude decreases by scaling each partial with a set of exponentially related amplitude envelopes. The model is implemented in the Csound software synthesis language. After implementing the model, several solos and orchestral excerpts that are commonly required in auditions were synthesized. The model is particularly effective on Baroque and classical excerpts.

9:50–10:00 Break

10:00

3aMU4. Implementation of loudness in a digital instrument. Hans G. Kaper (MCS Div., Argonne Natl. Lab., Argonne, IL 60439) and Sever Tipei (Univ. of Illinois, Urbana, IL 61801)

Loudness is a psychological term used to describe the magnitude of an auditory sensation. The loudness of a sound depends not only on the intensity of the sound and its physical composition (the partials in the sound and their amplitudes and frequencies), but also on the subjective conditions of the listener. In this talk, we discuss the algorithms implemented in DISCO to compute the loudness of a synthesized sound. DISCO (Digital Instrument for Sonification and Composition) is an additive synthesis program that produces sound files from score files. Score files are generated by an editor, which implements the composer's specifications of a piece. DISCO, written in C++, is being developed jointly by the authors at the University of Illinois at Urbana-Champaign (Computer Music Project) and Argonne National Laboratory (Mathematics and Computer Science Division). The program can be used for music composition, as well as for the auditory display of complex data sets—for example, in scientific computing (scientific sonification). The loudness routines in DISCO enable the user to design complex sounds to specified loudness. Additional routines are designed to prevent overflow in the sound file, which would cause ''clipping'' upon playing, while preserving the relative loudness of all sounds.

10:25

3aMU5. Vowel representation and synthesis of singer identity. Maureen Mellody (Appl. Phys. Prog., The Univ. of Michigan, Ann Arbor, MI 48109), Mark A. Bartsch (The Univ. of Michigan, Ann Arbor, MI 48109), Freda Herseth, George I. Shirley (School of Music, The Univ. of Michigan, Ann Arbor, MI 48109), and Gregory H. Wakefield (The Univ. of Michigan, Ann Arbor, MI 48109)

Professional singers shape the acoustical attributes of their vocal production to create the perceptual impression of a single instrument spanning a variety of vowel timbres over a wide range of pitch. Samples of the five Italian vowels over a two-octave range were analyzed using the modal distribution to yield a time-frequency image of the signal. From such images (high-order) representations of the composite transfer function of the signal were extracted. These representations provide little evidence for the existence of an acoustic ''signature'' that is invariant over the singer's entire range of production. Rather, the data, along with results from a perceptual study, suggest that singers create the impression of a single instrument by smoothly transitioning from one local region of invariance to the next. By identifying these regions, low-order, perceptually robust approximations are constructed which bear the identity of the singer when used to synthesize new instances of sung vowels. These low-order representations also serve as endpoints for interpolating between two different singers, as well as boundary points for extrapolating a singer from chest to head modes of production. [Work supported by the MusEn Project at the University of Michigan.]

10:50

3aMU6. Scanned synthesis. Bill Verplank, Max Mathews, and Rob Shaw (CCRMA, Dept. of Music, Stanford Univ., Stanford, CA 94305-8180, verplank@ccrma.stanford.edu)

''Scanned synthesis'' is done by scanning the slowly varying shape of an object and converting this shape to samples of a sound wave. The shape of the object is determined by the dynamic reactions of the object to forces applied by the performer. These forces vary at ''haptic'' rates (0–20 Hz). If the scanning path is closed, the sound wave is quasiperiodic and a fundamental pitch is perceived at the cycling frequency (20 Hz–20 kHz). Scanned synthesis provides direct dynamic control by the performer over the timbre of sounds as they are produced. The object can be real or simulated. With finite-element models, we have simulated the one-dimensional wave equation for a generalized slowly vibrating string. Timbres generated by manipulating the string at haptic rates are perceived as having a very pleasing live quality caused by the continually changing spectrum. To achieve additional richness, the performer can change the properties of the string in time and over the length of the string. [Work supported by Interval Research and CCRMA.]

11:15

3aMU7. Introduction to pulsar synthesis. Curtis Roads (CREATE, Dept. of Music and Media Arts and Technol. Prog., Univ. of California, Santa Barbara, CA 93106)

Pulsar synthesis (PS) is a method of electronic music synthesis based on the generation of trains of sonic particles. PS can produce either rhythms or tones as it criss-crosses perceptual time spans. The basic method generates sounds similar to vintage electronic music sonorities, with several important enhancements. The advanced method combines multiple pulsar trains and convolution with sampled sounds. Together with Alberto de Campo, the author has designed a program for pulsar synthesis called PulsarGenerator. Applications of pulsar synthesis in compositions by the author are noted.

11:40–12:00

Hands-on demonstration of continuum fingerboard

WEDNESDAY MORNING, 6 JUNE 2001 WABASH ROOM, 8:30 TO 10:05 A.M.

Session 3aNSa

Noise: Aircraft and Other Outdoor Noise Sources

Angelo J. Campanella, Chair
Campanella Associates, 3201 Ridgewood Drive, Columbus (Hilliard), Ohio 43026

Chair's Introduction—8:30

Contributed Papers

8:35

3aNSa1. Noise analysis of a new flight procedure for regional jets at Logan Airport. Nancy S. Timmerman (Massport, Noise Abatement Office, One Harborside Dr., Ste. 200S, East Boston, MA 02128-2909, ntimmerm@massport.com)

During the second half of 2000, the FAA Boston Tower tested a new approach path for regional jets to Logan International Airport's Runway 4L. The path was the same as that currently in use for turboprops. In this study, complaints, flight tracks, and noise levels were compared between the test period and previous time periods. Noise levels were reviewed at four of Massport's permanent noise monitors for three types of regional jet and one type of turboprop. Maximum levels of regional jets were slightly higher than the turboprop investigated for comparable operations.

8:50

3aNSa2. The effect of atmospheric gradients on aircraft noise contours. Kenneth J. Plotkin (Wyle Labs., 2001 Jefferson Davis Hwy., Ste. 701, Arlington, VA 22202)

Calculation of airport noise contours generally ignores atmospheric gradients. This is usually justified by the assumption that any effects will average out in the long run. An initial study has been conducted to evaluate this assumption. Noise contours have been computed for a single takeoff and landing of an F-16C under various atmospheric conditions. Atmospheric conditions included clear daytime, cloudy daytime, and nighttime temperature profiles, with winds at several speeds and directions. The heuristic propagation algorithms of L'Esperance *et al.* [AIAA Pap. 93-4405 (1993)] were incorporated into Wyle Laboratories NMSIM single-event noise simulation model. Calculated SEL footprints exhibited substantial changes in shape, particularly in sideline regions under crosswind conditions. Some trends of increased area under windy conditions were seen. Application of this analysis to a long-term set of actual weather data would validate (or invalidate) the ''it all averages out'' assumption. [Work supported by USAF Armstrong Laboratory. Atmospheric profiles and propagation routines provided by Andre L'Esperance.]

9:05

3aNSa3. Noise from aircraft at a condominium complex near Logan. Nancy S. Timmerman and Frank N. Iacovino (Massport, Noise Abatement Office, One Harborside Dr., Ste. 200S, East Boston, MA 02128-2909)

Massport undertook an aircraft noise survey at a condominium complex which is less than two miles from one of Logan International Airport's departure runways. The study took place from September 28 to October 19, 2000, at a time of year when the runway is commonly used. Aircraft operational and flight track information were obtained from Massport's Noise Monitoring System. Noise data collected at the survey location, which included both statistical and event data, were compared with two permanent monitors from that system. Results at this complex were consistent with the other data.

9:20

3aNSa4. Noise control of a compressor station. Halimat I. Alabi, Tim Krutz, Mats Jennische, Chiew Seng Goh (218-05 Nimitz Dr., West Lafayette, IN 47906, alacran@africana.com), Benjamin Dominik, and Justin Stricula (West Lafayette, IN 47906)

The noise emissions from a compressor station, part of a wastewater treatment plant, were investigated. The noise was a source of annoyance to a nearby residential subdivision. Initial measurements made on the site revealed salient tonal components. Further measurements suggested that the two main sources were leakage around the pipe intakes, and flow noise from the exhaust mufflers. Modifications were made to both sources in order to mitigate the problem. The leaks were sealed using dense caulking materials, and spoilers were inserted in the mufflers. The impact of these noise control solutions was measured, and the benefits to the surrounding community were assessed.

9:35

3aNSa5. Unusual experiences of two acoustical consultants. Marco Peres (Bruel & Kjaer, Rua Jose de Carvalho 55, 04714-020 Sao Paulo SP, Brazil, masperes@uol.com.br) and David Larson (S&V Solutions, Inc., Sycamore, IL 60178)

Some unusual projects worked on by two acoustical consultants will be shared with the audience along with the sound and vibration measurement techniques employed to solve the problems. A contrast will be drawn relating the paradox between the initial project objectives and high-tech instrumentation employed with the expectations of both the client and consultant. The presentation will be a great opportunity to share some of these unusual experiences with the audience and to reinforce the comic side of some of the situations we as acoustical consultants sometimes end up facing. Noisy ghosts, measurement of sound in the desert, an air horn that produces 125 dB(A) at 30 m (100 ft), and renting Lear jets for use as pink noise sources are among a few of the unusual projects that will be shared. A serious side will be presented for each project as well, primarily consisting of the application of modern noise measurement equipment (with sample data plots). The authors are sure the presentation will be both enlightening and entertaining.

9:50

3aNSa6. Noise generation by rotating blades in a riding lawnmower. Christian Skinner and Courtney B. Burroughs (Grad. Prog. in Acoust., The Pennsylvania State Univ., State College, PA 16802)

Measurements presented at the Acoustical Society of America Meetings in Newport Beach, CA in December 2000 showed that the dominant source of noise radiated from our riding lawnmower is the mower deck and that most of this noise is radiated directly as airborne noise from under the mower deck. In this paper, results from measurements of unsteady pressures on the tips of mower blades and baffling under the deck are presented. Frequency spectra are presented along with fluctuations in the pressure versus blade position. Measurements were conducted on three types of blades (flat, standard, and mulching) located at each of the three blade positions under the deck. To quantify the affects of the interaction between blades, measurements were conducted with one instrumented and two noninstrumented blades installed, and with only the instrumented blade installed. Results from these measurements are presented and discussed.

WEDNESDAY MORNING, 6 JUNE 2001 WABASH ROOM, 10:30 TO 11:50 A.M.

Session 3aNSb

Noise: Active Noise Control

Angelo J. Campanella, Chair
Campanella Associates, 3201 Ridgewood Drive, Columbus (Hilliard), Ohio 43026

Chair's Introduction—10:30

Contributed Papers

10:35

3aNSb1. Multichannel active noise control using numerically robust recursive least-squares algorithms. Martin Bouchard (School of Information Technol. and Eng., Univ. of Ottawa, 161 Louis Pasteur, Ottawa, ON K1N 6N5, Canada)

Recursive least-squares (RLS) algorithms and fast-transversal-filters (FTF) algorithms were recently introduced for multichannel active noise control (ANC) systems. It was reported that these algorithms can greatly improve the convergence speed of ANC systems using adaptive FIR filters, compared to steepest descent algorithms or their variants. However, numerical instability of the algorithms was an issue that needed to be resolved. In this presentation, extensions of stable realizations of recursive least-squares algorithms such as the inverse QR-RLS and the QR decomposition least-squares-lattice (QRD-LSL) algorithms are first introduced for multichannel ANC. A first set of simulations will verify that these algorithms have indeed a better numerical stability than the previously published recursive least-squares ANC algorithms. The case of underdetermined ANC systems (i.e., systems with more actuators than error sensors) is then considered, to show that in these cases it may be required to use constrained algorithms in order to have numerical stability. Constrained least-squares algorithms for multichannel ANC systems are therefore introduced for two types of contraints: minimization of the actuator signals power and minimization of the adaptive filter coefficients squares. A second set of simulations will verify the stabilized behavior of the constrained algorithms.

10:50

3aNSb2. New training algorithms for nonlinear active control of sound and vibration using neural networks. Martin Bouchard (School of Information Technol. and Eng., Univ. of Ottawa, 161 Louis Pasteur, Ottawa, ON K1N 6N5, Canada)

In recent years, a few articles describing the use of neural networks for nonlinear active control of sound and vibration were published. Using a control structure with two multilayer feedforward neural networks (one as a nonlinear controller and one as a nonlinear plant model), steepest descent algorithms based on two distinct gradient approaches were introduced for the training of the controller network. The two gradient approaches were sometimes called the filtered-x approach and the adjoint approach. Some recursive-least-squares algorithms were also published. In this presentation, a heuristic procedure is introduced for the development of recursive-least-squares algorithms based on the filtered-x and the adjoint gradient approaches. This leads to the development of new recursive-least-squares algorithms for the training of the controller neural network in the two networks structure. These new algorithms produce a better convergence performance than previously published algorithms. Differences in the performance of algorithms using the filtered-x and the adjoint gradient approaches are discussed in the paper. The computational load of the algorithms discussed in the paper is evaluated for multichannel systems of nonlinear active control. Simulation results are presented to compare the convergence performance of the algorithms, showing the convergence gain provided by the new algorithms.

11:05

3aNSb3. Development of a high bandwidth Hartmann–Whistle type actuator. Samer Khanafseh, Ganesh Raman (MMAE Dept., Illinois Inst. of Technol., 10 W. 32nd St., Chicago, IL 60616), and Alan Cain (ITAC, Chesterfield, MO 63006)

There is current interest in using Hartmann–Whistle type powered resonance tube actuators for aeroacoustic control. Previous experiments (Raman *et al.* (2000, 2001) AIAA 2000-1930; Stanek *et al.* (2000) AIAA 2000-1905) have shown that such actuators with no moving parts are very effective in the suppression of flow impingement and cavity noise. However, fixed geometry actuators can only produce effective actuation over a small range of frequencies. Our idea is to use a variable depth powered resonance tube actuator where the depth of the resonance tube and the distance between the supply jet and the resonance tube can be controlled actively during the experiment. The experimental setup consists of a 1/4″ diameter jet facing the open end of the resonance tube that is closed at the other end. The tube depth is changed by a computer controlled piston fitted inside the tube. Microphone and pressure sensors located at various locations in the vicinity of actuation were used to measure the frequency and amplitude of the fluctuations produced by this device. Due to the high amplitudes of pressure within the resonance tube the linear eigenfrequencies do not predict the measured frequencies. [Work supported by AFOSR.]

11:20

3aNSb4. Linear arrays of powered resonance tube actuators for aeroacoustic control. John Keidaisch, Ganesh Raman (MMAE Dept., Illinois Inst. of Technol., E1, 10 W. 32nd St., Chicago, IL 60016), and Valdis Kibens (The Boeing Co., St. Louis, MO 63166)

Experiments were conducted to characterize the performance of a linear array of powered resonance tube actuators developed jointly by Boeing and IIT. The actuator is an adaptation of the Hartmann whistle and devices of this type have been shown to be effective in suppressing flow generated acoustic tones that commonly occur in aircraft applications, such as impingement tones and cavity resonances (Raman *et al.*, AIAA Paper 2000-1930; Stanek *et al.*, AIAA Paper 2000-1905). The goal of this characterization is to provide a database that can be used to compare the performance of these actuators with other actuator designs (such as piezoelectric actuators) and to optimize the design of these actuators. Optimization involves modifying the geometric parameters of the actuator in order to get the maximum acoustic control with minimum input mass flow rate. In order to accomplish these tasks, detailed measurements of the fluctuating velocity and pressure were made at numerous locations in the flowfield at the actuator exit. Results indicate that the actuation could produce frequencies ranging from 4352 to 11 070 Hz (for the 1/2 inch depth tube chosen). Amplitudes were as high as 156 dB in the vicinity of actuation. [Work supported by Boeing/DARPA.]

11:35

3aNSb5. On silencers, asymmetry, and group theory. L. J. Eriksson (Eriksson Research, LLC, 6105 Fairfax Ln., Madison, WI 53718, erikssonresearch@execpc.com)

One of the most important factors in the design of effect noise control structures is the use of asymmetry. The design of silencers and mufflers provides an excellent illustration of the use of asymmetry to improve acoustical performance. Symmetric silencer designs often concentrate performance on a narrow range of frequencies at the expense of broadband performance. Asymmetric designs enable more effect performance over a broad range of frequencies. This is particularly important for applications with wide variations in operating speed and high-harmonic content. This work examines the role of asymmetry in the design of passive and active silencers including both plane-wave and higher-order mode effects. Certain unifying principles are identified and their role in system design is discussed. It concludes with suggestions for future research on asymmetry in acoustics related to the theory of iterative structures, the Noether conservation theorem, and the mathematical theory of groups.

WEDNESDAY MORNING, 6 JUNE 2001 SALONS III/IV, 8:00 TO 10:30 A.M.

Session 3aPAa

Physical Acoustics: Thermoacoustics

Ralph T. Muehleisen, Chair

Civil, Environmental and Architectural Engineering, University of Colorado, Campus Box 428, Boulder, Colorado 80309-0428

Contributed Papers

8:00

3aPAa1. Acoustic waveform dependence on resonator shape in thermoacoustic engines. Mark F. Hamilton, Yurii A. Ilinskii, and Evgenia A. Zabolotskaya (Dept. of Mech. Eng., Univ. of Texas, Austin, TX 78712-1063)

At high amplitudes, sound fields in thermoacousic refrigerators and prime movers are rich in harmonics and possess nonsinusoidal waveforms. In empty cylindrical resonators the natural frequencies are equidistant, the pressure waveforms are sawtooth in appearance, and they dissipate considerable energy at the shocks. Noncylindrical resonators possess nonequidistant natural frequency spectra that impede harmonic generation and reduce losses. The relative phases of the harmonics also depend on the natural frequency spectrum, mainly the relation of the second harmonic to the closest natural frequency, which usually corresponds to the second mode. Pressure waveforms are U shaped when the second harmonic is greater than this natural frequency, sawtooth when the second harmonic coincides with the natural frequency, and inverted-U shaped when the second harmonic is less than the natural frequency. These phenomena are illustrated by results of numerical calculations performed with a 2-D nonlinear model that approximates a thermoacoustic engine consisting of a stack in a resonator with varying cross section. Both pressure and particle velocity waveforms are presented at different locations in resonators of different shapes, possessing correspondingly different natural frequency spectra. Attention is devoted especially to waveforms near and inside the stack. [Work supported by ONR.]

8:15

3aPAa2. A miniaturized micro-machined thermoacoustic cooler. Reh-Lin Chen, Ya-Chi Chen, Chung-Lung Chen (Appl. Computational Phys., Rockwell Sci. Ctr., 1049 Camino Dos Rios, Thousand Oaks, CA 91360), Chialun Tsai, Jeff DeNatale, and Jeff Nelson (Rockwell Sci. Ctr., Thousand Oaks, CA)

A thermoacoustic cooler of miniature scale with a micro-machined stack has been numerically modeled and experimentally realized. DELTAE and Fluent were used to perform the numerical modeling and simulation. The acoustic pressure, velocity, and temperature profiles in the stack and exchangers were simulated. The stack is manufactured using MEMS technology for precise dimension control and flexible geometry design. The MEMS stack as thick as 4.2 mm with high aspect ratio (1/50) thin rib structures, 10 to 15 μm, have been realized. Commercially available PZT drivers were initially used to deliver acoustic power into the resonator at around 4 kHz. A more powerful acoustic driver design using PLZT is under development. The simulation result, device performance, and system integration will be reported. [Work supported by DARPA.]

8:30

3aPAa3. Performance of a high-frequency thermoacoustic refrigerator. Ehab Abdel-Rahman, Orest G. Symko, and Angela M. Frates (Dept. of Phys., Univ. of Utah, Salt Lake City, UT 84112-0830)

A thermoacoustic refrigerator pumped at an acoustic frequency of 4.4 kHz has been investigated for optimal geometric configuration for heat pumping. The device consists of a piezoelectric driver with a random stack in a wave cylindrical resonator with air at 1 atm as the working fluid. The temperature difference between the heat exchangers, at each end of the stack, has been studied as a function of acoustic intensity, stack position, stack length, and stack surface area. Results show that with sound levels of 160 dB the C.O.P. can be optimized to be around 3. Such performance strongly depends on the stack length with a compromise between effective stack area and quality factor of the resonator. Results will also be presented on the efficiency of the driver as a function of load.

8:45

3aPAa4. Admixture of traveling wave and standing wave components in a miniature thermoacoustic heat engine. Young Kwon, Orest G. Symko, and Matt Emmi (Dept. of Phys., Univ. of Utah, Salt Lake City, UT 84112)

Because of the importance of the thermoacoustic prime mover for all sorts of applications, we have investigated its performance at different operating conditions. Of special interest is the onset of oscillation temperature difference and its dependence on the geometry of the device. This was investigated in miniature heat engines, with resonant frequencies of 2 − 5 kHz, with temperature differences where one end of the device was below room temperature and also when the other end was above room temperature. All the devices were operated with air at 1 atm in the wave resonator configuration. The onset temperature ratios depended strongly on geometry, stack, heat exchangers, and on the degree of traveling wave admixture to the standing wave produced by the resonator.

9:00

3aPAa5. Thermoacoustic–Stirling model refrigerator. Matthew E. Poese and Steven L. Garrett (Grad. Prog. in Acoust., Penn State Univ., P.O. Box 30, State College, PA 16804, poese@psu.edu)

The design and performance of a model refrigerator that uses a stainless-steel woven regenerator (4.1 cm diameter, 5.0 mm thick) and an unconventional Helmholtz resonator to provide proper phasing for the regenerator are described. Cooling is produced in air at atmospheric pressure with sound pressures in excess of 6 kPa in a transparent 60 cm long standing wave tube. The sound power is generated using an inexpensive 5 in loudspeaker enclosed in standard PVC plumbing hardware. The resonance frequency of the Helmholtz resonator, and thereby the refrigerator's cooling capacity, can be tuned by adjusting the volume between the hot end of the regenerator and the end of the tube. As described in Swift *et al.* [J. Acoust. Soc. Am. **105**, 711 (1999)], performance is improved significantly by the addition of a latex diaphragm to suppress streaming. Due to its low cost and flexibility, this device is useful as both a lecture demonstration and for the measurements that will be presented and compared with theoretical performance. [Work supported by ONR and the PSU Applied Research Laboratory. Donation of the regenerator material by N. V. Bekaert S.A. of Belgium and discussions with R. M. Keolian, R. W. Smith, and R. S. Wakeland are gratefully acknowledged.]

9:15

3aPAa6. Numerical simulation of minor losses in thermoacoustic devices. Chingwei M. Shieh, Philip J. Morris (Dept. of Aerosp. Eng., The Pennsylvania State Univ., 233P Hammond Bldg., University Park, PA 16802), and Victor W. Sparrow (Grad. Prog. in Acoust., The Pennsylvania State Univ., 157 Hammond Bldg., University Park, PA 16802)

As thermoacoustic devices become more efficient, minor losses through sudden expansions or contractions in such acoustic resonators become more important. To further improve the efficiency of thermoacoustic devices it is necessary to understand in detail the fluid dynamic mechanisms involved in these minor losses. In the present study, a parallel numerical simulation of a three-dimensional acoustic resonator, typical of a thermoacoustic device, is presented. A computational aeroacoustic (CAA) approach is used. The Navier–Stokes equations are discretized in space with the fourth-order dispersion-relation-preserving (DRP) scheme of Tam and Webb and are integrated in time with a fourth-order Runge–Kutta scheme. In the attached regions of the flow the Spalart–Allmaras one-equation turbulence model is used. For separated flow regions this transitions automatically to a detached eddy simulation (DES). In the present study a high amplitude standing wave is generated in a resonator with a sudden change in cross-sectional area. Numerical results for the flow field and minor losses are presented and initial results are compared with experimental data. [Work supported by ONR.]

9:30

3aPAa7. Nonlinear properties of jet pumps. Andi Petculescu and Larry Wilen (Dept. of Phys. and Astron., Ohio Univ., Athens, OH 45701, wilen@helios.phy.ohiou.edu)

Traveling-wave thermoacoustic devices employ jet pumps to cancel mass streaming which develops in toroidal geometries [G. W. Swift, D. L. Gardner, and S. Backhaus, J. Acoust. Soc. Am. **105**, 711–724 (1999)]. We are using a simple nonresonant technique to measure nonlinear effects in jet pumps. The method involves a lumped-element analysis of a compliant region in parallel with the sample. The acoustic impedance of the sample is determined from the measured overall impedance and the impedance of the compliant region by a simple subtraction. At high amplitudes, (1) the resistance of the sample is amplitude dependent, and (2) a DC pressure develops across the sample. From these quantities, we determine the minor loss parameters K_{in} and K_{out} for a variety of jet pumps using an analysis similar to that described in Thurston [J. Acoust. Soc. Am. **30**, 452–455 (1958)]. A physical picture will be presented which explains the dependence of these parameters on geometry. Time permitting, results for some exotic geometries will also be discussed. [Work supported by ONR.]

9:45

3aPAa8. Thermoviscous functions of wire mesh and RVC stacks. Ralph T. Muehleisen and C. Walter Beamer (Dept. of Civil, Environ., and Architectural Eng., Univ. of Colorado, Boulder, CO 80309)

Thermoacoustic stacks made of stacked wire mesh elements or reticulated vitreous carbon (RVC) are becoming popular because of their performance, cost, and ease of construction. The thermoviscous Rott functions f_ν and f_κ [related to the porous media function $F(\lambda)$] have been measured for wire mesh and RVC stacks of 20, 30, 45, 60, 80, and 100 pores-per-inch. The thermoviscous functions were determined from measurements of the characteristic impedance and propagating wave number

of the stacks using the 4 microphone method of Song and Bolton. The measurements were then used to develop empirical models for the stacks. The measurement technique, measurement results, and emprical models are presented. [Work supported by ONR.]

10:00

3aPAa9. Numerical model of heat exchangers in oscillating flow with oscillating pressure. Ray Scott Wakeland and Robert M. Keolian (Grad. Prog. in Acoust., Penn State, P.O. Box 30, State College, PA 16804, wakeland@psu.edu)

Results are presented from a numerical model of heat transfer between two heat exchangers in oscillating flow and with pressure oscillations of the type expected in a standing-wave-based thermoacoustic device. The model assumes that each parcel of gas undergoes adiabatic compression/expansion when outside the exchangers; when inside an exchanger, the parcel temperature change ΔT in each small time step Δt is taken to be a superposition of the change that would have occurred for adiabatic pressure change Δp plus a change due to thermal relaxation toward the exchanger temperature T_{hx}. That is, $\Delta T=[(\gamma-1)/\gamma](T/p)\Delta p-\alpha(T-T_{hx})\Delta t$, where γ is the polytropic coefficient and α is an arbitrary relaxation time constant characterizing the exchanger. The pressure p fluctuates as $p/p_0=(x/x_0)^{-\gamma}$, where x is the distance from the parcel to the pressure antinode (hard end) and subscripts zero indicate equilibrium values. This relation is expected to hold close to pressure antinodes, where heat exchangers operate in thermoacoustic devices. After iteration, the model results in a closed cycle of parcel temperature as a function of time, allowing calculation of cycle work and cycle thermal efficiency. [Work supported by ONR and the Pennsylvania Space Grant Consortium.]

10:15

3aPAa10. Thermodynamics of open cycle thermoacoustic engines. Nathan T. Weiland and Ben T. Zinn (Schools of Aerosp. and Mech. Eng., Georgia Inst. of Technol., Atlanta, GA 30332, gte852f@prism.gatech.edu)

Recent advances in thermoacoustics have included the design, fabrication, and testing of an open cycle thermoacoustic refrigerator where the cold heat exchanger can be replaced by a slowly flowing gas which is cooled as it passes through the stack [R. S. Reid and G. W. Swift, J. Acoust. Soc. Am. **108**, 2835–2842 (2000)]. In addition to removing the irreversibilities associated with the cold heat exchanger, an important thermodynamic advantage is provided by cooling the working fluid as it flows through the stack, because the removal of heat from the fluid at higher temperatures increases the efficiency of the refrigeration process. This work investigates the possibility of similar efficiency improvements in open cycle thermoacoustic engines and heat pumps, where a steady flow is superimposed on the working fluid. The benefits of adding mean flow to a thermoacoustic engine rely heavily on the chosen method of heat transfer to the hot side of the stack or regenerator. Ideal and nonideal heat transfer from flow-based and electric-based heat exchangers will be analyzed and compared to an open cycle thermoacoustic engine driven by a steady flow of hot gas through its stack or regenerator.

Session 3aPAb

Physical Acoustics: Propagation in Inhomogeneous Media

D. Keith Wilson, Chair
U.S. Army Research Laboratory, 2800 Powder Mill Road, Adelphi, Maryland 20783-1197

Contributed Papers

10:40

3aPAb1. Sound propagation in the nighttime atmospheric boundary layer: The CASES-99 experiment. John M. Noble, D. Keith Wilson, and Mark A. Coleman (U.S. Army Res. Lab., Attn: AMSRL-CI-EP, 2800 Powder Mill Rd., Adelphi, MD 20783-1197)

A unique series of acoustical measurements was performed in calm, nighttime conditions in conjunction with the Cooperative Atmospheric-Surface Exchange Study (CASES-99). Conducted near Wichita, KS in October 1999, CASES-99 was one of the most ambitious experiments yet undertaken in boundary-layer meteorology. The purpose was to study processes in the nighttime boundary layer and the morning/evening transitions. Included among the instrumentation were a 60-m tower, free-flying and tethered balloons, and an FM-cw radar to image turbulence and wave motions. The concurrent sound-propagation experiment involved a series of five 6-m towers placed at distances between 300 and 1200 m from a loudspeaker. Each tower had microphones at heights of 0.5, 1, 2, and 3 m. Three transverse linear arrays were also placed on the ground at distances out to 1300 m from the speaker. Each data run consisted of broadcasting a 50-Hz square wave for a 2-h block of time. The acoustic and meteorological data are currently being analyzed. Preliminary results suggest that received sound levels are determined by both the gradual evolution of the wind and temperature gradients (due to cooling of the ground) and by discrete episodes associated with downward mixing of warm air.

10:55

3aPAb2. Sound refraction characteristics near the ground based on turbulence similarity theory. D. Keith Wilson (U.S. Army Res. Lab., ATTN: AMSRL-CI-EP, 2800 Powder Mill Rd., Adelphi, MD 20783-1197, dkwilson@arl.army.mil)

Sound refraction in the near-ground atmosphere over uniform terrain is systematically addressed with the Monin–Obukhov turbulence similarity theory. The gradient of the effective sound speed (actual sound speed plus the component of the wind vector in the propagation direction) is parametrized with two nondimensional ratios, one proportional to the height and the other representing the magnitude of sound-speed fluctuations relative to wind-speed fluctuations. Based on these parameters, three distinct propagation regimes (with transitions) can be identified: (1) strongly unstable atmospheric stratification where refraction is upward at all propagation angles relative to the mean wind, (2) strongly stable atmospheric stratification where refraction is downward for all angles, and (3) neutral atmospheric stratification where refraction is determined by the wind direction for small values of the nondimensional height ratio and is upward for greater heights. The relative contributions of temperature and humidity to the sound-speed gradient are found to depend on the Bowen ratio, defined as the ratio of the sensible and latent heat fluxes at the surface. Weather conditions appropriate for measuring sound-pressure levels are discussed in connection with the specifications in ANSI S12.18-1994.

11:10

3aPAb3. The effects of turbulence anisotropy and inhomogeneity on line-of-sight sound propagation. Vladimir E. Ostashev (NOAA/ETL, 325 Broadway, Boulder, CO 80305 and Phys. Dept., NMSU, Las Cruces, NM 88003, vostashe@nmsu.edu) and D. Keith Wilson (U.S. Army Res. Lab., Adelphi, MD 20783)

Line-of-sight wave propagation through isotropic, homogeneous turbulence has been studied by many previous researchers. A treatment for sound propagation, valid in the presence of both temperature and velocity fluctuations, was presented by Ostashev [*Acoustics in Moving Inhomogeneous Media* (E & FN SPON, London, 1997)]. Results for the statistical moments most frequently of interest were derived, in particular the correlation functions and variances of the log-amplitude and phase fluctuations, the mean sound field, and the coherence function of the sound field. In the present paper, these equations are generalized for the case of line-of-sight sound propagation through anisotropic, inhomogeneous turbulence. The obtained equations do coincide with those known in the literature in the limit when only temperature fluctuations are present or turbulence is isotropic and homogeneous. Using Mann's spectrum of shear-produced atmospheric turbulence, the effects of turbulence anisotropy on the statistical moments of a sound field are studied. [This material is based upon work supported in part by the U.S. Army Research Office under Contract No. DAAG55-98-1-0463.]

11:25

3aPAb4. Optimization of time-reversal focusing in a multiple scattering environment. Arnaud Derode, Arnaud Tourin, and Mathias Fink (Universite Paris 7, LOA, ESPCI, 10 rue Vauquelin, 75005 Paris, France)

We present new experimental results about the limits of time-reversal focusing through a two-dimensional multiple scattering medium. In a weakly heterogeneous medium, the quality of focusing is limited by the array aperture and bandwidth. It gets better in the presence of high-order multiple scattering: the long-lasting scattered signals are recompressed to form a high-power pulse, and the spatial resolution becomes practically independent from the array's aperture. With a 16-element 3-MHz array, the resolution was found to be 30 times thinner than in a homogeneous medium and resolutions of the order of the wavelength were attained. Therefore one could think that the more multiple scattering, the better. Experimental results show that this is not true. As the order of scattering is progressively increased, the system first takes advantage of multiple scattering, then reaches a limit; for higher orders of scattering, the focusing is degraded both spatially and temporally. The origin and consequences of this phenomenon are shown and discussed in relation with the transport parameters of the medium. There is therefore an optimal amount of multiple scattering. This is of high importance for applications of time-reversal focusing in a scattering or reverberating environment, whether for imaging or communication purposes.

11:40

3aPAb5. Scattering of sound by a core vortex: Numerical simulations using a wide angle parabolic equation. Ph. Blanc-Benon and L. Dallois (Ecole Centrale de Lyon, Ctr. Acoustique, LMFA UMR 5509, BP 163 69131 Ecully Cedex, France, philippe.blanc-benon@ec-lyon.fr)

The generation and scattering of sound by flow inhomogeneities such as vortices are basic problems which have received much attention in the efforts to develop methods to detect the strength and location of noise sources. Recent experimental studies have demonstrated that acoustic scattering can serve as an efficient (direct and nonintrusive) probe of the vorticity field for the characterization of turbulent flows. In 1978 Candel presented the results of a numerical study based on a standard parabolic equation in which the sound speed variations are included through an effective celerity $c_{\text{eff}}=c_0+v_x$ where the x axis give the direction of propagation of the acoustic wave. However, the effective celerity model does not include the effects of the perpendiculary components of the velocity field and the importance of this components increases with the angle of propagation. A wide angle parabolic equation has been recently proposed [Dallois *et al.*, 7thAIAA/CEAS paper N 2001-2256] to take into account the mean velocity effects up to second order in Mach number. To estimate the validity of this new parabolic equation (MW–WAPE), we considered a configuration with both strong gradient and high Mach number. Our results are compared to a reference solution obtained by solving the linearized Euler equations. The MW–WAPE solutions are in good agreement with the reference solutions. This parabolic equation gives accurate results up to $M\approx 0.5$ whereas the standard equation using an effective celerity failed to give good results above $M\approx 0.2$.

11:55

3aPAb6. Backscattering of acoustic waves in a cavity. Mark J. Beran and Shimshon Frankenthal (Faculty of Eng., Tel Aviv Univ., Ramat Aviv, Israel)

We discuss here the transmission of an acoustic signal in a cavity with random fluctuations in the sound speed. We consider pulse transmission in the single-mode case which consists of one forward and one backward propagating wave and cw signals in cavities with two propagating forward and backward waves. We begin by reviewing the cw single-mode case which has been treated in depth by a number of authors. For the pulse propagating in a single-mode cavity we then show how the cw results may be used to obtain useful information. Finally, we extend the governing equations to the two-mode case and consider possible methods of solution.

Session 3aPP

Psychological and Physiological Acoustics: Behavioral Studies and Physiological Correlates

Ruth Y. Litovsky, Chair
Department of Biomedical Engineering, Boston University, 44 Cummington Street, Boston, Massachusetts 02215

Chair's Introduction—9:00

Invited Papers

9:05

3aPP1. Physiological and behavioral assessments of an auditory pathway for processing spectral cues for sound localization. Brad May (Dept. of Otolaryngol.–HNS, Johns Hopkins Univ., 720 Rutland Ave., Baltimore, MD 21205)

Many species, including humans and cats, have head-related transfer functions (HRTFs) that introduce sharp features to the amplitude spectrum of broadband sounds as auditory stimuli propagate to the eardrum. Our behavioral results indicate that HRTF-based spectral cues, in particular mid-frequency spectral notches, are crucial for accurate spatial hearing in cats. This talk will summarize a series of physiological and behavioral assessments that follow the neural encoding of biologically relevant spectral information from a distributed representation in the auditory nerve to notch-selective neurons in the inferior colliculus. Just as anatomically defined pathways have been described for binaural directional cues, our results support the existence of a functionally segregated auditory pathway that is specialized for the processing of spectral cues for sound localization. [Work supported by NIDCD Grant No. DC00954.]

9:35

3aPP2. Psychophysical and physiological studies of the precedence effect and echo threshold in the inferior colliculus of the behaving cat. Daniel J. Tollin, Luis C. Populin, and Tom C. T. Yin (Dept. of Physiol., Univ. of Wisconsin, 1300 University Ave., Madison, WI 53706)

We are studying the relationship between single-unit responses in the inferior colliculus (IC) of behaving cats and their perception of spatial location using stimuli evoking the precedence effect (PE), an illusion where the perceived location of two similar sounds delivered from different locations but separated by a time delay approximates the perceived location of only the leading sound. Psychophysically, our cats experienced the PE: for 1–15-ms delays, they localized paired transients to nearly the same location as the leading sound alone. For delays greater than 15–20 ms, they often localized the leading and/or lagging location, an indication of echo threshold. Physiologically, for the delays that the cats experienced the PE the IC responses to the lag were reduced relative to the responses to the lag presented alone but with little effect on the response to the lead. The responses of these cells correlate with the behavioral perception of location: for delays encompassing the PE the cats' spatial perception of the sound is similar in the presence or absence of the lag, while at echo threshold the IC responses to the lag are nearly fully recovered to the response obtained in the lag presented in isolation. [Work supported by DC00116, DC02840, and DC00376.]

10:05

3aPP3. Behavioral and physiological studies of binaural detection with reproducible noise in rabbit. Laurel H. Carney (Dept. of Biomed. Eng., Boston Univ., 44 Cummington St., Boston, MA 02215, carney@bu.edu)

This presentation will review studies of binaural detection in rabbit. Behavioral results include detection experiments with narrow band and wide band reproducible noise maskers using the Pavlovian conditioning technique. These results showed similar trends across bandwidths and binaural conditions as in a parallel study in human listeners. Comparison of the responses to the ensemble of reproducible noises across bandwidths demonstrates the influence on performance of spectral components that are outside a narrow frequency band (approximately a critical bandwidth) surrounding the tone. Physiological responses of neurons in the inferior colliculus to the ensemble of reproducible noises will be presented. These responses can be compared to decision variables provided by computational models that are being developed to predict the behavioral results. Strategies for ongoing combined studies of behavior and physiology related to detection will be presented. [Work supported by NIH-NIDCD.]

10:35–10:50 Break

10:50

3aPP4. Spectral and temporal integration in the auditory cortex of alert primates. Xiaoqin Wang (Dept. of Biomed. Eng., Johns Hopkins Univ., 720 Rutland Ave., Ross 424, Baltimore, MD 21205, xwang@bme.jhu.edu)

Lesion studies in humans and primates have shown that the auditory cortex is necessary for the perception of spectrally and temporally complex sounds. Our knowledge on signal processing mechanisms in this cortical area, however, has been largely based on experimental observations obtained under anesthetized conditions. Anesthetics have long been known to have severe side effects on auditory cortical responses. We have systematically studied single-unit activities in the auditory cortex of alert primates using a wide range of spectrally and temporally complex sounds. In the spectral domain, our results showed that neurons in the auditory cortex integrate inputs as far away as several octaves from their central excitatory receptive fields. Such spectral integration appears to be supported by an extensive network of long-range horizontal connections. In the temporal domain, we showed that stimulus-synchronized discharges are on average limited to 20–30 ms inter-stimulus intervals in the unanesthetized cortex. Acoustic transients occurring at shorter time scales are integrated by cortical neurons and transformed into discharge-rate based representations. These findings suggest that (1) processing of acoustic signals by individual cortical neurons is modulated by spectral context and (2) time-varying signals are represented in nonisomorphic forms in the auditory cortex.

11:20

3aPP5. Auditory spatial resolution in the barn owl under echoic and anechoic conditions. Terry T. Takahashi, Matthew W. Spitzer, and Avinash Bala (Inst. of Neurosci., Univ. of Oregon, Eugene, OR 97403)

Experiments were designed to allow direct comparison of auditory spatial resolution, measured behaviorally, with that of single space-specific neurons in the barn owl's midbrain. Behavioral measurements of spatial discrimination were obtained using habituation and recovery of the pupillary dilation response (PDR). The acoustically evoked PDR habituates to repeated presentation of a sound, and recovers if the location changes. Thus, the difference in magnitudes of PDRs evoked by a sequence of test and habituating stimuli can be quantified using signal detection theory to provide a measure of discrimination [''$p(c)$,'' computed from empirical ROC curves]. The minimum audible angles (MAAs) for single sound sources separated in azimuth and elevation were 3° and 9°, respectively. Under simulated echoic conditions, MAAs increased by a factor of 2 for direct sources, and 4 for simulated reflections. Neuronal discrimination was similarly quantified using ROC curves to calculate $p(c)$ for spike discharges evoked by sources at different virtual locations. Preliminary analysis of neuronal data from anesthetized owls suggests that MAAs of the most sensitive neurons are similar to behavioral MAAs. Simultaneous measurement of neuronal and behavioral performance in awake owls is currently underway. [Work supported by the McKnight Foundation, NIH DC03925 and DC00448.]

Contributed Paper

11:50

3aPP6. Frequency selectivity estimated using stimulus-frequency otoacoustic emissions and psychophysical masking. Andrew J. Oxenham (Res. Lab. of Electron., MIT, Cambridge, MA 02139) and Christopher A. Shera (Massachusetts Eye and Ear Infirmary, Boston, MA 02114)

Psychophysical studies of frequency selectivity have concluded that the relative bandwidths of auditory filters decrease only slightly with increasing center frequency above 1 kHz. In contrast, estimates based on measurements of stimulus-frequency otoacoustic emissions (SFOAEs) in humans and other mammals suggest that frequency selectivity improves significantly with increasing center frequency. Psychophysical measures have mostly used simultaneous masking, which can be influenced by suppression. Since suppressive effects are stronger at high frequencies, these techniques may systematically overestimate filter bandwidths at high center frequencies. This study investigates whether psychophysical and physiological estimates of filter bandwidth can be reconciled if suppressive effects are eliminated by using forward masking. Filter bandwidths at center frequencies of 1, 2, 4, 6, and 8 kHz were measured using the notch-noise technique in simultaneous and forward masking with the signal level fixed at 10 and 30 dB above threshold in quiet. Relative bandwidths estimated in the same subjects using measurements of SFOAE group delay at 40-dB SPL were consistent with earlier results. The psychophysical and physiological estimates of relative bandwidth are generally consistent, suggesting that the relationship between center frequency and filter bandwidth found in previous psychophysical studies may have been influenced by suppression. [Work supported by NIDCD.]

Session 3aSAa

Structural Acoustics and Vibration: Complexity and Energetics

J. Gregory McDaniel, Chair

Department of Aerospace and Mechanical Engineering, Boston University, 110 Cummington Street, Boston, Massachusetts 02215

Contributed Papers

8:00

3aSAa1. An energy formulation for high-frequency radiation into an unbounded acoustic medium. Aimin Wang, Nickolas Vlahopoulos (Dept. of Naval Architecture and Marine Eng., Univ. of Michigan, 2600 Draper Rd., Ann Arbor, MI 48109-2145), and Kuangcheng Wu (Newport News Shipbuilding, Newport News, VA 23607-2770)

A new energy formulation for computing high-frequency radiation from structures into an unbounded acoustic medium is presented. The outer surface of the radiating structure is divided into elements. Surface intensity values on each element comprise the boundary conditions. Similar to the panel method [J. L. Hess and A. M. O. Smith, *Calculation of Potential Flow About Arbitrary Bodies*, Progress in Aeronautical Sciences Vol. 8, edited by D. Kuchemann (Pergamon, New York, 1967)] a distribution of fictitious energy sources and sinks is considered over the surface of the model. Appropriate strengths are computed for the sources and the sinks in order to satisfy the intensity boundary conditions. Once the distribution of the surface sources/sinks has been completed, the acoustic energy density and the acoustic intensity can be evaluated at any point on the surface of the radiating object, or at any field point within the acoustic medium. The new development does not require one to define damping properties for the infinite fluid and no assumptions are made about representing the radiated acoustic field as reverberant. The new developments are validated through comparisons of numerical results to analytical solutions. [Research supported by ONR, Contract No. N00014-00-I-0382.]

8:15

3aSAa2. Energy source identification using reactive structural intensity. Paulo Sergio Lima Alves and Jose Roberto Franca Arruda (State Univ. of Campinas, P.O. Box 6122, 13083-970 Campinas, SP, Brazil, psl@fem.unicamp.br)

The analysis of wave propagation phenomena in structures, especially energy flux, can be a powerful approach to noise and vibration control problems. The active mechanical intensity corresponds to the period-averaged value and is related to the propagative wave field. The reactive intensity is related to the reverberance of energy in the structure, which originates the standing waves or mode shapes. Properly separated into its wave components, reactive intensity maps agree well with operational modes, indicating the location of the nodal lines. On the other hand, active intensity is the quantity usually addressed when solving vibration and noise problems, since it provides information about the main energy flow paths, thus enabling identification of the energy sources and sinks. For highly reverberant structures, however, measuring active intensity becomes awkward, and intensity plots fail to indicate the region where the energy is injected or dissipated in the structure. A new technique to localize energy sources based on the divergence of the reactive intensity and the distribution of the potential and kinetic energy densities within the structure is presented. Numerical results are given for a beam with a point excitation using finite-element and spectral element models based on the Bernoulli–Euler beam theory.

8:30

3aSAa3. Determination of the frequency response functions of complex systems using spectral-based inverse substructuring approach. Teik C. Lim (Dept. of Mech. Eng., The Univ. of Alabama, 290 Hardaway Hall, P.O. Box 870276, Tuscaloosa, AL 35487, tlim@coe.eng.ua.edu)

The structural-acoustic response of discretely connected complex structures, such as a motor vehicle, are determined by inverting the spectral-based substructuring matrix equation in dynamic compliance form, which is classically used to compute coupled system response in terms of the dynamic characteristics of free substructures. The resulting formulation backs out the response spectra of the free substructures and the dynamic stiffness matrix representing the coupling elements. This provides a useful way to dissect a fairly complex system into two or more substructures. When analyzing the dynamic interaction between two adjacent substructures, the proposed theory can be conveniently used to express the transmissibility and path contribution functions directly as measurable coupled system response functions. Therefore, critical elements of a complex system can be identified easily without having to physically disconnect the substructures. The salient features of this approach are demonstrated using a motor vehicle example. The study specifically addresses the road noise problem by examining the sensitivity of the motor vehicle structure response to tire patch excitation. In the analysis, the body structural-acoustic functions, and excitation transfer functions between the spindle and chassis mounts are determined. [Work supported by CAVT and Ford.]

8:45

3aSAa4. Accounting for manufacturing variability in interior noise computations. Michael J. Allen and Nickolas Vlahopoulos (Dept. of Naval Architecture and Marine Eng., Univ. of Michigan, 2600 Draper Rd., Ann Arbor, MI 48109-2145)

A formulation that accounts for manufacturing variability in the analysis of structural/acoustic systems is presented. The methodology incorporates the concept of fast probability integration with finite element (FEA) and boundary element analysis (BEA) for producing the probabilistic acoustic response of a structural/acoustic system. The advanced mean value method is used for integrating the system probability density function. FEA and BEA are combined for producing the acoustic response that constitutes the performance function. The probabilistic acoustic response is calculated in terms of a cumulative distribution function. The new methodology is used to illustrate the difference between the results from a probabilistic analysis that accounts for manufacturing uncertainty, and an equivalent deterministic simulation through applications. The probabilistic computations are validated by comparison to Monte Carlo simulations. Significant computational time savings are achieved by the new probabilistic algorithm in comparison with the Monte Carlo simulations. Based on its computational efficiency and its accuracy, the new methodology is concluded to be a viable method of calculating numerically the probabilistic response of structural/acoustic systems due to manufacturing variability. [Research supported by the National Defense Science and Engineering Graduate Fellowship program.]

9:00

3aSAa5. On the emergence of the Green's function in the correlations of a diffuse field. Richard L. Weaver and Oleg I. Lobkis (Theoretical and Appl. Mech., Univ. of Illinois, Urbana, IL 61801)

It is shown that a diffuse field is not devoid of phase information, but has a correlation function equal to the Green's function. More specifically, the cross correlation between diffuse signals in two transducers is very nearly equal to the direct response of one transducer to an impulse applied to the other. Plausibility arguments for this assertion are followed by a detailed proof. Estimates are constructed for the amount of averaging necessary before clear convergence can be expected. The assertion is then confirmed in laboratory experiments using a single impulsive ultrasonic point source to generate the diffuse field. The correlation function is constructed by averaging over the long ring down time of the resulting diffuse field. Comparison with direct responses shows that the correlation function reveals the strongest features, but smaller features do not emerge above the grain noise. The convergence is found to further improve with additional averaging over source position. [Work supported by NSF.]

9:15

3aSAa6. Can the spatial distribution of damping be measured? Sondipon Adhikari, J. Woodhouse, and A. Srikanth Phani (Eng. Dept., Cambridge Univ., Trumpington St., Cambridge CB2 1PZ, UK)

Our knowledge regarding the detailed description of damping in a vibrating structure is very primitive. The most common approach involves assuming proportional viscous damping, then determining damping factors for each mode. The next step, followed by many authors, is to obtain a full nonproportional viscous damping matrix. Both methods may misrepresent the true damping behavior, as viscous damping is assumed from the outset. In this study a new method for identification of the spatial distribution of damping is proposed. The approach is based on general linear damping models and is not confined to the traditional viscous damping model. Algorithms have been developed for taking measured frequencies, damping factors and complex mode shapes, and fitting the parameters of a damping model with an exponential relaxation function. The viscous damping model is a special case of this model. Simulation studies show that the method seems to perform quite robustly. Preliminary experimental results on a beam with a localized patch of damping are very encouraging and show that the method can indeed indicate the spatial distribution of damping correctly. The approach also gives some information on the question of whether the chosen exponential damping mechanism is correct or not.

9:30

3aSAa7. Application of fuzzy structures analysis to aircraft panels. Victor W. Sparrow (Grad. Prog. in Acoust., Penn State, 157 Hammond Bldg., University Park, PA 16802, sparrow@helmholtz.acs.psu.edu) and Ralph D. Buehrle (Structural Acoust. Branch, NASA Langley Res. Ctr., Hampton, VA 23681)

In 2000 a research implementation of Christian Soize's fuzzy structures analysis procedures were implemented by Sparrow in the commercial finite element program MSC/NASTRAN [MSC Software Corp. (Los Angeles, CA)] using its direct matrix abstraction program (DMAP) capability coupled with MATLAB [Mathworks, Inc. (Natick, MA)]. The background and implementation details of the new DMAP code were recently given [AIAA paper 2001–1320, Non-Deterministic Approaches Forum, 42nd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conf., Seattle, WA, 2001]. The present talk will describe the application of the new code to two prototypical aircraft panels: a small 0.96 by 0.19 m panel with one longitudinal stringer and a larger 1.8 by 1.2 m panel including six longitudinal stringers and four equally spaced frame stiffeners. The smaller and larger panels were modeled in NASTRAN as having approximately 7900 and 60 000 degrees of freedom, respectively. In these studies the attachment points between the stringers and skin were assumed to be fuzzy. Results will be described comparing the fuzzy and nonfuzzy finite element predictions to experimental vibration results. [Work supported by NASA Research Cooperative Agreement NCC-1-382.]

WEDNESDAY MORNING, 6 JUNE 2001 SALON V, 10:00 A.M. TO 12:05 P.M.

Session 3aSAb

Structural Acoustics and Vibration, Engineering Acoustics, Noise and Signal Processing in Acoustics: Machinery Prognostics

David C. Swanson, Chair
520 East Hamilton Avenue, State College, Pennsylvania 16801

Chair's Introduction—10:00

Invited Papers

10:05

3aSAb1. Calculation of residual signal features for gear fault detection: Physics and signal processing. Karl M. Reichard (Appl. Res. Lab., The Pennsylvania State Univ., P.O. Box 30, State College, PA 16804-0030, kmr5@psu.edu)

The vibration signal produced by mating gears can be decomposed into two components—a ''mean'' component due to surface imperfections present when the gear is manufactured and elastic tooth deformations, and a ''dynamic'' component due to wear and gear tooth damage. The mean component is responsible for the gear mesh frequency and its harmonics and typically dominates the measured vibration signal; the dynamic component, or residual error signal, contains information about evolving damage. A number of techniques exist for extracting the residual error signal from measured vibration signals that require removal of the great mesh frequency and its harmonics from the measured vibration signal. Several techniques for computing gear health based on the residual error signal are compared and the effect of the technique used to remove the mean signal component is examined. [Work supported by ONR through the *Accelerated Capabilities Initiative* (Grant Number N00014-96-1147.]

10:35

3aSAb2. Gear fatigue crack size estimation using embedded modeling. C. James Li, Hyungdae Lee, and Suk-Hwan Choi (MEAEM, RPI, 110 8th St., Troy, NY 12180)

This paper describes an embedded modeling methodology for identifying gear meshing stiffness from measured gear angular displacement or transmission error. An embedded model integrating a physical based model of the gearbox and a parametric representation, in the form of truncated Fourier series, of meshing stiffness is established. A solution method is then used to find the meshing stiffness that minimizes the discrepancy between model output and the measured output. Furthermore, an algorithm is also developed to estimate the size of tooth crack from identified meshing stiffness. Both simulation and experimental studies were conducted to evaluate if identified tooth meshing stiffness can reveal a tooth crack more effectively, and if the crack size can be estimated with an adequate level of accuracy. [Work supported by ONR under the Multidisciplinary Research Program of the University Research Initiative.]

11:05

3aSAb3. Structural assessment aided by low level response measurements. Ziyad H. Duron (Dept. of Eng., Harvey Mudd College, 301 E. 12th St., Claremont, CA 91711)

The acquisition of low-level response measurements on large civil structures has been aided in recent years by significant advances in instrument and computer technologies. Measurement techniques applied to structures like dams and bridges have traditionally focused on known, measurable excitation forces (e.g., forced vibration) applied directly to the structure. An alternative way of thinking about the conduct of testing large structures is presented in which very low-level response measurements are used to evaluate dynamic characteristics, to evaluate system characteristics, and to evaluate changes in structural health or damage detection. Samples of actual field experiences are discussed and the ongoing development of a national database for large civil structural response characteristics is presented.

11:35

3aSAb4. Research and education program in machinery monitoring and prognosis at the University of Tennessee. Tom Byerley and Belle Upadhyaya (Maintenance and Reliability Ctr., The Univ. of Tennessee, 505 East Stadium Hall, Knoxville, TN 37996-0750)

An effective strategy for modern plant operation and maintenance requires the integration of plant monitoring, incipient detection and isolation of faults in critical equipment, and life prediction or prognosis to enable plant personnel in making decisions about maintenance scheduling and equipment replacement. This paper presents some of the recent techniques being developed for equipment prognosis and their implementation. Examples of current techniques include data-driven modeling, wavelet analysis, and generalized trending techniques. Applications to vibration signature trending, steam generator tubing degradation trending, accelerated aging of induction motors and their prognostics will be presented, demonstrating that the tools may be used effectively in minimizing unscheduled plant downtime. In addition, the paper will discuss how The University of Tennessee (UT), through its Maintenance and Reliability Center (MRC), is developing both the above-mentioned cutting edge tools and the personnel qualified to implement them in plant operations. UT and the MRC have created additional curriculum and a certification program at the BS, MS and PhD levels in order to prepare engineering graduates with knowledge and experience in the use of these emerging techniques and tools. The MRC also performs research to further develop these new health monitoring/life prediction systems.

Session 3aSC

Speech Communication: Glottal Models and Mechanisms, F0 and Prosody (Poster Session)

David A. Berry, Chair

Head and Neck Surgery, Unversity of California, Los Angeles, 31-24 Rehab Center, 1000 Veterans Avenue, Los Angeles, California 90095-1794

Contributed Papers

All posters will be on display from 8:30 a.m. to 12:00 noon. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 8:30 a.m. to 10:15 a.m. and contributors of even-numbered papers will be at their posters from 10:15 a.m. to 12:00 noon.

3aSC1. Computational aeroacoustics of phonation: Effect of subglottal pressure, glottal oscillation frequency, and ventricular folds. Cheng Zhang, Wei Zhao, Steven H. Frankel, and Luc Mongeau (School of Mech. Eng., Purdue Univ., West Lafayette, IN 47907, frankel@ecn.purdue.edu)

Numerical simulations of flow and acoustics in an idealized axisymmetric model of the human vocal tract have been conducted. The compressible Navier–Stokes equations were numerically integrated using highly accurate numerical treatments together with a forced glottal oscillation model and a moving grid. The effects of subglottal pressure and glottal oscillation frequency on velocity, vorticity, wall pressure, wall shear stress, and acoustic signal of the pulsating jet were investigated. An acoustic analogy was used to predict the far-field sound. Excellent agreement between the predicted and directly simulated far-field sound was obtained. The acoustic analogy was also used to decompose the acoustic source and identify monopole, dipole, and quadrupole contributions for analyses. The results show significant effects of subglottal pressure and glottal oscillation frequency on the jet vortical structure, wall forces, and acoustic radiation. The effect of including the ventricular (false) folds downstream of the oscillating glottal region was also investigated. Jet impingement on the ventricular folds introduces additional dipole sound sources. [Work supported by NIH DCO 3577-02, RO1 grant from NICDC.]

3aSC2. Acoustic correlates of breathy and clear phonation types. Ratree Wayland and Mohamed Al-Khairy (Univ. of Florida, P.O. Box 115454, Gainesville, FL 32611-5454, ratree@lin.ufl.edu)

Only a decade ago were types of phonation shown to have a phonological function in a language. Several languages spoken in Southeast Asia were claimed to have contrastive phonation type (i.e., clear, breathy and creaky) in their vowel systems. Several different techniques were employed to investigate this phenomenon. These included examining the physical appearance of the acoustic wave form, inverse-filtering, measuring the difference between amplitude of the first and the second harmonics (H1–H2), or between the first harmonic and the most prominent harmonic in the $F1$ and $F3$ regions (H1–A1, H1–A3), harmonic-to-noise ratio (HNR) as well as some aerodynamic techniques. This present study proposes to investigate the breathy versus clear phonation contrast phenomenon in Chong, a Mon-Khmer language spoken in Thailand using several of the above-mentioned techniques. Preliminary analysis suggested that only the H1–H2 values obtained for breathy vowels were greater than for clear vowels. The nonsignificant difference between breathy and clear vowels in Chong in other dimensions will be discussed in relation to the question of whether or not Chong should be classified as a phonation-type language.

3aSC3. Phonation type controls the source rate. Hansang Park (Dept. of Linguist., Univ. of Texas, Austin, TX 78712, phans@mail.utexas.edu)

The ratio of the source characteristic between the two harmonics in the relation of an octave was derived from the combined source and radiation characteristics, which shows that the source rate varies with frequency. A general term for harmonics was established from the ratio, where the higher harmonics are anchored to the fundamental component, such that the source and radiation characteristics of the higher harmonics can be represented in terms of those of the fundamental component. A new measure of phonation type was designed incorporating the varying source rate. Furthermore, it was claimed that phonation type controls the source rate. A phonation-type index k was proposed to account for variation in phonation type. The rationale for proposing the phonation-type index k is that phonation type is independent of the specific values of fundamental frequency but anchored to the fundamental component, and phonation type systematically contributes to the harmonics but not to the fundamental component. The phonation-type index k involves both open quotient and spectral tilt. The phonation-type index k may show unique distributions across speakers and contexts, and may account for modes of phonation type.

3aSC4. Methodology for straining engineered vocal fold tissues at audio frequencies. Sarah A. Klemuk, Ingo R. Titze (Dept. of Speech Pathol. & Audiol., The Univ. of Iowa, 330 Wendell Johnson Bldg., Iowa City, IA 52242), and Steven Gray (University of Utah, Salt Lake City, UT)

Engineered vocal fold tissues will be grown under conditions of shear strain, the magnitude of which is determined by known strains during human vocal fold oscillation. A cup and plate system is used with the Bohlin CVO 120 rheometer to apply a shear strain of approximately 0.4 radians at frequencies of 20–60 Hz. Baseline viscoelastic measurements are recorded for the synthetic scaffolding material into which different cell lines will be injected and grown to form their own matrix. The methodology allows duration and repetitions of strain exposure to seeded scaffolding to be varied and for subsequent viscoelastic measurements to be compared. [Work supported by NIDCD.]

3aSC5. Geometrical deformation of the canine vocal fold induced by formalin fixation. Miwako Kimura, Niro Tayama (Dept. of Otolaryngol., Univ. of Tokyo, Tokyo, Japan), Roger W. Chan (Purdue Univ., West Lafayette, IN 47907), and Ingo R. Titze (Univ. of Iowa, Iowa City, IA 52242)

Many existing data on vocal fold geometry are based on anatomical measurements made on histologically prepared vocal fold tissues, commonly by formalin fixation. However, the reliability and validity of these geometrical data are questionable because of the potentially significant tissue deformation associated with histological fixation. As we have shown that reliable geometrical data may be obtained on tissue samples quickly frozen with liquid nitrogen [Tayama *et al.*, Ann. Otol. Rhinol. Laryngol. (in press)], this study attempted to quantify the geometrical deformation of vocal fold tissues fixed with formalin with respect to quick-frozen tissues. Excised canine larynges were divided into two halves along the sagittal plane, one of which was quickly frozen and the other was fixed with formalin. Coronal sections of the frozen halves were thawed and histological slides were prepared from the fixed halves. Measurements of vocal fold geometry (depth and thickness) were made on images of midmembranous coronal sections of the samples. Results showed that significant shrinkage of the vocal fold body (thyroarytenoid muscle) and considerable distortion of the vocal fold cover (free edge) were associated with formalin fixation. These artifacts should be carefully considered when any geometrical data based on histologically fixed tissues are used for laryngeal biomechanical modeling.

3aSC6. Finite-element analysis of passive vocal fold posturing using nonlinear muscle responses. Eric J. Hunter, Ingo R. Titze, and Fariborz Alipour (Dept. of Speech Pathol. & Audiol., The Univ. of Iowa, 328G Wendell Johnson Bldg., Iowa City, IA 52242)

A finite-element model of the passive properties of laryngeal muscles was created, based on nonlinear stress–strain responses of laryngeal muscular tissue [E. J. Hunter and I. Titze, J. Acoust. Soc. Am. **108**, 2531 (2000)]. Using this finite-element muscle model as a building block, groups of muscles are used to construct a 3-D tissue model of vocal fold posturing. Vocal fold configurations were studied with assumed muscle stresses applied to the adductor–abductor muscles. Results comply, in the linear region, with earlier results on a 2-D linear model that had constant thickness [E. J. Hunter, J. Acoust. Soc. Am. **106**, 2540 (1999)], but now take into account more accurately the high strains. [Work supported by NIDCD.]

3aSC7. An *M*-mass vocal-fold excitation model and its relation to source–tract interaction. Jun Huang, Stephen Levinson (Univ. of Illinois, Urbana, IL 61801, jhuang@ifp.uiuc.edu), Don Davis, and Scott Slimon (Electric Boat Corp., Groton, CT 06340)

In this work, we propose an *M*-mass model of the vocal-fold excitation for articulatory speech synthesis. The vocal fold in this model is composed of *M* coupled oscillators. Each oscillator consists of a mass, a spring stiffness, and a dumper. The adjacent masses are coupled by a spring stiffness. The dynamic response of the *M*-mass model is described by a set of second-order ordinary differential equations called the equations of the motion. This *M*-mass model is combined with the Navier–Stokes equations to provide numerical solutions in an iterative way. First, the glottal volume velocity in the current iteration is used to determine the aerodynamic force exerted on each mass. Then the glottal area is calculated by solving the equations of motions given the aerodynamic force. This glottal area information is used to compute the glottal volume velocity in next iteration. We use this *M*-mass model to compute the glottal excitation signals including volume velocity and particle velocity. Experimental results show that we can observe ripples in the positive glottal opening interval of the particle velocity and volume velocity, which is strong evidence of source–tract interaction during speech production.

3aSC8. The effects of vowel quality and pitch on spectral and glottal flow measurements of the voice source. Melissa A. Epstein (Dept. of Linguist., UCLA, 3125 Campbell Hall, Los Angeles, CA 90095-1543) and Blas G. Payri (UCLA School of Medicine, Los Angeles, CA 90095-1794)

In efforts to evaluate voice quality, the voice source is frequently assessed using measurements of the voice spectrum and/or measurements of the glottal flow. In normal dynamic speech, though, speakers not only change their voice quality, but also their vowel qualities (i.e., they say different words) and their pitch (i.e., they use different intonation contours). The purpose of this study is to explore how vowel quality and pitch affect both spectral and glottal flow measurements of source characteristics. In the experiment, speakers are asked to say several words on three pitches each. Since these are real words, speakers should produce them in their normal, but undefined, voice quality. Samples are selected from the center of each vowel. Differentiated glottal flow is obtained by inverse filtering and is then matched with an LF model. The voice spectrum is evaluated using the measurements H1-H2, H1-A1 and H1-A3 (as proposed by Stevens and Hanson 1995). These measurements are then compared across vowels, pitches and speakers. [Work supported by the French Ministry of Foreign Affairs and NIH/NIDCD.]

3aSC9. Measuring voice quality in dynamic speech. Melissa A. Epstein (Dept. of Linguist., UCLA, 3125 Campbell Hall, Los Angeles, CA 90095-1543)

Different voice qualities are used in normal speech to convey phonological contrasts (e.g., Zapotec, Gujarati, etc.), prosodic information, emotions, etc. When measuring changes in voice quality in dynamic speech, it is necessary to use a technique that can capture small, relative differences under quickly changing conditions. Although a number of techniques are available for assessing voice quality (e.g., perceptual assessment, qualitative assessment of the time amplitude waveform, qualitative assessment of the spectrogram, quantitative measurement of the spectrum, and quantitative measurement of the voicing source), they have, for the most part, been designed for capturing voice quality differences in sustained vowels or carefully matched speech conditions. This study develops a technique for relative voice quality measurement which ascertains voice quality differences across diverse words, pitches, and speakers. Specifically, inverse filtering is used to obtain the differentiated glottal flow for each vowel in the dynamic speech, which is theoretically robust across different vowel qualities and pitches. Then, for comparison, a baseline voice quality measurement is obtained for each word from each speaker to control for segmental and personal voice quality differences. Thus, the relative voice quality measurement will be resistant to the variations in dynamic speech. [Work supported by NIH and NIH/NIDCD.]

3aSC10. Effect of hydration on the dynamic mechanical properties of vocal fold tissues. Roger W. Chan (Dept. of Audiol. and Speech Sci., Purdue Univ., West Lafayette, IN 47907, rchan@purdue.edu) and Niro Tayama (Univ. of Tokyo, Tokyo, Japan)

Previous research has shown that vocal fold hydration level is an important physiological variable affecting vocal fold vibration and vocal health. For instance, unbalanced laryngeal water transport has been associated with benign vocal fold lesions such as nodules and polyps, while hydration treatment is frequently used as a preventive or therapeutic strategy for the management of a variety of voice disorders. The theoretical basis for the role of hydration in voice production is that it may affect phonation threshold pressure or the ease of phonation, presumably by changing the viscous properties of vocal fold tissues. However, there is little empirical evidence showing how hydration might actually affect the biomechanical properties of vocal fold tissues. This study attempted to assess the effect of hydration on the viscoelastic shear properties of canine vocal fold mucosa. Osmotic changes in the hydration level of excised canine larynges were induced by incubation in hypertonic, isotonic, and hypotonic solutions. Viscoelastic shear properties of vocal fold mucosal tissues were measured with a rheometer. Significant changes in both issue

elasticity and viscosity were observed with the induced changes in hydration. These findings supported the hypothesis that hydration affects voice production by altering tissue viscoelastic properties.

3aSC11. Analysis of asymmetric vocal fold vibrations in a finite-element model. David A. Berry (UCLA School of Medicine, Div. of Head and Neck Surgery, 31-24 Rehab Ctr., Los Angeles, CA 90095-1794, daberry@ucla.edu) and Fari Alipour (Univ. of Iowa, Iowa City, IA 52242)

Asymmetric vibrations of a finite-element model of the vocal folds were analyzed using the method of empirical eigenfunctions. In a previous study of symmetric vibrations from the model, periodic vibrations yielded two dominant eigenfunctions, which were related to the theoretical normal modes of the model. In a more recent study, the method of empirical eigenfunctions was applied to high-speed endoscopic imaging of vocal fold vibration on human subjects, suggesting several mechanisms of irregular vibration. However, in the *in vivo* study, the investigation of empirical eigenfunctions was limited to the superior view provided by endoscopy. The results of the present computational study were compared with the *in vivo* study, while providing additional information regarding vocal fold vibration along the medial surface of the folds. Moreover, the computational study allowed the asymmetric vibrations to be studied systematically as a function of tissue stiffness. [Work supported by Grant No. R29 DC03072 from the National Institute on Deafness and Other Communication Disorders.]

3aSC12. Wavelet analysis of turbulence in the glottal jet flow. Fariborz Alipour (Dept. of Speech Pathol. and Audiol., Univ. of Iowa, Iowa City, IA 52242-1012)

Airflow from the lungs drives the vocal folds into oscillation and exits the glottis as a turbulent jet, even though laminar flow is documented entering the glottis. Incomplete glottal closure may result in air leakage that becomes turbulent. This turbulence is thought to be the main source of breathiness in the voice. Few studies have measured degree of turbulent noise in the glottis. This turbulent noise adds significant amounts of energy to the voice signal and cannot be extracted easily without changing the voice. In this study, the turbulent jet was investigated in an excised canine larynx model with simultaneous recordings of air particle velocity, subglottal pressure, airflow rate, and EGG signal for various conditions of phonation. The velocity was measured with a constant-temperature hot-wire anemometer system. To separate the turbulence from the periodic component of the velocity signal, the technique of wavelet denoising was employed. In this method, unlike the Fourier transform which decomposes the signal into periodic components, wavelet denoising uses template matching transform to approximate the signal with different resolution or scale. Results indicate that the high resolution part contains the turbulent noise. [Work supported by NIDCD Grant No. R01 DC03566.]

3aSC13. Effects of syllable position on *F*0 and duration in Mandarin disyllabic words. Ching X. Xu and Yi Xu (Dept. of Commun. Sci. and Disord., Northwestern Univ., Evanston, IL 60208, xxq@nwu.edu)

In a previous attempt to fit a pitch target approximation model for *F*0 contours in Mandarin disyllabic words [Xu *et al.* (1999)], it was found that the implementation of tonal targets required two sets of values for the model parameters, one for each of the two syllable positions. This suggests that syllable position may have an effect on the production of the syllable. The present study is designed to further examine this position effect. One-hundred-eleven disyllabic words with various tonal combinations were read by native Mandarin speakers with two carrier sentences. Duration and *F*0 analyses were performed. Preliminary results indicate that a syllable has longer duration, higher *F*0, and closer-to-ideal *F*0 shape when it is in the first position than in the second position. These results agree with previous acoustic-phonetics findings about Mandarin disyllabic words, and together they seem to suggest that the first syllable in a disyllabic word is more stressed than the second one. Interestingly, this appears to be at odds with current phonological theories about Mandarin prosody, which maintain that the second syllable carries greater weight than the first in a disyllabic foot. Possible causes for this discrepancy will be discussed.

3aSC14. Experiment on pitch target approximation model for generating Mandarin *F*0 contour. Xuejing Sun, Ching X. Xu, and Yi Xu (Dept. of Commun. Sci. and Disord., Northwestern Univ., 2299 N. Campus Dr., Evanston, IL 60208, sunxj@northwestern.edu)

In the present study, an experiment was conducted to generate *F*0 contours for Mandarin with a pitch target approximation model proposed in Xu, Xu and Luo (1999). In this model, *F*0 contours in speech are assumed to be resulting from asymptotic approximation to underlying pitch targets that are either static or dynamic. The model parameters were estimated through nonlinear regression using the Levenberg–Marquardt algorithm. The speech corpus consisted of sentences from Voice of America broadcasting news. After the regression analysis, sentences were re-synthesized with the generated *F*0 using the TD-PSOLA technique. Preliminary results indicate that *F*0 contours generated by the model are close to the original both numerically and perceptually. Furthermore, most underlying pitch targets obtained through the regression analysis seem to match the models basic assumptions. However, it is also apparent that information about both higher-level linguistic functionality and additional low-level articulatory constraints is needed to account for the numerical variations in the estimated parameters. In general, the results are encouraging as they show that the model can generate close-fitting *F*0 contours even with strong linguistic assumptions. This suggests that it has the potential to evolve into a system with the predictive power desirable for intonation modeling.

3aSC15. Pitch peak alignment as a function of lexical and pragmatic factors in two dialects of Serbian/Croatian. Rajka Smiljanic (Univ. of Illinois at Urbana–Champaign, 707 S. Mathews, Urbana, IL 61801)

This study examines the patterns of pitch peak retraction under focus in Serbian/Croatian (S/C). Acoustic measures of pitch peak alignment were used to identify the effect of pragmatic focus on pitch realization in two dialects of S/C with typologically distinct prosodic systems. In the Belgrade dialect there is a lexically determined contrast between two tonal alignments: early peak (''falling accent'') and late peak (''rising accent''). In the Zagreb dialect there is no such lexical distinction; instead, each word typically receives a pitch prominence on the stressed syllable. Subjects produced target words with neutral intonation and with narrow/contrastive focus. The peak alignment under these two pragmatic conditions was shown to vary in the two dialects. Zagreb shows peak retraction from late (neutral intonation) to early peak (narrow focus). In this way it is similar to other stress languages (e.g., Spanish). Belgrade, with lexically contrastive peak alignment, shows much less pragmatically conditioned variation in the position of accentual peaks. The two types of pragmatically conditioned alignment in Zagreb correspond closely to the two lexically determined alignments of Belgrade. This suggests that the existence of the lexical contrast limits the pitch variation due to the pragmatic influence, while in the absence of such contrast, pragmatic factors entirely determine peak alignment.

3aSC16. Acoustic/prosodic characteristics of repetitions and substitutions in medical dictations. Guergana K. Savova (Univ. of Minnesota, Minneapolis, MN and Lernout & Hauspie, 5221 Edina Industrial Blvd., Edina, MN 55439)

This paper describes the results of a pilot study on the acoustic characteristics of two types of disfluencies exact repetitions and exact substitutions. The study was conducted at Linguistic Technologies, Inc., a com-

pany that employed automatic speech recognition (ASR) to medical dictations. Disfluencies present a significant problem to the current state-of-the-art ASR (Nakatani and Hirschberg, 1994; Lickley, 1994; Shriberg, 1994; Plauche and Shriberg, 1999). The acoustic/prosodic characteristics of disfluencies have been proposed as an additional knowledge source for improving recognition results (SRI studies, Price and Ostendorf, 1996). Disfluencies are of interest to psycholinguistics (Levelt, 1983; Clark and Wasaw, 1998). The study focuses on two talkers with a total of 67 disfluencies. The disfluency dissection is based on the Repair Interval Model (Nakatani and Hirschberg, 1994). Measurements of $F0$ were taken at the onset and offset peaks of the reparandum and the repair and the filled pause at the disfluency site. The duration of each word, silence, and filled pause was also measured. Entropic Cambridge Research Labs XWAVES was used for the hand-collected $F0$ measurements. The data gathered allowed the evaluation of several criteria: values of $F0$ variations, intonational contours, comparison of intonational contours, and comparisons of time durations.

3aSC17. Effects of semantic focus on Taiwanese intonation in spontaneous and read speech. Shu-hui Peng (Natl. Chi Nan Univ., Dept. of Foreign Lang. and Lit., 1 University Rd., Puli, Nantou Hsien, Taiwan, shpeng@ncnu.edu.tw)

Previous studies on prosodic effects of semantic focus indicated that sentences produced with narrow semantic focus had different intonation patterns from sentences produced with broad semantic focus. Theses effects include expansion of the pitch range for the focused words and compression of the pitch range for unfocused words in sentences produced with narrow focus. The present study examined the prosodic effects of semantic focus on the intonation patterns of Taiwanese by comparing pragmatically controlled spontaneous speech with read speech. Spontaneous speech was obtained from the dialogue between pairs of speakers participating in a game. The same speakers were asked to read scripts prepared from the original dialogues. Both the read and spontaneous versions were analyzed to show prosodic phrasing and global pitch shapes for different types of sentences. Sentences produced with a narrow focus in the spontaneous version were compared with their read counterparts produced with a broad focus for expansion and compression of pitch range and other tonal features.

3aSC18. The intonation of yes/no and focus constructions in Navajo. Joyce M. McDonough (Dept of Ling., Univ. of Rochester, Rochester, NY 14627, jmmcd@ling.rochester.edu)

This study is a preliminary investigation of tone and intonation in yes/no and focus questions in Navajo (Southern Athabaskan). Navajo is a tone language. The Navajo verb stands as complete proposition, with complex morphological and morpho-syntactic structure. The word final stem is a content morpheme preceded by functional morphemes filling out its propositional content. This stem is marked by acoustic properties commonly associated with stress: longer, louder segments and pitch range expansion, resulting in a striking end-prominent profile [McDonough, Anthro. Ling. **41**, 503–539 (1999)]. Preliminary studies indicated no apparent boundary tones. Research has shown that tone languages may use intonational and/or stress-related strategies (boundary tones, pitch-range expansion) to mark focus and yes/no questions. To investigate the acoustic properties of these constructions in Navajo, 12 native speakers were recorded reciting short statements followed by either yes/no questions (three types) or focus constructions. Both type constructions in Navajo are marked by pro and enclitic-like particles on/surrounding the (nonutterance final) NPs. Results indicate no differences in the $F0$ contour between statements and the contrasting constructions. This fact is arguably related to the NPs nonargument status in the grammar, possibly indicating a consequential interaction of morphology, morpho-syntax, and intonation in these type languages. [Work supported by NSF.]

3aSC19. Prosody in spontaneous and read speech: Some preliminary findings from a tone language. Chao-Yang Lee (Dept. of Cognit. and Linguistic Sci., Brown Univ., Providence, RI 02912)

Intuitively it appears easy to identify whether someone is speaking spontaneously or reading out loud from a text. Previous research has shown that prosody plays a role for this perceptual distinction. The current study examines a number of acoustic prosodic parameters in the production of the two speaking styles in Mandarin Chinese. We ask whether the use of contrastive pitch in this tone language is implicated in the prosodic parameters related to the two speech modes. Spontaneous speech from three female speakers of Mandarin was recorded during interviews with the experimenter. For each speaker, ten declarative sentences were selected and transcribed. The speakers read the transcribed sentences in a second recording session 2 weeks later. Thus, ten pairs of lexically and syntactically identical sentences were obtained. Acoustic analyses were conducted of a number of intonational and temporal parameters (i.e., $F0$, $F0$ range, $F0$ standard deviation, $F0$ declination, and speaking rate). Preliminary analyses show that among the acoustic parameters, only the speaking rate consistently distinguishes between the two speaking styles. This result contrasts with previous research showing the contribution of $F0$-related parameters to the spontaneous-read distinction.

3aSC20. A comparison of conversational and clear speech versions of three Green Mong tones. Jean E. Andruski and Scott Turner (Audiol. and Speech-Lang. Pathol., Wayne State Univ., Detroit, MI 48202)

Green Mong has a phonological tone system that incorporates both fundamental frequency ($F0$) and phonation-type distinctions. Previous work using discriminant analyses examined three tones which have similar $F0$ contours, but which are characterized by the distinctive use of breathy, creaky and modal phonation [Andruski and Ratliff, JIPA (in press)]. Results indicated that the relative amplitude of lower and higher harmonics (measured as the H1–H2 amplitude difference) most reliably discriminates productions of these three tones. The contributions of jitter, duration, $F0$ and shimmer were also significant. This study compares the same three tones in one male speaker's clear speech and conversational speech productions. Clear speech is typically more intelligible than conversational speech. The increase in intelligibility appears to result from exaggeration of acoustic cues that contribute to phonological distinctions in the language. The question explored here is the extent to which cues relating to phonation type are exaggerated in these three tones, as compared to $F0$ and $F0$ contour cues. Results of acoustic analyses of clear and conversational speech tokens will be presented, as well as results of a discriminant analysis using the discriminant functions generated in the previous study from the productions of six other speakers.

3aSC21. Prosody Workstation—Version 2. George D. Allen and John B. Eulenberg (Michigan State Univ., East Lansing, MI 48824)

The Prosody Workstation, Version 2 (PW2) is a low-cost, PC-based system for interactively teaching and researching prosody. This device does four different, inter-related jobs. On output from the host PC, it (1) synthesizes model utterances from stored LPCs and prosodic control parameters. On user input, it (2) extracts the $F0$ and intensity of the user's productions, for comparison with the model as presented. The host PC then (3) visually displays the original model and the user's response, with similarities and differences highlighted for the user's information. Finally, it (4) resynthesizes either the original model or the user's response for auditory comparison. Whereas PW1 used an expensive and no longer available signal processing board, PW2 runs on any PC running Windows and with SoundBlaster compatible sound I/O. PW2 features of interest are (a) scoring algorithms and visual displays are user selectable, permitting (e.g.) time and/or $F0$ normalization, selective focus on specific prosodic elements, and use of different $F0$ scales; (b) data acquisition is efficient, with one $F0$ and intensity estimate calculated for each 10 or 20 ms frame; (c) data extraction and resynthesis routines are coded as C++ DLLs,

while the command modules are visual basic, permitting efficient extension of functionality. [Work supported in part by NIDCD and Michigan State University.]

3aSC22. Broadband sound generation by flow through a dynamic physical model of the larynx. Zhaoyan Zhang, Luc Mongeau, and Steven Frankel (School of Mech. Eng., Purdue Univ., West Lafayette, IN 47907, mongeau@ecn.purdue.edu)

The random component of sound generated by confined jet flows through rubber larynx models was measured for stationary and nonstationary orifice areas. Anechoically terminated rubber hoses upstream and downstream of the orifice minimized acoustic resonance, thereby isolating the source. Radiated sound pressures were recorded, together with static pressures, orifice area, and the mean flow rate. Empirical scaling laws for the radiated sound spectra of jets through stationary orifices of different shapes and areas were first obtained. The walls of the orifice plate were then mechanically driven at frequencies between 50–120 Hz. The periodic component of the radiated acoustic pressure was removed from the microphone signals using ensemble averaging procedures. Statistics of the broadband component of the sound signals were compared for the stationary cases and the nonstationary cases for the same orifice geometry using conditional sampling methods. Very good agreement between the probability density functions was obtained. The instantaneous spectrum of the turbulence-generated sound for nonstationary flows was also found to be comparable to the spectrum of corresponding stationary flows. This confirms the validity of the quasisteady assumption used in aerodynamic models of voice production.

3aSC23. Effect of focus and inter-accent-interval on lexical pitch accents in Japanese. Ito Kiwako (Dept. of Linguist., Univ. of Illinois at Urbana–Champaign, 707 S. Mathews Ave., Urbana, IL 61801, b-creek@staff.uiuc.ed)

The effect of narrow focus and the size of inter-accent-interval (IAI) on the acoustic prominence of pitch accents in Japanese was investigated by measuring $F0$ maxima and variance. Subjects produced utterances containing two accented adjectives, with either 0, 1, or 2 syllables between the accents. Each sentence was produced with focus on the first, second, or none of the adjectives. The results show $F0$ maxima enhancement for the accented syllables under focus regardless of the word position, and attenuation in $F0$ movement peri-focally. Such effect resembles the pattern in Dutch [Rump and Collier, Language Speech **39**, 1–17 (1996)] more than that in English [Eady and Cooper, J. Acoust. Soc. Am. **80**, 402–415 (1986)]. The $F0$ variance during the first accent was significantly smaller in the presence of an accent clash, i.e., IAI=0. This pattern bears an interesting similarity to the stress-clash phenomenon [Cooper and Eady, J. Memory Language **25**, 369–384 (1986)]. The focus effect was larger than the IAI effect, and they did not interact, i.e., the focus effect was not larger for further apart accents. This suggests that, in Japanese, the pragmatic function of pitch prevails over its lexical contrastive function.

3aSC24. Intrinsic effects of obstruent consonants on *F0*. Helen M. Hanson (Sensimetrics Corp., 48 Grove St., Ste. 305, Somerville, MA 02144-1630, hanson@sens.com)

This study is part of a project to improve generation of $F0$ contours in rule-based speech synthesis using HLsyn. When a vowel follows an obstruent consonant, the fundamental frequency in the first few tens of milliseconds of the vowel is influenced by the voicing characteristics of the consonant. In an attempt to model this influence, pilot data were recorded by one subject. The alveolar stops /d,t,t^h/ and the nasal /m/ were paired with the vowels /i,a/ to form CVm syllables. The syllables mVm served as a baseline with which to compare the stop consonants. The target syllables were embedded in carrier sentences. Intonation was varied so that each target syllable was produced with either a H*, L*, or no pitch accent. $F0$ was increased relative to the baseline following all three obstruents, with the greatest increase for the voiceless aspirated stop and the least for the voiced stop. The $F0$ increase and the disparity among the three stops varied with pitch accent: H* showed the greatest influence, while L* showed the least. $F0$ may be increased for up to 80 ms following voicing onset. Data will be presented for additional subjects and obstruents. [Work supported by NIH Grant No. DC04331.]

3aSC25. Synthesis of nonperiodic features of pathological voices. Brian C. Gabelman, Jody Kreiman, Bruce R. Gerratt (UCLA Voice Lab., Div. of Head and Neck Surgery, 31-24 Rehab Ctr., Los Angeles, CA 90095-1794, gabelman@ucla.edu), and Abeer Alwan (UCLA, Los Angeles, CA 90095)

The question has been raised as to how successfully measures of nonperiodic vocal energy such as spectrally shaped aspiration noise, jitter, and shimmer can be used in modeling pathological voices, and how well listeners distinguish between these quantities. This work studies how two such quantities, spectrally shaped aspiration noise and jitter, can be used to synthesize pathological voices. Automatic algorithms for analyzing and modeling these measures are described. Tokens of the vowel [a] were inverse filtered and analyzed using a cepstral-domain comb filter algorithm to extract the aperiodic component of the vocal source. The pitch trajectory was established with an interpolating tracker and was high-pass filtered and analyzed to measure jitter. Synthetic tokens were then computed using the calculated levels of source noise alone, jitter alone, and a ratio of both, and the results compared with the original token in listening experiments. The success of each method at modeling the original token across a variety of voice samples will be described, as will implications for speech synthesis and voice perception. [Work supported by NIH/NIDCD.]

3aSC26. Comparison of *in vitro* growth and tissue behavior of two different fibroblast cell lines embedded on an artificial scaffold. Sanyukta Jaiswal, Ingo R. Titze (Dept. of Speech Pathol. and Audiol., Wendell Johnson Speech and Hearing Ctr., Univ. of Iowa, Iowa City, IA 52242), James A. Martin (Univ. of Iowa, Iowa City, IA 52242), and Steven Gray (Univ. of Utah, Salt Lake City, UT)

The molecular and cellular composition of the vocal folds and the relation of these to the biomechanical and physiological properties of the tissue have been of considerable interest in the current vocal fold research. The fibroblast cells establish and maintain the extra-cellular matrix of the vocal folds. These influence the visco-elastic properties of the tissue, and ultimately its functioning. The present study compares the *in vitro* tissue behavior of two fibroblast cell lines from the human vocal fold tissue and three from human foreskin tissue. Histological and biochemical analyses were used to investigate the inherent cytological differences between the two cell lines when cultured in a synthesized ''matrixlike'' scaffold. [Work supported by NIDCD.]

3aSC27. Reducing expressive variation in speech with synchronous speech. Fred Cummins (Dept. of Computer Sci., UCD, Belfield, Dublin 4, Ireland)

Any two utterances of the same word (or other unit) will differ. The variation arises from the summation of both linguistic and nonlinguistic factors. A major problem in studying prosody has been deciding what are the linguistic, and what are the nonlinguistic factors which collectively influence a measured quantity such as duration or pitch excursion. A substantial reduction in variability can be obtained by having two or more

speakers read a text in synchrony. This task demands that they eliminate most of the nonessential prosodic variability in their speech. Speakers have little difficulty with this task, and can produce highly synchronized speech with little effort and little or no previous practice. A parallel exists with musical performance: ensemble playing is much less variable than the performance of a soloist. Synchronous speech so obtained is potentially useful for distinguishing between linguistic and nonlinguistic sources of variability. It thus may be a useful tool for the experimental phonetician, where speech relatively untainted by individual expressiveness is required. It may also be of use for applications in which a minimization of expressive variation is required. Its suitability for collecting speech to use in a database for concatenative synthesis is currently being assessed.

3aSC28. Running autocorrelation method of *F*0 estimation. Alain de Cheveigné (Ircam-CNRS, 1 place Igor Stravinsky, 75004 Paris, France) and Hideki Kawahara (Wakayama Univ., Wakayama 640-8510, Japan)

A new method of fundamental frequency ($F0$) estimation is proposed that builds upon the standard autocorrelation method. Several principled steps are introduced that considerably reduce the number of errors. Formal evaluation on a database of speech recorded together with a laryngograph signal revealed an order of magnitude fewer gross errors than other standard methods. The method can be implemented efficiently, and can be extended to the task of estimating the periods of multiple concurrent sounds. A parallel may be drawn with time-domain models of auditory perception.

WEDNESDAY MORNING, 6 JUNE 2001 GRAND BALLROOM, 8:00 A.M. TO 12:05 P.M.

Session 3aSP

Signal Processing in Acoustics, Underwater Acoustics, Biomedical Ultrasound/Bioresponse to Vibration and Architectural Acoustics: Coded Signals for Acoustic Applications

Ning Xiang, Cochair
National Center for Physical Acoustics, University of Mississippi, P.O. Box 1848, University, Mississippi 38677

Wing T. Chu, Cochair
Acoustics/IRC, National Research Council, Montreal Road, Ottawa, Ontario K1A 0R6, Canada

Chair's Introduction—8:00

Invited Papers

8:05

3aSP1. Sequences from number theory for physics, signal processing, and art. Manfred R. Schroeder (Drittes Physikalisches Institut, Univ. of Goettingen, Buergerstr. 42-44, 37073 Goettingen, Germany and AT&T Bell Labs. (ret.), mrs@physik3.gwdg.de)

This work will discuss numerical sequences of potential interest to physicists, engineers, and artists. Foremost among these are the maximum-length sequences, based on the prime number 2 and larger primes, which have found wide use in physics, including acoustics, and engineering: (1) the measurement of impulse responses in concert halls, radar echoes from planets (to check the general theory of relativity) and travel times in the deep-ocean sound channel (to monitor water temperature and global warming); (2) the spatial diffusion of sound waves, coherent (laser) light, and electromagnetic waves; (3) algebraic error-correcting codes (SIMPLEX and HAMMING codes); (4) minimizing peak factors for radar signals, synthetic speech, and computer music; and (5) the formation of x-ray images with 2-D masks (in x-ray astronomy). Other sequences to be described include quadratic-residue sequences for the construction of wideband diffusing reflection phase gratings in one and two dimensions; the MorseThue, Fibonacci, and rabbit sequences and their musical potential; and certain self-similar sequences from number theory that engender attractive visual patterns, rhythms, and melodies.

9:05

3aSP2. Application of number theory sequences in architectural acoustics. Peter D'Antonio (RPG Diffusor Systems, Inc., 651-C Commerce Dr., Upper Marlboro, MD 20774) and Trevor J. Cox (Univ. of Salford, Salford M5 4WT, UK)

One of the challenges in architectural acoustics is to develop contemporary diffusing surfaces that complement contemporary architecture in the way that statuary, columns, and relief ornamentation complemented classic architecture. Number theory sequences invented by Schroeder [J. Acoust. Soc. Am. **65**, 958–963 (1979)] and further developed by D'Antonio *et al.* [Appl. Acoust. **60**, 143–165 (2000)] have been used for two decades to design reflection phase gratings, which provide useful sound diffusion in architectural acoustics. The use of modulation techniques, to enhance performance, and the development of proof-of-performance techniques, to evaluate their performance, will be described. The development of flat and curve optimized binary amplitude gratings will also be presented. The current state-of-the-art utilizing boundary-element multidimensional optimization to design 1- and 2-D diffusors, which incorporate aesthetic constraints, will offer a view of future design possibilities. Photos of installations using these acoustic tools will also be presented to illustrate the usefulness of these devices and their impact on architectural acoustics.

9:25

3aSP3. Phase reflection diffusers design using Huffman sequences. James Angus (School of Acoust. & Electron. Eng., Univ. of Salford, Greater Manchester M5 4WT, UK, jamesangus@aol.com)

Phase reflection diffusers based on integer sequences are, unfortunately, limited in their frequency response. The purpose of this paper is to present an alternative form of diffusion structure using noninteger sequence based phase reflection gratings. This paper will present a new method of generating such diffusers based on Huffman sequences. The theory, design, advantages, and limitations of these structures will be discussed and simulation results of their performance will be presented. In particular, their independence from the frequency response effects typical of integer based sequences. Methods of extending and improving their performance will also be suggested. Methods of construction will also be discussed. The paper will show that it is possible to develop diffusing structures which have a better frequency performance than conventional phase reflecting structures.

9:45–9:55 Break

9:55

3aSP4. Exploiting number-theoretical properties of binary *M* sequences for acoustical measurements. Ning Xiang (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, MS 38677) and Manfred R. Schroeder (Univ. of Goettingen, D-37073 Goettingen, Germany)

Due to their excellent number-theoretical properties, binary maximum-length sequences (*M* sequences) have received wide acceptance in many scientific and engineering fields. Since the introduction of *M*-sequence measuring technology into acoustics community in 1979 [J. Acoust. Soc. Am. **66**, 497–500 (1979)] by one of the authors (MRS), advanced measurement techniques have emerged in fields such as architectural acoustics, physical acoustics, ocean acoustics, and ultrasonics. As excitation, binary *M* sequences carry much more signal power within a single period than signals used in the pulse-echo method. Together with its inherent cross-correlation mechanism, this technique enables high noise immunity. A fast algorithm, called a fast *M*-sequence transform, can be applied for performing the cross correlation. A number of noteworthy properties of binary *M* sequences such as shift-and-add, Zech's logarithm, decimation, reciprocity, permutation, auto and cross correlation will be described with the emphasis of their applications in various acoustical measurement.

10:15

3aSP5. The application of *M* sequences to sonar. Harry DeFerrari (Rosenstiel School of Marine and Atmospheric Sci., Univ. of Miami, 4600 Rickenbacker Cswy., Miami, FL 33149, hdeferrari@rsmas.miami.edu) and Theodore Birdsall (Univ. of Michigan, Ann Arbor, MI 48104)

M sequences have long served as enabling technology for numerous basic research experiments to observe the fluctuations of acoustic propagation and for many acoustical oceanographic studies including ocean acoustic tomography. They provide coherent processing gain without loss of temporal resolution. Simply stated *M* sequences combine the energy of CW with the resolution of a pulse. In addition to processing gain, the *M* sequences can eliminate ping stealing and allow for environmentally friendly transmissions (lower peak pressure levels). Yet, early attempts (circa 1960) to apply *M* sequences to monostatic active sonar were unsuccessful. In hindsight, the inability to process Doppler and possibly an inappropriate application of a matched filter correlation are the reasons. Two approaches for *M*-sequence applications to sonar are discussed here, one for passive, whereby clandestine background experiments can evaluate sonar conditions, and the other for a multisource bistatic active. Ultrafast Hadamard transforms make possible full range Doppler processing for both applications and the combination of orthogonal sampling and cancellation by coordinate zeroing are applied to reduce the direct arrivals and some of the stable reverberation components for the active application. The same approach can be used to process and separate simultaneous transmissions from more than one source that transmit nearly orthogonal signals.

10:35

3aSP6. Development of the MLS technique at IRC/NRC, Canada. Wing T. Chu (Inst. for Res. in Construction, Natl. Res. Council, Ottawa, ON K1A 0R6, Canada)

In the early days of digital electronics, MLS (maximum-length sequence) with a very long period has been used primarily as a substitution for the conventional white noise generators employing gas discharge tubes or temperature-limited diodes. Hence, it is also called pseudorandom sequence. But MLS is in fact periodic and deterministic. When exactly one period of data is used, no ensemble averaging is required. It seems that such an application has been overlooked in the field of architectural acoustics until Schroeder

published his works in the late 1970s. This potentially very useful technique was immediately recognized at IRC/NRC and a program was set up to develop the technique with special emphasis on applications to the various architectural acoustics measurements. This talk will trace our development work and present results obtained from the various applications of the technique.

10:55

3aSP7. Explaining angels: The effects of nonlinearity on MLS measurements. Matthew Wright (ISVR Inst. of Sound and Vib. Res., Univ. of Southampton, Southampton SO17 1BJ, UK, mcmw@isvr.soton.ac.uk)

Maximum length sequences are widely used for system identification because their autocorrelation functions are similar to those of white noise. When used to identify systems with weak nonlinearities (in particular those with convergent Volterra series representations) the resulting impulse response estimate can contain spurious echoes, or ''angels'' in radar terminology. These spurious echoes are shown to be a consequence of the fact that maximum length sequences are indeed of maximum length. It is shown that when the interaction between the nonlinearity and the number of theoretic properties of the sequence in question is understood these artefacts can be used to obtain information about the nonlinearity by calculating the Zech's logarithm of the sequence. Examples from audiological research are discussed and ways of choosing optimal sequences are examined.

11:15

3aSP8. Effects of time variance on MLS measurements. U. Peter Svensson and Aslak Bjerkvik (Dept. of Telecommunications, Norwegian Univ. of Sci. and Technol., NO-7491 Trondheim, Norway)

When a system is measured with the MLS technique, any time variance in the system during the measurement will corrupt the resulting impulse response. The effects will look like extraneous noise and, depending on the type of time variance, the spectrum of the noise might be flat or increasing with frequency. Using an analogy with modulation, these effects are analyzed, and examples from measurements in rooms are presented. It is shown that the noise extracted from the initial time delay of a measured impulse response can be analyzed and compared with the background noise. This way, it is possible to detect the presence of extraneous, time-variance induced noise in an impulse response that is measured without averaging. Both intra- and interperiod time variance will cause similar effects, and a second side of the lowered signal-to-noise ratio is that the system's impulse response has had some of its energy transformed into noise. Studies of reverberation are especially sensitive to this since the later parts of the impulse response are affected the most in typical room situations.

Contributed Papers

11:35

3aSP9. Error analysis of HRTFs measured with complementary (Golay) codes. Corey I. Cheng and Gregory H. Wakefield (Dept. of EECS, Univ. of Michigan, 1101 Beal Ave., Ann Arbor, MI 48109, coreyc@umich.edu)

Head-related transfer functions (HRTFs) characterize the spatial filtering properties of the external ear, and can be measured using various system identification techniques. These techniques generally assume that the system under study is time invariant (e.g., motionless). In a recent study, Zahorik presented empirical results concerning the effects of head motion on the accuracy of a popular system identification procedure based on pairs of binary stimulus signals called complementary, or Golay, codes [P. Zahorik, J. Acoust. Soc. Am. **107**, 1793–1796 (2000)]. In this paper, we present a simple model for Zahorik's work which handles not only the effects of head motion, but also measurement noise and dc bias as well. This theoretical formulation confirms the observations made by Zahorik to the extent that head motion can introduce significant errors in both the contour and fine structure of the measured HRTF log magnitude response. Further analysis of this model suggests why another popular system identification procedure based on maximal length sequences (MLSs) does not suffer from these errors. Finally, an unexpected finding is that these errors depend on the spectral shape of the specific Golay pair used during measurement. [Work supported by a grant from ONR.]

11:50

3aSP10. Application of pulse coding to low-level atmospheric turbulent scattering. Stuart G. Bradley and Sabine von Hunerbein (School of Acoust. and Electron. Eng., Univ. of Salford, Salford M5 4WT, UK, S.G.bradley@salford.ac.uk)

A SODAR uses atmospheric turbulence backscatter from several narrow inclined beams to estimate turbulence and wind profiles. Simple averaging-over-range and averaging-over-frequency schemes have been used in some SODAR systems, but generally code techniques are problematic because of the high fractional Doppler shift of 0–0.04. Results are given of detailed simulations, using weather-like targets, of a sequenced comb of frequencies, a chirp, and phase-encoding methods. Three Doppler-adaptive matched filters are described and evaluated against the simulated noisy atmosphere. It is found that the comb of frequencies produce the least variance in estimated Doppler wind speed, and under expected typical turbulence levels, the filter should provide Doppler winds to about 1%. These methods are evaluated in a compact SODAR configuration, with the emphasis on extracting high spatial resolution data over the lowest 20–40 m altitude. This height range has applications in pollutant and agricultural investigations.

Session 3aUW

Underwater Acoustics: The Effect of Environmental Fluctuations on Acoustic Propagation

Mohsen Badiey, Chair
College of Marine Studies, University of Delaware, Newark, Delaware 19716

Contributed Papers

8:00

3aUW1. Intensity fluctuations of refracted bottom reflected arrivals in shallow-water propagation. Harry DeFerrari, Neil Williams, and Hien Nguyen (Rosenstiel School of Marine and Atmospheric Sci., Univ. of Miami, 4600 Rickenbacker Cswy., Miami, FL 33149, hdeferrari@rsmas.miami.edu)

Recent measurements of shallow-water acoustic propagation off the coast of south Florida have found persistent arrivals from low loss propagation paths. The environmental conditions that support this anomalous transmission are common at midlatitudes—specifically the combination of a downward refracting sound-speed profile with upward refracting bottom having a high critical angle, greater than 10 deg. When a source transmits from deep in the channel, refracted modes/rays are trapped above the bottom and reflect from the bottom with grazing incidence less than the critical angle. The combination accounts for low loss propagation. Further, the modes/ rays tend to reinforce and form lines of caustics at every range. The most intense caustic occurs for a receiver located at the source depth. Broadband pulse transmissions yield a single ''focused arrival'' at every range that is usually 25 dB more intense than other arrivals. Intensity fluctuations for the focused arrivals tend to have log-normal distributions with large scintillation even at short ranges (e.g., scintillation index, SI=8 at 10 km), whereas the fluctuations of intensity for other paths exhibit classical statistics with exponential distributions (Rayleigh) and SI=1. Fluctuations are analyzed for several frequencies of transmission and receiver depths.

8:15

3aUW2. Predictability of shallow water sound transmission in selected shallow water experiments. William M. Carey (Dept. of Aero and Mech. Eng., College of Eng., Boston Univ., Boston, MA 02215) and William L. Siegmann (Rensselaer Polytechnic Inst., Troy, NY 12180)

This paper summarizes our efforts to compare calculations in several shallow water areas using range dependent models with results from coherent sound transmission experiments. The approach is to use detailed environmental and geophysical measurements to determine acceptable ranges of input parameters for the range dependent calculations, and then to determine within these bounds the ability to calculate the pressure amplitude, the phase of the signal, and the modal wave number spectrum. In addition, the coherence of the transmitted signal and the measurement of relative array signal gain were estimated. Although shallow water is known for its variability, these investigations clearly demonstrate predictable metrics of sound transmission, provided a sufficient type and an amount of environmental data are known. These investigations also show the key factors in this variability for acoustic predictability and highlight the areas of uncertainty. These results further suggest that shallow water variability and sound transmission can be calculated deterministically with dynamic changing environmental oceanographic variables.

8:30

3aUW3. The PDF of intensity for Ocean WPRM—15 years later. Terry E. Ewart (Appl. Phys. Lab. and School of Oceanogr., Univ. of Washington, Seattle, WA 98105)

Spatial intensity correlations for wave propagation in random media (WPRM) can be predicted in terms of the index of refraction space-time autocorrelation function, with scattering parameters Gamma (strength) and X (scaled range) in the iso-velocity case. In ocean WPRM the temporal intensity correlations are also well predicted from the modeled index of refraction spectrum for the lower path in sub-CZ propagation with a depth-dependent sound speed profile (Eastern Pacific). For both of these cases the intensity pdf's have been shown [Ewart and Percival, **80**, 1745–1753 (1986)] to be very well fit by generalized Gamma distributions. In that work quantile–quantile goodness of fit tests were used to test the generalized Gamma hypothesis. The upper path intensity correlations, in sub-CZ propagation, however, have not been predicted. The pdf's of upper path data taken during MATE will be discussed. Several authors have proposed functional forms of intensity pdf's, e.g., one distribution modulated by another. Quantile-quantile tests will be applied to see if such functions model ocean WPRM intensity pdf's. [Work supported by ONR.]

8:45

3aUW4. Time-varying broadband interference patterns in the acoustic field during a shallow water, fixed-fixed experiment. Gerald L. D'Spain, Lewis P. Berger, Jerry A. Smith, and Eric J. Terrill (Marine Physical Lab., Scripps Inst. of Oceanogr., La Jolla, CA 92093-0704)

During a joint experiment between the Marine Physical Laboratory and the Applied Research Lab, Penn State University in Fall, 2000, continuous-broadband signals in the 100–400 Hz band were transmitted from a moored source to a large-aperture horizontal hydrophone array on the ocean bottom in 175-m water. Spectrograms over a 24-hour period were created from the data from several array elements. These long-duration spectrograms show sinuous interference patterns formed by the regions of modal constructive interference that can be easily traced throughout the entire 24-hour period. The time dependence of the amplitude and center frequency of these individual serpentine structures are correlated with variations in environmental conditions. In particular, their frequency-meandering nature is associated both with tidally driven changes in water depth and the watch circle of the moored source. Temporal oscillations in their amplitude occur with the same periods as those of internal waves. Abrupt jumps in their frequency content occur at the same time as rapid thickening of the upper mixed layer. A simple analytical model is used to predict many of these characteristics. [Work supported by ONR, Code 321US.]

9:00

3aUW5. Coherence of high-frequency signals with emphasis on environmental effects. Joe Keranen, Richard L. Culver, David L. Bradley, Jon Reeves (Appl. Res. Lab., Penn State Univ., P.O. Box 30, N. Atherton St., State College, PA 16804), and Steven D. Lutz (Penn State Univ., State College, PA 16804)

The relationship between the coherence of high-frequency, underwater acoustic signals and the upper ocean acoustic environment has been investigated experimentally. Acoustic and environmental measurements were made under ONR sponsorship (the ARL Program) during October–

November 2000 at a site 30 miles off the coast of San Diego in water depths of 200 to 700 m. The transmitted signals consisted of 0.1- to 1.0-ms CW pulses and linear frequency modulated sweeps. Center frequencies of 18, 32 and 46 kHz and bandwidths from 500 Hz to 13 kHz were employed. The projector was suspended from the floating instrument platform (FLIP) at a water depth of 170 m while the receive hydrophones were suspended from a moored buoy at two different ranges of 640 and 2350 m with water depths of 200 and 700 m, respectively. Environmental measurements included water temperature, current speed and direction, directional wave height and wind speed. Temporal and spectral coherence are reported and compared with environmental patterns, theoretical predictions and prior research. [Work supported by ONR Code 321 (ARL Project).]

9:15

3aUW6. Effects of bandwidth, center frequency, and prefiltering on passive ordinary and higher order correlation detection of broadband transients. Marcella E. Dean, George E. Ioup, Juliette W. Ioup (Dept. of Phys., Univ. of New Orleans, New Orleans, LA 70148, mdean@uno.edu), and Lisa A. Pflug (Naval Res. Lab., Stennis Space Center, MS 39529)

Earlier work done by the authors for passive detection of broadband signals using various functional inputs has resulted in only incremental improvements over the ordinary correlation detection of received signals. Prefiltering of the signal before input to cross-correlation, bicorrelation, and tricorrelation detectors improves detection by as much as 12 dB at a probability of detection (Pd) of 0.5 and a fixed false alarm rate of 0.001. This research is a systematic investigation using various filter widths for broadband chirp transients with bandwidths of 50, 150, and 250 Hz and center frequencies of 130, 430, and 730 Hz. To investigate detector performance, Monte Carlo simulations and hypothesis testing are done to generate ROC curves of Pd vs SNR. Results show that center frequency does not change the optimal filter width; it is solely dependent on the signal width, with decreasing filter size for narrower signals. Finally, the best detection performance is achieved with the tricorrelation detector, for which optimal filter widths vary less with input signal width than those of the optimal filters for the cross-correlation detectors. The tricorrelation detectors are about 3 dB better than the cross-correlation for each input width and center frequency. [Research supported by ONR.]

9:30

3aUW7. Acoustic propagation through surface ship wakes. R. Lee Culver, Xiao Di, Brian Rapids, Steven Lutz (Appl. Res. Lab. and Acoust. Grad. Prog., Penn State Univ., P.O. Box 30, State College, PA 16804, rlc5@psu.edu), and David Bradley (The Penn State Univ., P.O. Box 30, State College, PA 16804)

To predict acoustic propagation through a ship wake, an understanding is required of the physical properties (void fraction, spatial and temporal variation of the bubble structure, oceanic ambient dynamics, for example) of the wake, as well as an acoustic propagation model capable of handling the changes in these properties. The wakes of large surface ships contain high concentrations of air bubbles of various sizes and nonuniform distributions, which give the wake decidedly inhomogeneous acoustic properties. The extant literature contains little quantitative information. Limited material has been collected and combined with a notional hypothesis regarding the random structure to produce a physical model for some of the relevant physical properties. The model has been used as a basis for propagation calculations for this complex situation. Two methods have been employed for those calculations: range-dependent, Gaussian beam ray tracing and a 2-D parabolic equation technique, the latter having previously proven useful for turbulent atmospheric conditions. The acoustic model predictions provide a clear indication of the complexity of the problem and motivation for development of a 3-D propagation model. [Work supported by ONR Code 321 (ARL Project).]

9:45–10:00 Break

10:00

3aUW8. Comparisons of high-frequency acoustic and shallow-water ocean temperature fluctuations. Roger W. Meredith, Steven J. Stanic (Naval Res. Lab., Stennis Space Center, MS 39529), Ralph R. Goodman (Penn State Univ., State College, PA 16804), Edgar T. Kennedy (Naval Res. Lab., Stennis Space Center, MS 39529), and Nickolas G. Pace (NATO SACLANT Undersea Res. Ctr., La Spezia, Italy)

Dynamic ocean processes produce spatial and temporal fluctuations in the ocean's index of refraction along the propagation path that cause acoustic propagation fluctuations. This paper reports a second set of measurements designed to determine environmental limits on ping-to-ping amplitude and phase variability of direct path acoustic fluctuations in very shallow water. In September 1997, the Naval Research Laboratory in conjunction with the NATO SACLANT Undersea Research Centre conducted this series of experiments, in shallow water off American Beach, located between Pisa and Livorno, Italy. Established equations for the variance of the acoustic amplitude and phase as a function of the index of refraction variance and the ocean characteristic length scale under conditions of isotropic turbulence are used to predict acoustic fluctuations for comparison with measurements. In these results, measured phase fluctuations are more consistent with sea swell predictions than isotropic turbulence predictions. Phase fluctuations are relatively insensitive to water column stability and anomalous phase fluctuations correlate with wind direction. Amplitude fluctuations are consistent with isotropic turbulence predictions and anomalous amplitude fluctuations correlate with changing stability. [Work supported by ONR.]

10:15

3aUW9. Internal wave effects on ocean acoustic propagation in shallow water. Duncan P. Williams, Gerald L. D'Spain, and William A. Kuperman (Marine Physical Lab., Scripps Inst. of Oceanogr., Univ. of California at San Diego, La Jolla, CA 92093-0238)

Complementary models are developed to investigate the effect of internal waves on ocean acoustic transmissions in shallow water. An empirical model, based on temperature variations, and a modified Garrett–Munk spectral model of internal wave generated sound speed variations are in turn inserted into an acoustic transmission model. The Garrett–Munk model is itself an empirical model most often used in deep water; in shallow water a different parametric fit is needed to model the complexity found in the environment. The internal wave effects on the acoustic field and, in particular, the patterns of the field extrema, obtained from the numerical modelling are then compared with the results of 2–3 and 5–7 kHz broadband transmissions collected during a shallow water acoustics experiment off the coast of San Diego. The two modelling approaches are used to gain insight into the effects of internal waves on the motion of dislocations; dislocations are places in the sound field where the amplitude goes to zero and so the phase becomes undetermined. [Work supported by ONR.]

10:30

3aUW10. Bispectra of internal waves in shallow seas. Tao Wang (Dept. of Marine Sci., Univ. of Georgia, Athens, GA 30602), Tian-Fu Gao, and Li Ma (Chinese Acad. of Sci., Beijing 100080, PROC)

Densely sampled thermistor chain data have been obtained from several shallow-water acoustics experiments in the Yellow Sea, East China Sea and South China Sea. The data are analyzed to study bispectra of the internal waves observed in this paper. The magnitudes, real and imaginary parts of the bispectra are presented. It is found that strong quadratic interactions occur within the power spectral peak, and also between the peak and its first three harmonics. The real and imaginary parts of the bispectra represent the contributions to the vertical and horizontal asymmetry (skewness and asymmetry) of the internal wave elevations from each fre-

quency pair, respectively. The nonzero asymmetry values indicate the sawtoothlike waveforms. Bispectral analysis is thus a useful means for investigating nonlinear properties of shallow-water internal waves. [Work supported by NSFC.]

10:45

3aUW11. Moments of internal gravity waves. Tao Wang (Dept. of Marine Sci., Univ. of Georgia, Athens, GA 30602), Tian-Fu Gao, and Li Ma (Chinese Acad. of Sci., Beijing 100080, PROC)

Observations have revealed the widespread distribution of internal waves in deep oceans, shallow seas, lakes and reservoirs. Much theoretical, numerical, laboratory and field work has studied their generation and propagation, and in particular significant progress has been made in internal solitary wave research. However, statistical properties of internal waves, especially in shallow water, are still little understood. This paper is to investigate the second-, third- and fourth-order moments of internal waves. Results of internal waves in a strong seasonal thermocline of the Yellow Sea are presented and analyzed. It is found that the skewness coefficients are nonzero and the kurtosis coefficients are greater than 3, indicating skewed waveshapes and non-Gaussian distributions of the internal wave elevations. Statistics have been widely used to describe nonbreaking shoaling sea surface waves. It is suggested that the present study on internal gravity waves could be extended to construct nonlinear models for shallow-water internal waves. [Work supported by NSFC.]

11:00

3aUW12. On acoustic wave field response to internal waves in shallow water for variable source position. Yongke Mu, Mohsen Badiey, Reza Ghias (College of Marine Studies, Univ. of Delaware, Newark, DE 19716), and James Lynch (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

Acoustic variability is examined using a subset of the SWARM'95 experiment data in our earlier work [J. Acoust. Soc. Am. **108**, 2577 (2000)]. Acoustic signals are highly modulated both temporally and spatially due to the presence of internal waves. In addition to energy fluctuations, a large level change in transmission loss was found in the airgun data due to source position changes. The mechanism that contributes to acoustic modulations is mainly due to acoustic mode coupling induced by internal wave propagation. The acoustic modes are decomposed from the data. Mode decorrelation time is examined for different propagation tracks. Fluctuations of the modal correlation indicate an acoustic recorrelation at a longer lag time, which also indicate acoustic intensity variations due to internal waves. Source location was found to be the major factor for high-energy level variations. It is found that the airgun is depth position sensitive. The acoustic intensity and frequency variations due to source position changes are investigated in detail. A depth insensitive source (linear frequency-modulated sweep) was used for comparisons in this study. In order to understand the physics of this problem, acoustic mode shape variations and mode arrival patterns are modeled numerically for a variety of ocean environment profiles.

11:15

3aUW13. Current tomography in coastal regions. Mohsen Badiey, Yongke Mu, Luc Lenain (College of Marine Studies, Univ. of Delaware, Newark, DE 19716), and Stephen E. Forsythe (Naval Undersea Warfare Ctr., Newport, RI 02341)

Reciprocal acoustic transmissions have proved successful for tomography in deep water. However, relatively little effort has been focused on shallow-water environments where sound propagation and scattering takes place often over very short ranges in both the ocean (due to volume and surface) and the seabed. In these regions the effect of depth variations on sound speed is negligible and the salinity and temperature are dominant causes of volume fluctuations. Nevertheless, the methodology can still be utilized to determine the current. In a well-calibrated high-frequency (0.6–18 kHz) acoustic experiment, the range averaged horizontal current in two different directions is determined based on the differential travel times recorded from three reciprocal source–receiver stations. These tripod stations were arranged in a triangular configuration. Data were obtained in nearly isovelocity condition with slight variations in the salinity profile during several tide cycles. The feasibility of obtaining range averaged current profile in very shallow-water regions for long-term measurement is demonstrated. It is also shown that the current can be obtained accurately by utilizing not only the refracted (direct) path, but also the surface reflected arrivals during relatively calm surface conditions. The accuracy of this approach depends on the choice of center frequency and bandwidth of the signal. [Work supported by ONR and Sea Grant.]

11:30

3aUW14. Noise generation by breaking waves in the surf zone. Ali R. Kolaini (Mech. and Environ. Eng. Dept., Univ. of California, Santa Barbara, CA 93106) and Steve Means (Naval Res. Lab., Washington, DC 20375)

During the fall of 1997 numerous researchers investigated the physical processes of the surf zone on a barred sandy beach on the coast of North Carolina. The Naval Research Laboratory participated by deploying a 24-phone bottom-mounted hydrophone array. Measurements to characterize the sound generated by breaking waves in the surf zone were made. The sea surface elevations were simultaneously measured with pressure gauges deployed near the array. A video camera was used to capture the breaker images. The spectral analysis of the data demonstrates the presence of a strong acoustic field in the frequency range of 20 Hz to 3 kHz. Features of the spectral shapes and their relationship to the waveheight spectra are discussed. We present convincing data to support the generation of larger bubbles as being the dominant source of the low-frequency sound in the surf zone. In the stormy days the spectral level begins to attenuate at frequencies as low as 400 Hz. This provides evidence of the existence of the large number of bubbles with radii as large as a few centimeters that were kept in the water column by the turbulent structures for an extended period. Finally, the waveguide effect on the propagation of the low-frequency sound is discussed.

11:45

3aUW15. High-frequency ocean ambient sound: Estimating wind speed using a 150-kHz ADCP. Len Zedel (Phys. and Physical Oceanogr. Dept., Memorial Univ. of Newfoundland, St. John's, NF A1B 3X7, Canada, zedel@physics.mun.ca)

Ocean ambient sound is well correlated with wind speed for frequencies between 2 and 50 kHz. At higher frequencies it is conventional wisdom that thermal noise will dominate the wind-generated sound. Data are presented showing that background sound levels recorded by a 150-kHz ADCP demonstrate strong correlation with wind speed and with independent measurements of ambient sound (below 70-kHz frequency). When the ADCP data are calibrated using scaling factors available from the manufacturer, the observed 150-kHz sound levels are consistent with the expected $\simeq 20$ dB/decade slope of the high-frequency ambient sound spectrum. Using calibrations based on two independent deep ocean deployments, 150-kHz ADCP background sound levels can be used to estimate ocean wind speeds with an accuracy of 1.5 ± 1.5 m s^{-1} for wind speeds up to 15 m s^{-1}. [This research was funded by R. D. Instruments.]

Session 3pAO

Acoustical Oceanography and Underwater Acoustics: Inverse Methods for Sub-Bottom Surveys II

Altan Turgut, Cochair
Naval Research Laboratory, Acoustics Division, Code 7127, 4555 Overlook Avenue, SW, Washington, DC 20375

David N. Lambert, Cochair
Naval Research Laboratory, Code 7431, Stennis Space Center, Mississippi 39529

Contributed Papers

1:00

3pAO1. Minimum samples sizes required to minimize variance and bias of geoacoustic estimates of seafloor properties. Ian Ingram, Aaron M. Thode, and Nicholas C. Makris (MIT, 77 Massachusetts Ave., Cambridge, MA 02139)

Analytic expressions for the first order bias and second order covariance of a maximum-likelihood estimate (MLE) [Naftali and Makris, J. Acoust. Soc. Am. **107**, 2859 (2000)] are used to rigorously determine the sample sizes necessary for a geoacoustic MLE to become unbiased and attain the CRLB. The analysis is applied to the specific case of a narrowband deterministic signal recorded on a vertical array. The biases and variances are computed for bottom sound speed, density, and attenuation in a Pekeris waveguide, for situations where the source location is either known, or must also be estimated from the acoustic data. The case of a two-layer bottom with a sound-speed gradient is also investigated. As the beamformed signal-to-noise ratio (SNR) descends below 0 dB, the bottom sound speed and density bias can exceed 100 m/s and 1 g/cc, repsectively, and at least 100 data samples are required to attain the CRLB. An optimum source range exists for minimizing bottom speed bias, and the bias sign is strongly dependent on source depth. These formulas can provide powerful tools for the design and interpretation of geoacoustic experiments.

1:15

3pAO2. Experimental design for geoacoustic parameter inversion. Mischa F. P. Tolsma, Adriaan van den Bos, and Gerrit Blacquière (Dept. of Appl. Phys., TU Delft, The Netherlands, m.tolsma@tn.tudelft.nl)

Both model-based and nonparametric classification benefit from a theoretical investigation into the information contained in reflections from the subbottom. Specifically, it is difficult to find out if the properties of a certain bottom type are identifiable with a given acoustical signal due to the complexity of the reflection models used. This problem is approached with the Fisher information matrix of the stochastic model for the measured acoustic response. The problem is unidentifiable when this matrix becomes singular. Its inverse, the Cramér–Rao lower bound, gives the best possible precision of the parameter inversion. This lower bound can be used to design appropriate signals for precise measurement of the subbottom. This has been done for a model of a surface layer with varying density given by Lyons and Orsi [IEEE J. Oceanic Eng. **23**, 411–422 (1998)]. It is found that precise measurement of the reported range of model parameters requires a signal with a frequency range of 1–50 kHz. This supports the finding that broadband or multifrequency techniques are needed for the classification of layers with continually varying properties. [Work supported by TNO TPD and Rijkswaterstaat, The Netherlands.]

1:30

3pAO3. Modal mapping in range and azimuth for shallow-water waveguides. Kyle M. Becker (MIT/WHOI, Joint Prog. in Oceanogr. and Oceanogr. Eng., Woods Hole Oceanogr. Inst., Woods Hole, MA 02543) and George V. Frisk (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

It is well established that the normal mode solution for sound propagation in shallow water is a function of the local waveguide properties. Much work has been done on the inverse problem where the local modes are measured and used as input data for geoacoustic inversion methods. In the range-dependent case, short sliding-window transform methods based on the Hankel transform have been implemented to extract the corresponding range-varying wave number spectra. A major objective of this work is to extend the range-dependent work to fully three-dimensionally varying environments. Recently, experiments were conducted using cw sources operating in the frequency range 20–475 Hz. The acoustic field was measured on an array of freely drifting hydrophone buoys equipped with GPS navigation. Using the complex pressure measured on these synthetic aperture horizontal arrays, local mode information is extracted as a function of range and azimuth from the source. Emphasis is placed on modal variability for tracks both along and across bathymetric contours at several frequencies. [Work supported by ONR.]

1:45

3pAO4. Complete acoustic recording of water column scatters and sea bed characteristics. Hans Petter Knudsen, Melle Webjrn, Misund Ole Arve, Ona Egil (Inst. of Marine Res., P.O. Box 1870, N-5024 Bergen, Norway), Aksnes Dag, and Sejrup Hans Petter (Univ. of Bergen, N-5020 Bergen, Norway)

A new vessel for marine science in Norway has been commissioned for delivery in 2003. The 78-m long, 16.5-m wide, 6000-kW powered vessel is designed for multifunctional operation. Facilities for sampling the physical and chemical characteristics of the water column, recording and sampling plankton and fish stocks, mapping the sea bottom topography, and marine geological and geophysical research constitute the scientific framework of the vessel. To eliminate fish avoidance, the vessel will be silent according to the recommendation given in ICES CRR 209, and at night the ship can be darkened. The acoustic instruments for plankton and fish recordings, sea bottom mapping and characterization are described.

Session 3pID

Interdisciplinary: Hot Topics in Acoustics

Robin O. Cleveland, Chair
Department of Aerospace and Mechanical Engineering, Boston University, 110 Cummington Street, Boston, Massachusetts 02215

Chair's Introduction—1:55

Invited Papers

2:00

3pID1. Hot topics in architectural acoustics: New materials. Brandon D. Tinianov (Johns Manville Tech. Ctr., 10100 West Ute Ave.) and Neil A. Shaw (Menlo Sci. Acoust., Inc., Topanga, CA 90290-1610)

While both new metrics for measuring sound and analytical software programs for predicting sound performance have proven to be powerful tools of architectural acoustics, new acoustical materials have allowed us to further engineer these sensitive soundscapes. These developments embrace several areas of contemporary acoustical materials. HVAC liners have seen the addition of fiber-free alternatives to traditional linings. New solutions include coated polyimides and melamine foams. In electroacoustics, materials that have improved performance are now replacing some old standbys. PEN materials such as Kaladex are finding more acceptance for use in tweeter domes and microphone elements. Liquid crystal polymer fibers, such as Vectran, which can be used instead of aramid and carbon fibers, offer similar performance improvement for wet lay-up direct radiator transducer cones. Absorptive panels have moved beyond fabric wraps. They are now paintable and can mimic typically reflective treatments. Finally, diffusers have shown advances in both acoustical performance and geometries that better blend with aesthetic goals. These new surfaces consist of a wide range of surface type including modulated reflection phase gratings, binary amplitude gratings, and most recently boundary-element optimized surfaces of constant or variable impedance. Each of these exciting developments are reviewed in this presentation.

2:20

3pID2. Hot topics in noise. Bennett M. Brooks (Brooks Acoust. Corp., 27 Hartford Turnpike, Vernon, CT 06066)

In response to increasing worldwide public concern about noise, new methods and means for assessing and controlling noise have been developed. Outdoors, advancements have been made in understanding the role that soundscapes play in the quality of life in our communities, which range from urban to rural environments. Recent efforts have been devoted to quantifying and preserving the resource of natural quiet in US National Parks. The impact of theme parks and amphitheaters on their host communities is an active area of research. International development of environmental noise policy is producing significant action. Indoors, new investigations on room noise descriptors are advancing that area of noise practice. The development of standards for classroom acoustics, including practical limits on noise, has the potential to improve the educational experience for all children. Product design research for sound quality will reduce perceived noise levels, and also make the resulting sound more pleasant. Important recent findings on occupational noise effects include new criteria for noise exposure, and the real-world effectiveness of hearing protectors. For many people, acoustics is chiefly represented by noise. These advances in the practical application of acoustics promise to contribute to a more civilized and hospitable society.

2:40

3pID3. Hot topics in structural acoustics: Optical imaging of surface vibrations—Some new developments and applications. Yves H. Berthelot (Woodruff School of Mech. Eng., Georgia Inst. of Technol., Atlanta, GA 30332-0405)

Laser Doppler vibrometry is routinely used for a vast range of applications to measure surface vibrations in the nanometer range. Some new developments include the use of laser vibrometers with continuous tracking on moving targets, the detection of the vector displacement of a surface (either with multiple beam interferometry or with a photo-emf probe), and the direct measurement of torsional vibrations on rotating shafts. Electronic speckle pattern interferometry (ESPI) is a full-field optical technique that is also commonly used to image surface vibrations. An interesting development of ESPI is electronic shearography. Some applications of these optical techniques will be discussed. These include calibrations of transducers and MEMS devices, imaging of rotating parts, mapping of structural intensity, characterization of the dynamic response of viscoelastic materials, as well as some more exotic applications such as damage assessment of frescoes and painting, three-dimensional vibration imaging of the ossicular chain in cats, and dynamic characterization of human teeth.

3:00

3pID4. An investigation of research trends in acoustics from 1970 to 1999. Frederick M. Pestorius (Office of Naval Res., Intl. Field Office, PSC 802 Box 39, 09499-0039) and John A. Viator (Beckman Laser Inst., Univ. of California, Irvine, CA 92612)

Text data mining is a burgeoning field in which new information is extracted from existing text databases. Computational methods are used to compare relationships between database elements to yield new information about the existing data. Text data mining software was used to determine research trends in acoustics for the years 1970, 1980, 1990, and 1999. Trends were indicated by the number of published articles in the categories of acoustics using the *Journal of the Acoustical Society of America* (JASA) as the article source. Research was classified using a method based on the Physics and Astronomy Classification Scheme (PACS). Research was further subdivided into world regions, including North and South America, Eastern and Western Europe, Asia, Africa, Middle East, and Australia/New Zealand. In order to gauge the use of JASA as an indicator of international acoustics research, three subjects, Underwater Sound, Nonlinear Acoustics, and Bioacoustics, were further tracked in 1999, using all journals in the INSPEC database. Reserach trends indicated a shift in emphasis of certain areas, notably Underwater Sound, Audition, and Speech. JASA also showed steady growth, with increasing participation by non-US authors, with about 20% in 1970 to nearly 50% in 1999. [Work supported by ONR.]

WEDNESDAY AFTERNOON, 6 JUNE 2001 CRYSTAL ROOM, 1:00 TO 2:10 P.M.

Session 3pMU

Musical Acoustics: New Synthesis Techniques II

James W. Beauchamp, Chair
School of Music, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, 2136 Music Building, 1114 West Nevada, Urbana, Illinois 61801

Invited Paper

1:00

3pMU1. Physically informed stochastic modal sound synthesis. Perry R. Cook (Dept. of Computer Sci. (also Music), Princeton Univ., 35 Olden St., Princeton, NJ 08544, prc@cs.princeton.edu)

A general and flexible sound synthesis method is described, which is based on controlled random excitations and parameter modifications of resonant filters. Parameters can be derived from simulations of particle/friction systems, from data extracted from physical experiments, from signal analysis (a developing yet still open topic of research), or from simple experimentation and ''tweaking.'' The algorithm has proven convincing and useful for modeling the sounds of many common objects and interactions in music and everyday life. Particle system examples include maracas, windchimes, coins in a pocket, ice cubes in a glass, shoveling gravel, breaking glass, etc. Friction systems include coins rolling along a table, squeaking floor panels and hinges, etc. Aside from being flexible and realistic, the algorithm is extremely efficient by nearly any standard of computation or memory use. Most exciting is the ability to drive and modify the parameters in real time, resulting in virtual objects and systems which are ''playable'' in musical performance, or easily useable for sound effects in games and virtual reality.

Contributed Papers

1:25

3pMU2. Analysis and resynthesis of the steady state portion of acoustic tones. Josef Jurek (Dept. of Computer Sci., The Univ. of Chicago, 1100 E. 58th St., Chicago, IL 60637-1581)

Variations in amplitude and frequency over time that exist in tones produced by acoustic instruments are often difficult to capture when attempting to resynthesize such tones. In the present study, to derive a successful solution to this problem and to further our understanding of such spectral variations over time, the amplitude envelopes of each overtone during the steady state portion of six musical tones were modeled as a spectra and used for resynthesis. A number of techniques for minimizing analysis artifacts in creating these envelope spectra are explored and evaluated as well as some techniques for manipulation of these spectra before resynthesis.

1:40

3pMU3. Scat singing generation using a versatile speech manipulation system, STRAIGHT. Hideki Kawahara and Haruhiro Katayose (Faculty of Syst. Eng., Wakayama Univ., 930 Sakaedani, Wakayama 640-8510, Japan)

A set of procedures to generate scat singing by manipulating a small set of seed voices using a speech manipulation system called STRAIGHT [Kawahara *et al.*, Speech Commun. **27**, 187–207 (1999)] is proposed. $F0$ adaptive spectral smoothing based on a second-order cardinal spline combined with an $F0$ extractor based on a fixed-point analysis from filter center frequencies to the output instantaneous frequency enables the STRAIGHT system to generate a highly natural manipulated singing sound. Group delay manipulation for generating the excitation source signal introduces new control flexibility in source characteristics. $F0$ trajectories are generated using a dynamical model based on $F0$ feed-forward

control regulated by auditorily mediated feedback [Kawahara *et al.*, *Vocal Fold Physiology*, edited by H. Fletcher and P. Davis, pp. 263–278 (1996)]. Effects of the spectral interpolation function and interactions between $F0$ and the spectral envelope are also discussed and demonstrations of a scat chorus are presented. [Work supported by CREST and MEXT, Japan.]

1:55

3pMU4. Enhanced digital music glove. Uipil Chong and Changwon Lee (Computer Eng. Dept., Univ. of Ulsan, Ulsan, South Korea 680-749)

Most of electronic music works are created off line using studio editing techniques and hand-crafted computed-generated or sequence control streams [M. Goldstein, Systems, Man, and Cybernetics, 1998 IEEE International Conference, Vol. 2, pp. 1076–1079]. In this paper a digital music glove (DMG) system playing musical scores in real time is implemented. Furthermore, the present commercial product, TouchGlove, connected to an I-CubeX Digitizer [http://www.infusionsystems.com/products] becomes a finger tip controller for drum machines, but this DMG, by selecting any instrument in the sound card through the interface with the player can be used for many other instruments including the drums. Additionally, this new system controls the pitch, duration, chorus, reverberation, and velocity of sound by updating the pitch wheel change message, polyphonic aftertouch, and channel pressure according to the MIDI file generation. The computer, according to the status of the fingers and foot switches, generates the sound signals and sends them to the speaker systems through the application C++ program and MIDI message. [Work supported by Korean Ministry of Communication and Information.]

WEDNESDAY AFTERNOON, 6 JUNE 2001 RED LACQUER ROOM, 1:00 TO 2:05 P.M.

Session 3pSP

Signal Processing in Acoustics: Distinguished Lecture

James V. Candy, Chair

Lawrence Livermore National Laboratory, University of California, L-156, P.O. Box 808, Livermore, California 94551

Chair's Introduction—1:00

Invited Paper

1:05

3pSP1. A microphone array for hearing aids. Bernard Widrow (Dept. of Elec. Eng., Stanford Univ., Stanford, CA 94305)

A directional acoustic receiving system is constructed in the form of a necklace including an array of two or more microphones mounted on a housing supported on the chest of a user by a conducting loop encircling the user's neck. Signal processing electronics contained in the same housing receives and combines the microphone signals in such a manner as to provide an amplified output signal which emphasizes sounds of interest arriving in a direction forward of the user. The amplified output signal drives the supporting conducting loop to produce a representative magnetic field. An electroacoustic transducer including a magnetic field pick-up coil for receiving the magnetic field is mounted in on the user's ear and generates an acoustic signal representative of the sound of interest. The microphone output signals are weighted (scaled) and combined to achieve desired spatial directivity responses. The weighting coefficients are determined by an optimization process. By bandpass filtering the weighted microphone signals with a set of filters covering the audio frequency range and summing the filtered signals, a receiving microphone array with a small aperture size is caused to have a directivity pattern that is essentially uniform over frequency in two or three dimensions. This method enables the design of highly directive hearing instruments, which are comfortable, inconspicuous, and convenient to use. The invention provides the user with a dramatic improvement in speech perception over existing hearing aid designs, and particularly in the presence of background noise and reverberation.

Plenary Session, Business Meeting, and Awards Ceremony

Katherine S. Harris, President

Acoustical Society of America

Annual Meeting of the Membership

Presentation of Certificates to New Fellows

Paul J. Abbas
Shira L. Broschat
Søren Bech
Angelo J. Campanella
René Caussé
Michael F. Dorman
Stan E. Dosso
E. Carr Everbach
Brian G. Ferguson
Michael P. Gorga
John H. Grose
Takayuki Hidaka
Paul C. Hines
Christy K. Holland
Jun-ichi Kushibiki
Nathan C. Martin
Cynthia F. Moss
Virginia M. Richards
Mario A. Ruggero
Christian Soize
Thomas L. Szabo
Bernhard R. Tittmann
Emily Tobey
William A. Watkins

Presentation of Awards

R. Bruce Lindsay Award to Andrew J. Oxenham

Helmholtz–Rayleigh Interdisciplinary Silver Medal to William M. Hartmann

Gold Medal to Herman Medwin

Session 4aAA

Architectural Acoustics: A Walking Tour of Three Chicago Theatres—An Acoustical Perspective

Ian B. Hoffman, Chair
The Talaske Group, Inc., 105 North Oak Park Avenue, Oak Park, Illinois 60301

8:30

Chair's Introduction and Tour Orientation

This session has been organized as a walking tour of three theaters in the Chicago Loop. The tour will include a mix of new and old, large and small theaters. At each facility, a tour and technical discussion will be led by the acousticians and architects who have most recently contributed to the design or renovation of the venue.

The theaters and speakers include:

Auditorium Theater—Carl Giegold of Kirkegaard Associates will discuss ongoing renovations of this large touring multi-use hall.

Oriental Theater—A member of Daniel Coffey and Associates, Architects, will discuss the rehabilitation of this vaudeville theater.

Goodman Theater—Gary Madaras of The Talaske Group will discuss the recent design, construction and opening of this 800-seat non-reinforced drama theater facility.

There is about a 15-minute walk between the hotel and between each of the theaters. A 45-minute lecture will be presented at each venue. The last lecture should end by noon or shortly thereafter.

There is no charge for the tour and no transportation will be provided. Space is limited so advance registration is required. Register for this tour by sending your name and contact information to Ian Hoffman (fax: 708-524-2818 or e-mail: ian@talaske.com).

Session 4aABa

Animal Bioacoustics and Signal Processing in Acoustics: Information Theory Analysis of Animal Acoustic Communication

John R. Buck, Chair
CMAST, 706 South Rodney French Boulevard, New Bedford, Massachusetts 02744

Chair's Introduction—8:00

Invited Papers

8:05

4aABa1. The relationship between the amount of information and the value of information in animal communication. Jack Bradbury (Cornell Lab. of Ornithology, 159 Sapsucker Woods Rd., Ithaca, NY 14850)

How much information should senders try to provide and receivers try to extract from signals? Using a simple but generalizable model of communication, it will be shown that the net payoff to either party for communicating, called the value of information, is not a linear, or even monotonic, function of the amount of information. Because receivers invariably have a default (no-signal) strategy that is better than chance, a signal must exceed a minimal threshold of accurate information before receivers should attend to it. Because costs of producing and analyzing increasingly accurate signals tend to rise faster than benefits above the minimal threshold, there is also an upper limit on how much information is worth the effort. These conclusions have important implications for how new signals are likely to evolve, and how much error we should expect to see in real animal communication systems.

8:30

4aABa2. Entropic structure of chickadee calls. Jack P. Hailman (Archbold Biological Station, Old SR 8 (County 17), Venus, FL 33960)

Language-like ''chick-a-dee'' calls of the black-capped chickadee are combinations of four note types with repetitions but no permutations. Written English has higher values of entropy (bits/letter) than calls (bits/note) at all orders of a Markov chain because it has 26 elements compared with only four. More importantly, the values of bits/letter falls evenly with increasing orders of Markov chain, reflecting long-range dependencies, whereas values of bits/note drop noticeably from entropy based on probability of occurrence to that based on probability of transition to the next note. This latter structure has been called ''semi-Markovian'' and reflects the lack of permutation in this chickadee's calls. The ''entropic structure'' of calls also differs from that of English words, which have a theoretically immense channel capacity. Nevertheless, constraints on English are so severe that actual words on the average have less than twice the source entropy of calls (11.8 bits/word vs 6.7 bits/call). English is far more informative than chick-a-dee calls, though, because words are themselves combined into meaningful sentences whereas successive chick-a-dee calls appear to be merely repetitive.

8:55

4aABa3. The behavioral ecology of animal vocal communication: An information theoretic approach. Brenda McCowan (Comparative Behavioral Biol. Lab., VMTRC-UC Davis, 18830 Rd. 112, Tulare, CA 93274, bmccowan@vmtrc.ucdavis.edu), Laurance Doyle, and Sean Hanser (SETI Inst., Mountain View, CA 94043)

Innovative approaches for analyzing animal vocal communication are needed to investigate the behavioral ecology underlying the composition and structure of animal vocal communication systems. This approach will require a quantitative comparative analysis of animal vocalizations from different populations and species. To date, few studies have quantitatively examined vocal and repertoire structure in relationship to social or ecological strategies in populations of mammalian species. This lack of information has hindered our understanding of the behavioral ecology and evolution of mammalian vocal communication. We will present and discuss quantitative comparative measures from information theory that provide important insights into the ecological and social processes that shape the structure and organization of vocal repertoires. Information theory provides a powerful and efficacious approach because it links the physical, biological, and social properties of the environment to the structure of communication systems on multiple analytical and conceptual levels. Using two phylogenetically unrelated species as examples (i.e., bottlenose dolphins, squirrel monkeys), we will consider each level of communication, from the acoustic structure within call types to the complexity of call sequences, for a comprehensive approach toward animal vocal communication and its relationship to human language systems. [Work supported by NIH, Marine World Foundation, and SETI Institute.]

9:20

4aABa4. Information entropy—interpretation and estimation. Ryuji Suzuki (Speech and Hearing Sci. Prog., Harvard–MIT Div. of Health, Science and Technol., 77 Massachusetts Ave., E25-518, Cambridge, MA 02139, rsuzuki@mit.edu)

Information theory originated in a series of papers by Shannon in 1948, where he sketched an important theoretical framework for characterizing information sources and communication channels. Subsequently, information theory has primarily studied theoretical problems based on known source and channel properties. Although Shannon's theory was in spirit applicable to a wider class of problems, information theory has paid less attention to analysis problems, where an observed signal is accessible to the researcher but the underlying true mechanism is unknown, e.g., statistical estimation and hypothesis testing problems. This paper reviews the properties of information entropy and common techniques for estimating and interpreting it. The performance and pitfalls of common entropy estimators are illustrated with examples. Specifically, the naive technique of substituting relative frequencies or empirical distribution into the definition of entropy is shown to produce inaccurate estimates. An alternative technique for analyzing observed data, the sliding window match length estimator, is relatively simple to implement, can be intuitively understood, and is mathematically well supported.

Contributed Paper

9:45

4aABa5. A comparison of two information theory approaches to animal communication systems. Laurance R. Doyle, Jon M. Jenkins (SETI Inst., 2035 Landings Dr., Mountain View, CA 94043, ldoyle@seti.org), Sean F. Hanser (Six Flags Marine World, Vallejo, CA 94589), and Brenda McCowan (Univ. of California, Davis, Tulare, CA 93274)

Information theory has been used to analyze and explore animal communication systems using two different approaches. The first approach—which is the original application outlined by Shannon and Weaver (1949)—considers the temporal distribution (i.e., probabilities of occurrence) of the components of a communication system to quantify the diversity and complexity of a system's (repertoires) organization. This quantification of the communication complexity is based on the information content, not just at the repertoire level, but also at the sequential (i.e., Markovian) organizational levels of the signals constituting the communication system. The second approach, which has been much more widely applied, with mixed results over several decades, considers a communication system's ability to essentially ''encode'' signaling between individuals in order to measure how much information has been ''transmitted.'' We will outline and contrast these two approaches, pointing out that until the communication complexity of the entire signaling system is quantified, there is little hope of correctly interpreting information exchange between individuals using that system of communication [C. Shannon and W. Weaver, *The Mathematical Theory of Communication* (University of Illinois Press, 1949)].

Session 4aABb

Animal Bioacoustics: General Topics

John R. Buck, Chair
CMAST, 706 South Rodney French Boulevard, New Bedford, Massachusetts 02744

Contributed Papers

10:15

4aABb1. Assessment of ecosystem biodiversity by acoustic diversity indices. Stuart H. Gage, Brian M. Napoletano (Dept. of Entomology, Michigan State Univ., East Lansing, MI 48824, napoleta@msu.edu), and Michael C. Cooper (Ferris State Univ., Big Rapids, MI 49307)

Assessment of ecosystem biodiversity by measurement of acoustic diversity was explored [B. L. Krause, Explorers J., Winter, 156–160 (1993)]. Specific acoustic indices (e.g., based on frequency spectrum) were developed and correlated with standard diversity indices (e.g., standard species abundance indices). Necessary technological infrastructures and analytic processes to measure acoustic dynamics of ecological biodiversity were explored. An automated web-based infrastructure capable of capturing, processing, and relaying real-time field measurements from multiple ecosystems to desktop and home computers was developed, tested, and implemented. Key infrastructure components were remote field instrumentation, remote computer processing, wireless digital relay instrumentation, Internet server, and automation, relational database, and website software. A dynamic digital library of ecological acoustic samples, correlated ecosystem attributes, and acoustic analysis methodologies was established. Library resources, including digital sound files captured from ecosystems, were made available to researchers and the public over the Internet [N. Metzger and M. Blumenthal, Realizing Info. Future, NAP, 113–119 (1994)]. Indices of acoustic ecosystem diversity were refined through application on existing digital ecosystem sound recordings and digital sounds captured from multiple ecosystems. The performance of these indices was compared to standard biodiversity indices applied to the same ecosystems.

10:30

4aABb2. Thermal damage analyses of human skin cells by scanning acoustic and atomic force microscopies. Chiaki Miyasaka, Jikai Du, and Bernhard R. Tittmann (Dept. of Eng. Sci. and Mech., Penn State Univ., University Park, PA 16802, cxm48@psu.edu)

The acoustic microscope (SAM) is a well-known tool for imaging biological cells with high resolution. Here images of living human skin cells (HaCat) in a culture liquid (dMEM+10% FCS) are presented. The cells were grown onto square glass plates (15.0 mm $\times$ 15.0 mm), which were mounted on the bottom of a plastic well, which in turn was placed on the temperature-controlled stage of our scanning acoustic microscope (Olympus UH-3). Images were obtained at a frequency of 600 MHz as the temperature was increased from 37.5 °C to the threshhold temperature for cell death (atoptosis). The temperature was increased 1 °C every 3 min. The acoustic beam from a high numerical-aperture (NA) lens was focused onto the cells by monitoring the absorption of the reflected waves. By using image processing (DSP), features could be identified, such as change in thickness, cell rounding, lateral shrinkage, and atoptosis. The shapes and dimensions were confirmed by Atomic Force Microscopy (AFM). The SAM and AFM are thus effective tools for characterization of living cells.

10:45

4aABb3. On the structural acoustic modeling of cetacean mammals by the finite element method. Martin H. Marcus and Charles F. Gaumond (Naval Res. Lab., 4555 Overlook Ave. SW, Washington, DC 20375, marcus@astro.nrl.navy.mil)

Requirements for the computational modeling of a mammal are presented. In general, meticulous attention to details of its anatomy is needed. The extent of detail must go beyond size and shape to include an inventory of each organ and bone. Also, some components are unique to a species, such as the melon in dolphins. The elastic properties of several biological materials must be determined. A few of these materials are bone, connective tissue, and fat. The compressional sound speeds of the above materials in humans are easily available. Transverse sound speeds and anisotropic information are somewhat less abundant. The compressional and transverse sound speeds are sufficient to model isotropic materials, such as fat, but anisotropic information is necessary for materials that have a grain, such as connective tissue. The modeling of bone may require an understanding of the separate layers of bone, which are quite different from each other. A simple FEM model of a dolphin is presented. The model is run over a range of frequencies to examine the nature of resonances that can be found in such a structure. It is found that the bone structure plays a significant role in the resonance behavior. [Work supported by ONR.]

11:00

4aABb4. A framework for annotating animal bioacoustic data. Kazuaki Maeda and Steven Bird (Ling. Data Consortium, Univ. of Pennsylvania, 3615 Market St., Ste. 200, Philadelphia, PA 19104-2608, maeda@ldc.upenn.edu)

Annotation graphs (AGs) provide an efficient and expressive data model for linguistic annotations of time–series data [Bird and Liberman, 2001]. The term annotation here refers to the association of structured symbolic information with an extent of signal data. This paper demonstrates that AGs also provide an excellent framework for annotating time–series animal bioacoustic data. The AG model has several desirable properties for annotating animal bioacoustic signals. Partial or incomplete annotations can be represented naturally. Hierarchical structures, such as internal structures of animal communicative signals, can be stored and manipulated effectively. The model permits efficient query using the relational database technology. The spreadsheet-like editor component described in the paper provides an intuitive user interface for creating annotations of bioacoustic signals, such as monkey vocalizations, as AG data. The paper describes the tool used for annotating the vocalization of vervet monkeys, and the preliminary findings of the project are presented. Finally, the paper lays out a complete infrastructure for supporting the rapid development of annotation tools for animal bioacoustic data [S. Bird and M. Liberman, ''A Formal Framework for Linguistic Annotation,'' Speech Commun. **33**, 23–60 (2001)].

11:15

4aABb5. Multiple scattering in a reflecting cavity: Application to fish counting in a tank. Philippe Roux and Julien De Rosny (LOA/CNRS, 10 rue Vauquelin, 75005 Paris, France)

Classical fisheries acoustics techniques are useless in the presence of multiple scattering or reflecting boundaries. A general technique is developed that provides the number and the scattering strength of fish in motion placed inside a highly reflecting cavity. This approach is based on multiple scattering theory. The idea is to measure the average effect of the fish on the acoustic echoes of the cavity interfaces. This leads to the measure of the scattering mean free path, a typical length that characterizes the scattering strength of the fish school. Numerical results are shown to agree with a simple theoretical analysis. Experiments are performed with fish in a tank at two different scales: ultrasonic frequency (400 kHz) in a 1.4-l beaker with 1-cm-long fish as well as fisheries acoustics frequency (12.8 kHz) in a 30-m^3 tank with 35-cm-long fish. These results have interesting applications to fish target strength measurement and fish counting in aquaculture.

THURSDAY MORNING, 7 JUNE 2001 PDR 17, 8:00 A.M. TO 12:00 NOON

Session 4aBB

Biomedical Ultrasound/Bioresponse to Vibration: Ultrasound Bioeffects

J. Brian Fowlkes, Chair

Department of Radiology, University of Michigan Medical Center, Kresge III, R3315, 200 Zina Pitcher Place, Ann Arbor, Michigan 48109-0553

Chair's Introduction—8:00

Invited Papers

8:05

4aBB1. Overview of ultrasound-induced lung hemorrhage, and a comment about the mechanical index. William D. O'Brien , Jr., Leon A. Frizzell (Dept. of Elec. and Computer Eng., Univ. of Illinois, 405 N. Mathews, Urbana, IL 61801, wdo@uiuc.edu), Douglas G. Simpson (Univ. of Illinois, Champaign, IL 61820), and James F. Zachary (Univ. of Illinois, 2001 S. Lincoln, Urbana, IL 61802)

It has been well documented that ultrasound-induced lung hemorrhage can occur in mice, rats, rabbits, pigs, and monkeys. Our own experimental studies have focused on mice, rats, and pigs as animal models. The characteristics of the lesions produced in mice, rats and pigs were similar to those described in studies by our research group and others, suggesting a common pathogenesis for the initiation and propagation of the lesions at the macroscopic and microscopic levels. We demonstrated that lung hemorrhage in mice was not caused by inertial cavitation. Threshold estimates and superthreshold behaviors were investigated as a function of species (adult mice and rats) and ultrasound frequency (2.8 and 5.6 MHz), and showed that mice and rats were similar in sensitivity to lung damage with similar and frequency-independent threshold estimates. A 3×5 factoral-designed study (3 exposure durations and 5 pulse repetition frequencies) at the same superthreshold exposure level in adult mice and rats showed that pulse repetition frequency and exposure duration both affect the percentage of animals with lesions. The research reported in these studies and others we have done proves increasing evidence that the mechanical index is inconsistent with observed findings. [Work supported by NIH Grant No. HL58218.]

8:30

4aBB2. Diagnostic ultrasound with contrast agent induces membrane damage in monolayer cells *in vitro*. Douglas Miller and Jawaid Quddus (Univ. of Michigan Medical Ctr., Ann Arbor, MI 48109-0553)

Interaction of diagnostic ultrasound with contrast agent gas bodies can induce membrane damage in monolayer cells. Epidermoid (A431) and phagocytic (RAW-264.7) cells were grown on 5 micron thick Mylar sheets, which formed the upper windows for 1 mm deep exposure chambers. For exposure, a 3.5 MHz transducer from an acoustic imaging (model 5200B) scanner in spectral Doppler mode was aimed upward at the chamber, 7 cm away, in a 37 °C water bath. Damage was indicated by sonoporation (uptake of large fluorescent molecules) and cell lysis (trypan blue staining). Apparent thresholds for membrane damage with 1% optison and 1 min exposure after 30 s rise and warmup time were about 0.12 MPa for both cells lines. The gas bodies tend to stick to the phagocytic cells, but not to epidermoid cells. When incubated with 5% agent for 15 min and then rinsed to remove unattached gas bodies before exposure, the RAW cells retained the 0.12 MPa threshold, but effects were greatly decreased for epidermoid cells. These results indicate relatively low thresholds for membrane damage, and suggest a potential for targeting of phagocytic cells for membrane damage by diagnostic ultrasound with contrast agents. [Work supported by NIH Grant No. CA42947.]

8:55

4aBB3. Effects of ultrasound on the heart. Diane Dalecki (Dept. of Biomed. Eng. and The Rochester Ctr. for Biomed. Ultrasound, Univ. of Rochester, Rochester, NY 14627)

Ultrasound and lithotripter fields can affect the function of the heart. A single, pulse of ultrasound can produce two qualitatively different effects on the heart. When the pulse is delivered during systole, it can cause a reduction in the developed aortic pressure. Evidence suggests that this effect arises from an action of radiation force on the cardiac tissues. A premature cardiac contraction results when the ultrasound pulse is delivered during diastole. The threshold for producing a premature contraction with a 5-ms pulse at 1.2 MHz is on the order of 5 MPa. The presence of ultrasound contrast agents in the vasculature significantly enhances the effect. When Albunex is present, the pressure threshold for the production of a premature cardiac contraction is reduced by nearly an order of magnitude. In addition, when contrast agents are present, premature contractions can be produced with diagnostically relevant pulse durations and pressure amplitudes. Other researchers have observed premature contractions in humans using clinical imaging devices.

9:20

4aBB4. Bubble and cavitation-assisted hyperthermia (BACH). R. Glynn Holt, Patrick A. Edson, Xinmai Yang, and Ronald A. Roy (Dept. of Aerosp. and Mech. Eng., Boston Univ., Boston, MA 02215, rgholt@bu.edu)

Rapid hyperthermia resulting in tissue necrosis has proven to be a useful therapeutic modality for clinical application of high-intensity focused ultrasound. At therapeutic intensities, the hyperthermia is often accompanied by bubble activity. *In vitro* and *in vivo* experiments alike have shown that under certain conditions bubble activity can give rise to a doubling of the heating rate. With a view toward harnessing the energy-concentrating effects of bubbles to do useful clinical work, we report the results of experiments and modeling for the dynamic and thermal behavior of bubbles subjected to 1-MHz ultrasound at megapascal pressures. The dominant heating mechanism depends on bubble size, medium shear viscosity number, and frequency-dependent acoustic attenuation. The bubble size distribution, in turn, depends on insonation control parameters (acoustic pressure, pulse duration), medium properties (notably dissolved gas concentration), and bubble-destroying shape instabilities. The evidence obtained so far points to a range of control parameters for which bubble-enhanced heating can be assured. [Work supported by DARPA.]

9:45

4aBB5. Standardizing the estimation of *in situ* ultrasound exposure in beams with strong nonlinearity. Francis Duck (Medical Phys. Dept., Royal United Hospital, Bath BA1 3NG, UK)

This paper will discuss alternative approaches to the problem of predicting in situ exposure from acoustic measurements made in water. Ultrasonic pulses from medical diagnostic scanners commonly exceed sigma$=1$ in the focal region in water, and the associated amplitude-dependant loss of energy from the beam causes in situ exposure estimates from linear assumptions to be in error. Alternative approaches have been and will be explored in order to establish robust methods of universal applicability. One approach requiring low amplitude operation followed by linear scaling shows promise. Such an approach may overestimate in situ exposure, making it acceptable as a means for safety management. Standard techniques will need to be established to set upper thresholds during measurement. These may place a limit on the acoustic pressure at the source or at the focus, or on a derived quantity such as sigma(m). Alternatively a limit may be placed on the allowed harmonic level at the measurement point. Other approaches will be reviewed, including those using tissue equivalent media, and those computing in situ exposure directly from high-amplitude free-field measurements in water. It is noted that procedures appropriate for manufacturing standards may not be appropriate for individual patient dosimetry.

10:10–10:30 Break

Contributed Papers

10:30

4aBB6. Role of pulse repetition frequency and exposure duration in mice and rats with lung hemorrhage induced by exposure to superthreshold pulsed ultrasound. William D. O'Brien, Jr., Leon A. Frizzell (Dept. of Elec. and Computer Eng., Univ. of Illinois, 405 N. Mathews, Urbana, IL 61801, wdo@uiuc.edu), and James F. Zachary (Univ. of Illinois, Urbana, IL 61802)

Superthreshold behavior for ultrasound-induced lung hemorrhage was investigated in adult mice and rats at an ultrasound center frequency of 2.8 MHz to assess the role of pulse repetition frequency (PRF) and exposure duration (ED). One hundred fifty 6-to-7-week-old mice and 150 10-to-11-week-old rats were each divided into 15 exposure groups (10 animals per group) for a 3-by-5 factorial design (3 EDs: 5, 10, 20 s; 5 PRFs: 25, 50, 100, 250, 500 Hz). The *in situ* (at the pleural surface) peak rarefactional pressure of 12.3 MPa [MIs of 6.3 (mice) and 7.1 (rats)] and the pulse duration of 1.42 μs were the same for all ultrasonically exposed animals. The lesion depth and surface area were measured for each animal as well as the percentage of animals with lesions per group. The proportion of lesions in both species was related statistically to PRF and ED, with the exception that PRF in rats was not quite significant; the PRF$\times$ED interaction (number of pulses) for lesion production was not significant for either species. The PRF, but not ED, significantly affected lesion depth in both species. Both PRF and ED significantly affected lesion surface area in mice. [Work supported by NIH Grant No. HL58218.]

10:45

4aBB7. Cardiopulmonary function in rats with lung hemorrhage induced by exposure to superthreshold pulsed ultrasound. Jeffery M. Kramer, Tony G. Waldrop (Dept. of Molecular and Integrative Physiol., Univ. of Illinois, 407 S. Goodwin, Urbana, IL 61801), Leon A. Frizzell (Univ. of Illinois, Urbana, IL 61801), James F. Zachary (Univ. of Illinois, Urbana, IL 61802), and William D. O'Brien, Jr. (Univ. of Illinois, Urbana, IL 61801)

The study's objective was to assess cardiopulmonary function in rats exposed to superthreshold ultrasound known to produce lung hemorrhage. In nine rats, five ultrasound-induced lesions were produced in the left lung, and in six rats, five ultrasound-induced lesions were produced on each lung. Lesions were induced using superthreshold conditions (3.1-MHz, 1.7-kHz PRF, 1.3-μs pulse duration, 60-s exposure duration, 17-MPa *in situ* peak rarefactional pressure, MI of 5.8). Lesions volumes were measured and approximately 3.4% and 7.9% of the lung was damaged in one-sided and two-sided exposures, respectively. Cardiopulmonary data

were quantified before and 30 min after the last exposure: measurements included pulsatile arterial pressure, heart rate, end-tidal carbon dioxide, respiratory rate, and blood gases. In one-sided exposures, there were no statistically significant differences in cardiopulmonary function, pre- and post-exposure. In two-sided exposures, statistically significant differences were detected for arterial pressure and arterial pO2. We concluded that damage of one lung does not affect peripheral indicators of lung function because of functional respiratory capacity retained in unexposed lung. However, when two lungs were damaged, the functional respiratory reserve was decreased to the point where rats were unable to maintain resting levels of arterial pO2. [Work supported by NIH Grant No. HL58218.]

11:00

4aBB8. Repair of lung lesions following exposure to pulsed ultrasound. James F. Zachary (Dept. of Veterinary Pathobiology, Univ. of Illinois, 2001 S. Lincoln Ave., Urbana, IL 61802, zacharyj@uiuc.edu), Leon A. Frizzell, and William D. O'Brien, Jr. (Univ. of Illinois, Urbana, IL 61801)

This is the first study to examine the long-term effects of ultrasound-induced hemorrhage in lung. The reparative responses were characterized in a rat lung following exposure to pulsed ultrasound. Forty five adult rats were exposed at two sites, the left sixth and ninth intercostal spaces [3.1 MHz, in situ (at the pleural surface) peak rarefactional and compressional pressures of 17.0 MPa and 39.7 MPa, respectively, 1.7-kHz PRF, 1.2-μs pulse duration, and MI of 5.7]. Lung lesions were evaluated macroscopically and microscopically at 0, 1, 3, 5, 7, 10, 13, 16, 19, 23, 27, 30, 34, 37, and 44 days post exposure. Macroscopic lesions progressed between 0 and 44 days post exposure from large bright red ellipses of hemorrhage to small areas of light yellow–brown discoloration. Microscopic lesions progressed between 0 and 44 days post exposure from areas of acute alveolar hemorrhage, to areas of intense cellular proliferation, to areas with small quantities of hemosiderin pigment and septal fibrosis. Alveoli returned to near normal morphology and functional diameter by 44 days post exposure. Under our exposure conditions, lesions induced by ultrasound do not appear to have long term residual effects in lung. [Work supported by NIH Grant No. HL58218.]

11:15

4aBB9. Evaluation of the acoustic boundary conditions at the pleural surface that causes ultrasound-induced lung hemorrhage. William D. O'Brien, Jr., Rita J. Miller (Dept. of Elec. and Computer Eng., Univ. of Illinois, 405 N. Mathews, Urbana, IL 61801, wdo@uiuc.edu), Jeffery M. Kramer, Tony G. Waldrop, Leon A. Frizzell, and James F. Zachary (Univ. of Illinois, Urbana, IL 61802)

In a previous study [J. Acoust. Soc. Am. **108**, 1290 (2000)], the acoustic impedance difference between intercostal tissue and lung was evaluated as a possible explanation for the enhanced lung damage with increased hydrostatic pressure, but the hydrostatic-pressure-dependent impedance difference alone could not explain the enhanced occurrence of hemorrhage. A hypothesis was suggested that the animal's breathing pattern might be altered as a function of hydrostatic pressure, which in turn might affect the volume of air inspired and expired. The acoustic impedance difference between intercostal tissue and lung would be affected with altered lung inflation, thus altering the acoustic boundary conditions. In this study, the rat's respiratory system was inhibited, a ventilator was used to control lung volume, and superthreshold ultrasound exposures of the lungs were conducted (3.14-MHz, 1000-Hz PRF, 1.3-μm pulse duration, 10-s exposure duration, 16.9-MPa *in situ* peak rarefactional pressure, MI of 5.8). Deflated lung was more easily damaged than inflated lung. The acoustic impedance difference between intercostal tissue and lung is much less for the deflated lung condition, suggesting that the extent of lung damage is related to the amount of acoustic energy that is propagated across the pleural surface boundary. [Work supported by NIH Grant No. HL58218.]

11:30

4aBB10. Role of beamwidth in lung hemorrhage induced in adult rats by exposure to pulsed ultrasound. Leon A. Frizzell, William D. O'Brien, Jr. (Dept. of Elec. and Computer Eng., Univ. of Illinois, 405 N. Mathews Ave., Urbana, IL 61801), Douglas G. Simpson, and James F. Zachary (Univ. of Illinois, Urbana, IL 61802)

The objective of this study was to assess the role of beamwidth on ultrasound-induced lung hemorrhage in rats. A total of 144 10- to 11-week-old 257±21-g female Sprague–Dawley rats were randomly exposed to pulsed ultrasound at three exposure levels and four beamwidths (12 rats/group). The three in situ peak rarefactional pressures were 5.0, 7.4 and 10 MPa. Four 19-mm-diam focused transducers were used to obtain the −6-dB beamwidths of 470 μm (2.8 MHz, f/1), 930 μm (2.8 MHz, f/2), 310 μm (5.6 MHz, f/1), and 510 μm (5.6 MHz, f/2) at the pleural surface. Exposure durations were 10 s, pulse repetition frequencies were 1 kHz, and pulse durations were about 1 ms. The lesion surface area and depth were measured for each rat as well as the percentage of rats with lesions per group. Each lesion parameter exhibited a significant dependence upon beamwidth, but no dependence on frequency. Logistic regression analysis was used to determine an ''effective dose'' threshold for 5% (ED05) probabilities of lesions. The ED05 exhibited no dependence on the beamwidth. [Work supported by NIH Grant HL58218.]

11:45

4aBB11. Cavitation mediated rat lung bioeffects from diagnostic ultrasound. Christy K. Holland (Dept. of Radiol., Univ. of Cincinnati, 234 Goodman St., Cincinnati, OH 45267-0761), Ronald A. Roy (Boston Univ., Boston, MA), Paul W. Biddinger, Christopher J. Disimile, and Cheryl Cawood (Univ. of Cincinnati, Cincinnati, OH)

Several animal models have exhibited petechial hemorrhage in the lung within the current FDA output limit of diagnostic ultrasound systems. To elucidate the mechanism of damage, seven rat lungs were simultaneously exposed or sham exposed *in vivo* to 6-MHz pulsed Doppler ultrasound and interrogated with a confocally aligned 30-MHz active cavitation detector (ACD) [R. A. Roy, S. Madanshetty, and R. E. Apfel, J. Acoust. Soc. Am. **20**, 2451–2455 (1990)]. The right lung lobes of four 200 g rats were insonified with an ATL HDI 3000 6.0-MHz Doppler pulse for 1.5 min and three rats were sham exposed. At the termination of ultrasound or sham exposure, each animal was immediately euthanized with sodium pentobarbital (200 mg/kg IP) and lung tissues were removed intact, inflated with buffered formalin, and examined by a pathologist who was blind to the exposure conditions. Histologically, damage was observed in the lungs exposed to ultrasound as extravasation of erythrocytes into the alveolar spaces. The ACD signal received was analyzed for the presence of increased scattering and radiated noise from inertial cavitation. The variance of the sequentially obtained ACD received pulses correlated with the damage in the exposed lungs. We conclude that inertial cavitation mediated the damage from diagnostic ultrasound. [Work supported by NIH R29 HL58761-02.]

Session 4aEA

Engineering Acoustics and Underwater Acoustics: Modern Magnetostrictive Materials

Harold C. Robinson, Cochair
Naval Undersea Warfare Center, Code 2131, 1176 Howell Street, Newport, Rhode Island 02841

Stephen C. Butler, Cochair
Naval Undersea Warfare Center, 1176 Howell Street, Newport, Rhode Island 02841

Chair's Introduction—8:15

Invited Papers

8:20

4aEA1. Modern magnetostrictive materials. James B. Restorff, Marilyn Wun-Fogle (Naval Surface Warfare Ctr., Carderock Div. Code 645, 9500 MacArthur Blvd., West Bethesda, MD 20817-5700, restorffjb@nswccd.navy.mil), and Arthur E. Clark (Clark Assoc., Adelphi, MD 20783)

Two classes of magnetostrictive materials are emerging which can strongly impact magnetomechanical transduction and high-stress antivibration systems. One class utilizes highly magnetostrictive rare earths with anisotropic $4f$ electron shells. The second class is based upon $\alpha-$Fe, with substitutions of nonmagnetic Ga and Al for Fe. The first class contains three distinct types of materials: (a) hexagonal $Tb_{1-x}Dy_x$ alloys, (b) cubic CsCl type $Tb_{1-x}Dy_xZn$ alloys, and (c) cubic Laves phase Terfenol alloys. While the first two types produce extraordinarily high magnetostrictions only at cryogenic temperatures, Terfenol alloys exhibit large magnetostrictions (>1000 ppm) at 200 °C and above. This paper will focus on recent measurements on low-hysteresis Terfenol-DH alloys taken between -40 °C–$+80$ °C. The new class of promising high-strength, low-cost, Fe-based alloys will also be introduced. Here the emphasis is on the development of a ductile, high tensile-strength material with high strains (>300 ppm), high relative permeabilities (>100), and low excitation fields (<300 Oe) under high loads. Magnetostriction, magnetization and elastic constant measurements as a function of stress and magnetic field will be presented. Values of piezomagnetic constants and coupling factors, calculated from these measurements, will be included. [Work supported by ONR.]

8:45

4aEA2. TERFENOL-D manufacturing practices and their design impacts. Jon Snodgrass (Etrema Products, Inc., 2500 N. Loop Dr., Ames, IA 50010, jon.snodgrass@etrema-usa.com)

The manufacture of TERFENOL-D has a substantial impact on the possible designs of acoustic transducers. As with all transducer materials, a better understanding of the manufacturing process can enable the designer to avoid various pitfalls that impact not only performance but cost as well. This presentation will in essence be a primer on the manufacture of giant magnetostrictive materials. Substantial changes have taken place since the invention of the materials. The history of these changes and the subsequent design limitations will be discussed. The current state-of-the-art of TERFENOL-D production will be presented, and key features that have design impacts will be highlighted, especially where the design possibilities have been increased through the refinement of the production techniques. This will focus primarily on the crystal growth of TERFENOL-D but will also include details of characterizing the material and the machining of the driver to its final configuration, which has a substantial impact on cost. Current efforts to further refine and improve these processes will also be presented to indicate the future capabilities that can be expected.

9:10

4aEA3. Magnetostrictive transducer performance characterization. Alison Flatau and Rick A. Kellogg (Aerosp. Eng. & Eng. Mech., Iowa State Univ., IA)

An overview of work done on the characterization of magnetostrictive materials and magnetostrictive transducers will be presented. Experimental results obtained from a test device sized for use with 0.25 in. diameter by 2.0 in. long samples will be presented. Tests were conducted at constant temperatures ranging from 0–75 °C as the sample underwent cyclic magnetic fields and/or cyclic mechanical stresses. Test results include characterization of samples of Terfenol–D under varied temperatures, cyclic stress, and cyclic fields. The output force––strain relationship typical of the performance of Terfenol–D transducers under varied operating conditions is examined to study transducer blocked force characteristics. The design and construction of a transducer for testing under controlled thermal, magnetic, and mechanical load conditions are described. Results of compression tests used to generate load lines and the blocked force characteristics of the transducer will be discussed. Comparisons of the transducer's force and strain output are made with published data. This test data is also used to examine the variability in Young's Modulus with applied magnetic field, strain, and stress. [Work supported by NSF and the Carderock Division of the Naval Surface Warfare Center is gratefully acknowledged.]

9:35

4aEA4. Large strain particulate magnetostrictive composites. Geoff McKnight and Greg P. Carman (Mech. Aerosp. Eng. Dept., UCLA, 38-137M Eng. IV, Los Angeles, CA 90095, carman@seas.ucla.edu)

In this presentation we will describe recent progress on manufacturing particulate magnetostrictive composites that produce large strains ($>$1600 microstrain). The composites are manufactured by combining Terfenol-D particulate with a polymer resin in the presence of a magnetic field. The field aligns the particulates into continuous chains similar to continuous fiber composites. Magnetostrictive particulate composite research has been conducted by a variety of research institutions during the last decade. Most studies report values for saturation strain on the order of 800 microstrain, however, UCLA found that by combining selected particles the saturation strain increases to 1200 microstrain. While this represents an improvement, these strain levels (1200) were still below those reported for the monolithic material (i.e., 1800 microstrain). During the last year UCLA discovered a manufacturing technique that increases the saturation strains to values comparable with the monolithic material (1700 microstrain). A representative figure comparing the response of the monolithic material to a 20% volume fraction particulate composite is provided in the figure in this paper. Note that the strain outputs from the composite are similar to those from the monolithic material. Both these large strains and the large bandwidths provided by the particulate composites suggest opportunities exist for use in high power sonar devices.

10:00–10:15 Break

10:15

4aEA5. Potential new transducer materials: Ferromagnetic shape memory alloys showing 6% field-induced strain. Robert C. O'Handley, Christopher P. Henry, Jorge Feuchtwanger, Miguel Marioni, Pablo G. Tello, David Bono, and Samuel M. Allen (D.M.S.E., MIT, Cambridge, MA 02139)

Very large field-induced strains ($e=6\%$) have been reported for Ni–Mn–Ga single-crystal ferromagnetic shape memory alloys (FSMAs) at room temperature. While these materials show the conventional thermoelastic shape memory effect upon transformation to the high-temperature phase, the magnetic field-induced strain occurs fully within the low-temperature, martensitic phase by twin-boundary motion. Fields of 320–400 kA/m (4–5 kOe) are sufficient to produce the maximum strain under opposing stresses of order 1 MPa. The blocking stress observed here is approximately 2 MPa but can be up to 10 MPa. An ac test system is described that applies an axial load to the sample via a spring. A magnetic field applied transverse to the load causes axial extension. Result indicate that the effective magnetostrictivity, d31 (output strain per unit field), is little changed over the actuation-frequency range accessible with our present power supplies, namely 320 Hz. The hysteresis in the strain-versus-field curves can be as large as 100 kA/m but in some samples is much less. The field dependence and stress dependence of the strain are well accounted for by the simple thermodynamic model. All parameters for the model can be measured from stress-versus-strain curves and magnetization-versus-field curves.

10:40

4aEA6. A quantitative theory of magnetization and magnetostrictive hysteresis in Terfenol-D. William D. Armstrong (Dept. of Mech. Eng., Watson School of Eng. and Appl. Sci., State Univ. of New York–Binghamton, Binghamton, NY 13902, wda@binghamton.edu)

New theory presented successfully reproduces and explains the experimental magnetoelastic hysteresis behavior of Terfenol-D. It is assumed that a very large number of individual domain wall translation events combine to produce each measurable domain transformation in a macroscopic sized sample. Each individual domain wall may be expected to suffer some level of domain wall translation inhibition; however, the severity of the inhibition will spatially vary. It is therefore assumed that the presence of defects in the material increases the directional magnetization potential of subsequent domain states within a process, and distributes nontrivial probabilities of the occupation of $\langle 111 \rangle$ type domain states in a parameter selected, inverse exponential form familiar from the study of statistical thermodynamics. The increased magnetization potential of subsequent high magnetostriction and high magnetization states retards their occupation until higher intensity applied magnetic fields are produced, thus shifting the increasing applied magnetic field curve in a positive field direction and shifting the decreasing applied magnetic field curve in a negative field direction in a form consistent with experiment.

Contributed Papers

11:05

4aEA7. Investigation of eddy current losses in laminated TERFENOL-D drivers. Julie Slaughter (Etrema Products, Inc., 2500 N. Loop Dr., Ames, IA 50010)

One of the most significant loss mechanisms in TERFENOL-D is eddy currents. Performance of the active material is reduced both by heating of the magnetostrictive element, and thus decreased magnetostrictive capability, and also by reduced magnetic field intensities in the active material. Accurate predictions of eddy current losses in TERFENOL-D are necessary to avoid too few laminations, which reduces performance, or too many laminations, which unnecessarily increases costs. Eddy current losses in TERFENOL-D drivers of different configurations were investigated for realistic operating conditions. Drivers of the same size with different lamination configurations were tested at frequencies from 10 Hz up to 8 kHz. At each frequency, energy loss per cycle per unit volume due to eddy currents was calculated. Comparisons were made between the experimental data, theory for infinite sheets, and finite element analysis. Very good agreement was achieved between the FEA and the experiments. Poor agreement was shown between theory and experiment. Eddy current losses are over-predicted by infinite sheet theory. In order to achieve accurate estimations of eddy current losses, more sophisticated analysis methods, such as finite element analysis, must be used.

11:20

4aEA8. Biased resonance measurements as predictors of large-signal performance of PMN-PT. Elizabeth McLaughlin and Harold Robinson (NAVSEA Undersea Warfare Ctr., Div. Newport, Newport, RI 02841)

The large-signal performance of electrostrictive materials, such as lead magnesium niobate-lead titanate, is of critical importance to sonar and actuator designers. However, obtaining these large signal parameters properly, particularly under compressive prestress, is expensive and time consuming. The complexity of these measurements precludes them as a

method for quickly and easily screening materials. Traditionally, resonance measurements, which otherwise are relatively simple to perform, have been used, but they suffer from the drawback that the material parameters obtained are at the incorrect frequency and under no prestress. Furthermore, the significance of the results of resonance measurements for nonlinear materials is unclear. It has recently been suggested [Hom and Shankar, IEEE Trans. Ultrason. Ferroelectr. Freq. Control UFFC-46, p. 1422, 1999] that dc biased resonance measurements on electrostrictive ceramics would be an accurate predictor of the coupling factor and optimum bias point. Here, dc biased resonance measurements on several formulations, with varying dielectric maximum temperatures, are analyzed to determine which composition has the highest predicted coupling factor. This prediction is compared with large signal quasistatic measurements conducted on NAVSEAs SDECS (stress-dependent electromechanical characterization system). The predictive ability of the resonance measurements is analyzed as a function of temperature. [Work supported by ONR.]

THURSDAY MORNING, 7 JUNE 2001 WABASH ROOM, 10:00 TO 11:25 A.M.

Session 4aNS

Noise: Community Noise Ordinances—Topics for Chicago

Robert D. Hellweg, Chair
Compaq Computer Corporation, MR01-3D3, 200 Forest Street, Marlborough, Massachusetts 01752

Chair's Introduction—10:00

Invited Papers

10:05

4aNS1. State of Illinois noise regulations in 2001. Greg Zak (Illinois EPA, 1021 N. Grand Ave. E., Springfield, IL 62794-9276)

Since 1973, the State of Illinois has had noise regulations. Active enforcement of these regulations has been ongoing in different forms depending on what resources are available to accomplish this end. Prior to 1981, Illinois had the funding to actively investigate noise complaints and refer those cases that had merit to the Attorney General's office for prosecution. Loss of funding for both state and federal noise programs resulted in greatly reduced enforcement after 1981. Illinois addressed this situation by making best use of remaining resources in order to initiate innovative noise limit enforcement activities. The result has been more actual noise control from the 1990s to 2001 than was obtained prior to the 1981 programmatic cuts.

10:25

4aNS2. Cook County noise regulation. Louis DeRose (Cook County Dept. of Environ. Control, 69 W. Washington St., Ste. 1900, Chicago, IL 60602, ccdec@wwa.com)

Cook County is an urban county consisting of over 5-million people in a 700-square mile area which includes the city of Chicago. Some 130 villages and municipalities in the area outside the city of Chicago provide a governing and zoning mix that places commercial and industrial entities in close proximity to residential communities. In addition, a complex system of roadways and expressways impacts on nearly all citizens. Through enacting Article 9, Noise and Vibration Control, the Cook County Board of Commissioners has recoginzed the potential for jurisdictional disputes and conflicts and has charged the Department of Environmental Control with the responsibility of enforcing uniform property-line noise standards to ensure the quality of life for our citizens. The Department is often the only source of relief for problems of this nature. However, the Noise Ordinance is under increased scrutiny and confrontation from local governments due to jurisdictional conflicts and is in need of review. This presentation will discuss the status of this ordinance, the methods of enforcement and the status for review and update.

10:45

4aNS3. Noise ordinances: The good, the bad, and the ugly; An overview of more than 200 existing noise ordinances. Blomberg Les (Noise Pollution Clearinghouse, P.O. Box 1137, Montpelier, VT 05601)

The Noise Pollution Clearinghouse (NPC) has collected more than 200 community noise ordinances and state noise regulations and is in the process of putting them in the NPC Online Law Library. This paper provides an overview of existing noise ordinances, examining what's covered and what's missing, as well as what's enforceable and what's unenforceable. In addition, the paper will include a summary of how communities are handling various noise problems such as impulsive noise, boom cars, lawn equipment, air conditioners, etc. Finally, new and emerging issues in community noise control and enforcement are discussed.

11:05

4aNS4. Progress on a Model Community Noise Ordinance Standard. Bennett M. Brooks (Brooks Acoust. Corp., 27 Hartford Turnpike, Vernon, CT 06066) and Lawrence S. Finegold (Finegold & So, Community & Environ. Noise Consultants, Centerville, OH 45459)

The American National Standards Institute (ANSI) Accredited Standards Committee S12, Noise, has charged its Working Group (WG) 41 to develop a model community noise ordinance standard. This ordinance, or group of ordinances, should be suitable for various urban, suburban, and rural communities. This new standard will include, at a minimum, metrics, criteria, and an enforcement methodology. The model ordinance(s) could be used by local authorities around the United States as a basis for their own ordinances or zoning performance regulations. To date, WG 41 has developed a draft ordinance with three sets of criteria. The first set includes familiar A-weighted sound pressure level (SPL) limits for noise emissions at a receiver's property line. The other criteria sets include additional limits on low frequency noise emitters, such as industrial plants, musical entertainment at night clubs, and noise from automobiles. One criteria set uses C-weighted SPL noise limits. The other places limits on the allowable noise levels in specific octave bands, including the low frequency (bass) octave bands. The relative advantages of each criteria set are discussed and guidance is provided to assist local government officials in selecting the most appropriate approach for their community.

THURSDAY MORNING, 7 JUNE 2001 SALONS III/IV, 8:30 TO 11:45 A.M.

Session 4aPA

Physical Acoustics: General Topics in Physical Acoustics

Bart Lipkens, Chair
Mechanical Engineering, Virginia Commonwealth University, 601 West Main Street, P.O. Box 843015, Richmond, Virginia 23284-3015

Contributed Papers

8:30

4aPA1. James Clerk Maxwell and the propagation of sound in gases. Philip L. Marston (Dept. of Phys., Washington State Univ., Pullman, WA 99164-2814)

In the published abstract of his original presentation of statistical kinetic theory, James Clerk Maxwell noted that, ''The author intends to apply his mathematical methods to the explanation on this hypothesis of the propagation of sound . . . ,'' [British Association for the Advancement of Sciences (1859)]. Although it appears that he did not consider the propagation of sound in his subsequent publications on the kinetic theory of gases, Maxwell eventually examined the relationship between the speed of sound and the rms particle velocity in gases. His result for monatomic gases appears in a postscript of a paper communicated by Maxwell [S. T. Preston, Philos. Mag. **3**, 441–453 (1877)]. In Preston's publication, and in Preston's 1876 letter to Maxwell, the related early work of J. J. Waterston is noted. While the postscript has been previously noted [see, e.g., M. Greenspan, Phys. Acoust. **2** A, 1–45 (1965)], other aspects of Maxwell's acoustical activities are less obvious. These include Maxwell's referee reports on manuscripts. The manuscripts by Stokes (1868), Rankine (1870), and Rayleigh (1871) were published and becamse acoustical classics. Maxwell also suggested the title to Rayleigh's, ''The Theory of Sound,'' in an 1873 letter to Rayleigh.

8:45

4aPA2. Propagation of short pulse through arbitrary relaxing media with resonant properties. A new approach. German A. Maksimov and Vladimir A. Larichev (Dept. of High Density Energy Phys., Moscow Eng. Phys. Inst., Kashirskoe sh.31, Moscow 115409, Russia, maximov@dpt39.mephi.ru)

The new analytical representation of fundamental solution (Green's function) describing the short-pulse propagation in medium with single process of resonant relaxation is presented in this report. This analytical solution is based on the generalized local response function of linear media [V. A. Larichev and G. A. Maksimov, Acoust. Phys. **44**, 709–716 (1998)]. It contains well known Lorentz and Ziner models of relaxing media, as particular cases. The changing of pulse shape at propagation, described by obtained solution, shows a variety of forms of pulse propagation and general lows of pulse dynamics beginning from pure relaxation behavior and up to a resonant one. [Work supported by Russian Found Dasic Research (Project No. 00-02-16556).]

9:00

4aPA3. An experiment below the wave diffraction limits with an acoustic sink: the ideal time reversal experiment. J. de Rosny and M. Fink (Lab. Ondes et Acoust., ESPCI, Université Paris 7, CNRS UMR7587, 10 rue Vauquelin, 75231 Paris Cedex 05, France)

Time reversal has successfully demonstrated the ability to focus acoustic waves in the laboratory for several years. However, the minimal focal spot dimension is always larger than half the wavelength, which is in agreement with diffraction limit. We will show that this behavior is related to the presence of a diverging wave that is always created after the convergent wave collapse at the focus. In a time reversal experiment, by replacing the initial source by its time reversed image, we may eliminate this diverging wave. Such a process allows to create an acoustic sink which absorbs the converging wave. Thus, with the elimination of the diverging wave, the focal spot can become much smaller than the one predicted by the diffraction limit. In this paper, we present the first experimental demonstration of an acoustic sink and we observe focal spot smaller than 1/14 the wavelength.

9:15

4aPA4. Numerical investigation of sound diffraction by arbitrarily-shaped thin obstacles. S. Suh, L. Mongeau, and J. S. Bolton (Purdue Univ., 1077 Herrick Labs., West Lafayette, IN 47907)

The sound diffracted by rigid, thin, arbitrarily-shaped obstacles was investigated using a commercially available boundary element code. The numerical procedures were validated by computing the sound field in the shadow region of a rigid circular disk. The results were compared with predictions from available analytical expressions and with experimental data. Excellent agreement was obtained between the predicted and the measured insertion loss values. Scattering from objects having various

random shapes, previously suggested by other researchers [Menounou, Bailey, and Blackstock, J. Acoust. Soc. Am. **107**, 103–111 (2000)], was then investigated. The performance of these objects as sound barriers was compared to that of regularly shaped obstacles. Potential application to linear sound barriers will be discussed. [Work supported by INDOT, JTRP, FHWA, and SQDH.]

9:30

4aPA5. On radiation of sound from open end of an irregular wedge-shaped waveguide. Vladimir T. Matsipura and Victor T. Grinchenko (Inst. of Hydromechanics, NAS of Ukraine, 8/4 Zhelabov St., Kiev 252057, Ukraine, vin-igm@gu.kiev.ua)

The characteristics of sound field radiated from the open end of a wedge-shaped waveguide are studied. A characteristic feature of the waveguide geometry is that the edges of nonparallel half-planes are arranged on different distances from the wedge apex. The additional elongation of the wall is considered as a screen and is an important element in the mechanism of control of the near- and far-field characteristics. To solve the corresponding boundary problem an analytical method has been developed. The central idea of the analytical method and the possibility of one to solve other boundary problems are discussed. The radiated sound field is produced by an incident mode in waveguide. Qualitative data about space structure and energy characteristic of the radiated field are presented. The data about change of the sound pattern when the parameters of the waveguide are changed illustrate the complex structure of the radiated sound field and strong dependence of one on the structure of a normal wave carrying the energy to the edge of the waveguide. Qualitative description of the far-field structure is complemented by data about the concentration factor of sound energy and distribution of energy between transmitted and reflected waves.

9:45

4aPA6. Analog of the Fresnel reflection and refraction coefficients for smoothed boundary between two liquid medias. Nick Maltsev (MVM Intl., 10678D Maplewood Rd., Cupertino, CA 95014)

The Fresnel formulas for reflection and refraction coefficients at the boundary between two liquid medias assume discontinuity of media properties at the boundary. In many physical applications this assumption is too strong. A recent report considers a model problem with the following properties: $n^2(z)=3D[n_1^2\exp(\alpha z)+n_0^2\exp(-\alpha z)/\exp(\alpha z)+\exp(-\alpha z)]\rho(z)=3D[\rho_1\exp(\alpha z)+\rho_0\exp(-\alpha z)/\exp(\alpha z)+\exp(-\alpha z)]$ which are representing a smoothed boundary. An exact solution of Euler equations is derived in a form of plane waves and fast converging series. Physical effects are investigated for reflection and refraction of plane and spherical harmonic and pulse waves in such a model.

10:00–10:15 Break

10:15

4aPA7. Acoustic stop bands and discrete transmission in slender tubes. M. S. Kushwaha (Inst. of Phys., Univ. of Puebla, P.O. Box J-45, Puebla 72570, Mexico, manvir@sirio.ifuap.buap.mx), A. Akjouj, B. Djafari-Rouhani, L. Dobrzynski, and J. O. Vasseur (Univ. of Science & Technol. of Lille-I, 59655 Villeneuve D'Ascq Cedex, France)

Extensive band structures and transmission spectra for longitudinal (acoustic) wave propagation in a system made up of N' dangling side branches (DSB) periodically grafted at each of the N equidistant sites on a slender tube have been computed. A periodic pattern of large stop bands is obtained for the airy DSB on a slender water tube. Noteworthy are the interesting results of huge gaps and discrete transmission spectrum due only to the DSB grafted at a *single* site ($N=1$) on the slender tube. Designing the system with open tubes allows achievement of the lowest gap below a threshold frequency and extending up to zero—thereby providing an entirely discrete band structure and transmission spectrum. This should have important consequences for the suppression of low-frequency noise and for designing filters and transducers.

10:30

4aPA8. Giant acoustic band gaps in a two-dimensional periodic system of fluids. M. S. Kushwaha (Inst. of Phys., Univ. of Puebla, P.O. Box J-45, Puebla 72570, Mexico, manvir@sirio.ifuap.buap.mx) and B. Djafari-Rouhani (Univ. of Science & Technol. of Lille-I, 59655 Villeneuve D'Acsq Cedex, France)

Periodic binary systems can give rise to genuine acoustic stop bands within which sound and vibrations remain forbidden. Extensive band structures for two-dimensional (2D) periodic arrays of air cylinders in water have been computed. Complete, multiple, huge stop bands are found for both square and hexagonal arrangements. The lowest stop bands are largest for a range of filling fraction $10\%\leqslant f\leqslant 55\%$, with a gap/midgap ratio of 1.8. The most interesting finding of the present investigation is that the low-frequency, flat passbands for a perfectly periodic system correspond to the discrete modes of a single airy cylinder. This is attributed to the low filling fraction and huge density contrast in air and water. It is stressed that such a simple inhomogeneous system as made up of air and water exhibits the largest stop bands ever reported for 2D and/or 3D elastic as well as dielectric composites.

10:45

4aPA9. Rayleigh wave propagation in a curved elastic waveguide. John Harris (Ctr. of Quality Eng. & Failure Prevention, Northwestern Univ., Evanston, IL 60204)

A JWKB asymptotic expansion describing in-plane elastic waves is used to approximate a Rayleigh-like wave guided within an elastic waveguide whose curvature is small and changes slowly over a wavelength. The two lowest eigenmodes in a curved guide, taken together, constitute the Rayleigh-like wave. It is shown that this wave lies in the shadows of four closely spaced, virtual caustics. If the curvature becomes too large, the virtual caustic closest to the outer surface enters the guide, after which Rayleigh-like propagation ceases. This wave is dispersive: The dispersion arises both from the confinement within the guide and from the curvature. Propagation in both a thin and a thick guide are examined. Propagation into an environment of increasing curvature is studied, for both waveguides, to exhibit the influence of nearby caustics.

11:00

4aPA10. A coherent method for the measurement of the dispersive curves of elastic waveguides. Hailan Zhang, Shuwu Dai, and Chengyu Zhang (Inst. of Acoust., Chinese Acad. of Sci., Beijing 100080, PROC)

A coherent technique is presented to measure the dispersive curves of elastic waveguides. Waveforms received at several positions along the propagation path are transformed into the frequency domain by FFT. The phase thus obtained has been truncated into the range of the principle value. The true phase difference between the nearest and the farthest waveforms is restored from the truncated phases by summing all phase differences between the contiguous waveforms which are picked up at the positions close enough that the phase differences between contiguous waveforms are less than 4. The restored phase is then used to calculate the phase velocity at different frequencies. The use of the phase information makes the measurement more precise and stable than traditional noncoherent methods. The method is applied to measure the dispersive curves of an aluminum plate with rough surfaces and the result is compared with that of the smooth plate. It is found that the effect of the roughness is the attenuation of the amplitude and slight reduction of the phase velocity. These results agree with the prediction of a theoretical model based on the perturbation method. [Work supported by Nature Science Foundation of China, No. 19774062.]

11:15

4aPA11. Numerical study on 3D acoustic field in borehole generated by eccentric sources. Hailan Zhang and Jianguo Shen (Inst. of Acoust., Chinese Acad. of Sci., Beijing 100080, PROC)

Two new tools for well logging, circumferential acoustic device (CAD) and sector bond test (SBT), are used in oil fields in which the transmitting transducers are either pressed onto or located close to the borehole wall and the acoustic field thus generated is three-dimensional. We calculated the acoustic field along a fluid-filled cylindrical borehole in an elastic medium generated by a point source in the hole and close to the wall. The method called real axis integral is used. The spectrum of the acoustic field in the domain of the frequency and axial wave number is expressed as an infinite series of Bessel functions. Its truncation is analyzed which is dependent on the frequency and radius of the borehole. In the calculation Bessel functions up to the 60th order are used. The spectrum is first calculated and then transformed to the time-space domain. The waveforms at some receiving positions and some snapshots are given in the paper. It is shown that the field consists of the direct wave, reflected waves, and interface longitudinal and transverse waves. [Work supported by Nature Science Foundation of China, No. 19974059.]

11:30

4aPA12. Acoustic method for measurements of the coolant boiling on heat generating surfaces. Sergey V. Boldin (Dept. of Heat and Gaseous Supply, Arch. & Civil-Eng. Inst., 65 Ilinskaya St., 603600, Nizhny Novgorod, Russia) and Andrei B. Kolpakov (Dept. of Phys., 65 Ilinskaya St., 603600, Nizhny Novgorod, Russia)

Acoustic method for determining the period of different duties of the coolant boiling is presented. The foundation of this method is the damping phenomenon of ultrasonic waves travelling through the boundary between different media (steam–liquid) differing widely in their acoustic properties. The experimental setup and the results of the investigations into the period of water film boiling on spherical samples are also presented. The emphasis is on aluminum and copper samples differing in their diameter (4, 5, and 8 mm) for temperature interval $\Delta t = 180–690\,^{\circ}\mathrm{C}$ at atmospheric pressure and different surrounding water temperatures ($\Delta t = 15–75\,^{\circ}\mathrm{C}$). By analyzing the result, it is illustrated that the precision of the method is well above that of thermocouples. It is demonstrated that the measurements using this acoustic method excludes the errors typical of traditional methods.

THURSDAY MORNING, 7 JUNE 2001 RED LACQUER ROOM, 8:45 TO 10:45 A.M.

Session 4aPPa

Psychological and Physiological Acoustics: Speech Perception

Gary R. Kidd, Chair
Department of Speech and Hearing Sciences, Indiana University, Bloomington, Indiana 47405

Contributed Papers

8:45

4aPPa1. Auditory integration of bandpass filtered sentences. Brent Spehar, Nancy Tye-Murray (Central Inst. for the Deaf, 4560 Clayton Ave., St. Louis, MO 63110, bspehar@cid.wustl.edu), and Mitchell Sommers (Washington Univ., St. Louis, MO 63130)

Auditory integration is the ability to merge auditory inputs received from different channels into a single percept. The present experiments investigated auditory integration of speech using lists of filtered CUNY sentences. Within each set, separate filtering preserved either low-frequency or high-frequency spectral information. The three sets of bandwidths were 550–750 and 1650–2250 Hz; 550–800 and 1650–2350 Hz; and 550–850 and 1650–2450 Hz. Four conditions were tested for each set of bandwidths: (1) LF only: (2) HF only; (3) LF–HF monotic (LF and HF filtered sentences presented simultaneously to one ear); and (4) LF–HF dichotic (LF sentence presented to one ear, HF sentences presented simultaneously to the other ear). Scores in the single band conditions were near floor. Scores in the monotic conditions were significantly greater than would be predicted from an additive combination of the single band conditions (i.e., superadditive effects). Scores in the LF–HF dichotic condition were significantly less than those in the LF–HF monotic condition but were greater than those for either the LF-only or HF-only conditions. Taken together, these findings suggest that listeners are able to integrate information across two auditory channels but that this integration is less than optimal. Implications of these results for theories of sensory integration are discussed.

9:00

4aPPa2. Effect of harmonic structure of environmental noises on vowel perception. Kentaro Ishizuka and Kiyoaki Aikawa (NTT Commun. Sci. Labs., 3-1 Morinosato-Wakamiya, Atsugi City, Kanagawa 243-0198, Japan)

This work reports new findings on the effect of the harmonic structure of environmental noises on vowel perception. An experiment was conducted to examine how listeners perceive natural vowels under noisy environmental conditions at around -2 dB in SNR. Vowels were presented at 60-dB SPL. Eight types of noise were used: white noise and seven types of harmonic structured noises each having the same flat spectral envelope and energy. The fundamental frequency of the vowels was about 260 Hz. A pure tone of 260 Hz and its harmonics (A), a pure tone of 520 Hz and its harmonics (B), and a pure tone of 260 Hz and its odd harmonics (C) were used as noises. Mean identification scores for the white noise and noises (A), (B), and (C) were 33.73%, 59.93%, 66.80%, and 60.27%, respectively. The spectral envelope has been widely used for automatic speech recognition as the feature parameter. However, experimental results demonstrated that perceptual identification scores are significantly different depending on the detailed spectral shape of the noise. The difference between noises (B) and (C) suggests that the even harmonic components of the vowel make less of a contribution to noisy vowel perception than the odd harmonic components.

9:15

4aPPa3. Prediction from pure-tone thresholds of the effects of interaural time delays on the speech reception threshold in a multisource environment. John F. Culling (School of Psych., Cardiff Univ., P.O. Box 901, Cardiff CF10 3YG, UK), Monica L. Hawley (Massachusetts Eye and Ear Infirmary, Boston, MA 02114), Ruth Y. Litovsky, and H. Steven Colburn (Boston Univ., Boston, MA 02215)

Speech reception thresholds (SRTs) in $NoS\pi$ are predictable from the articulation index and pure-tone thresholds; the effective level of masking noise is assumed to decrease by the size of the binaural masking level difference (BMLD) at each frequency [H. Levitt and L. R. Rabiner, J. Acoust. Soc. Am. **42**, 820–829 (1967)]. This assumption was tested for the effect of interaural time delays in multisource environments. Thresholds for diotic pure tones were measured at third-octave intervals for three simultaneous, speech-shaped-noise interferers distributed in perceived azimuth using the temporal components of head-related impulse responses (ITD-only maskers). One masker (444) had all interferers directly ahead (position 4 of 7 in an arc). Pure-tone thresholds for other spatial configurations subtracted from corresponding thresholds in 444 gave BMLDs for each configuration and frequency. Copies of the 444 masker were filtered, attenuating the noise across frequency in accordance with the appropriate BMLDs for each configuration (367, 567, and 777). SRTs for speech from position 4 against each ITD-only masker and its corresponding filtered masker were measured and compared. SRTs matched for distributed configurations (367 and 567), but the filtered masker yielded lower SRTs than the ITD-only masker when all interferers were colocated away from the target speech (777).

9:30

4aPPa4. The effect of spectral contrast enhancement on speech perception in hearing-impaired listeners. Hannes Müsch and Chaslav Pavlovic (Res. and Core Technol. Ctr., GN ReSound Corp., Redwood City, CA 94063, hmuesch@gnresound.com)

The effect of enhancing the spectral contrast of clean and noisy speech on intelligibility and listening comfort in hearing-impaired listeners is explored. Spectral contrast refers to the difference between the amplitude of spectral peaks and valleys in a short temporal window. A method of enhancing spectral contrast similar to that suggested by Bunnell [H. T. Bunnell, J. Acoust. Soc. Am. **88**, 2546–2556 (1990)] was implemented in a simulated hearing aid with fast acting multiband wide-dynamic-range compression. The aid's frequency response was adjusted to accommodate the individual listeners' hearing losses. The effect of several degrees of spectral contrast enhancement on phoneme identification in quiet and in the presence of speech-spectrum-shaped noise was measured. In addition, listeners' preferences for several degrees of spectral contrast enhancement were assessed in an A/B comparison of spectrally enhanced connected speech in quiet and in the presence of cafeteria noise. The experimental results will be reported and their implications for the use of spectral contrast enhancements in hearing aids will be discussed.

9:45

4aPPa5. Speech enhancement based on aspects of the auditory process. Kathryn Arehart and John Hansen (Speech, Lang. and Hearing Sci., Univ. of Colorado, 409 UCB, Boulder, CO 80309, kathryn.arehart@colorado.edu)

The effectiveness of a speech enhancement algorithm on both the intelligibility and quality of speech-in-noise by listeners with cochlear-based hearing loss of moderate severity was tested. The enhancement scheme employed is based on an approach that uses the auditory masked threshold in conjunction with a version of spectral subtraction to adjust the parameters used in the subtraction process based on the masked threshold of the noise across the frequency spectrum. The ability of listeners with hearing loss to perceive stimuli from the nonsense syllable test (NST) and from the diagnostic rhyme test (DRT) in the presence of communication channel noise and automobile highway noise was measured at several different signal-to-noise ratios in both degraded and enhanced conditions. The algorithm effectiveness was also assessed through subjective ratings of speech quality. Results of the listening tests show that some but not all of the listeners with hearing loss benefited from the enhancement algorithm. We will discuss how the algorithm—which is based on models of auditory processing in normal-hearing listeners—might be reformulated to address specific auditory deficits in listeners with cochlear hearing loss. [Work supported by National Organization of Hearing Research, Council on Research and Creative Work, University of Colorado.]

10:00

4aPPa6. A model for predicting speech intelligibility. Hannes Müsch[a] and Søren Buus (Ctr. for Commun. and Digital Signal Processing, Northeastern Univ., Boston, MA 02115, hmuesch@gnresound.com)

A statistical-decision theory based model for predicting speech intelligibility is introduced. This model, which we call the speech-recognition sensitivity (SRS) model, aims at predicting speech-recognition performance from the long-term average speech spectrum, the masking excitation in the listener's ear, the linguistic entropy of the speech material, and the number of response alternatives available to the listener. Unlike articulation-index (AI) models, the SRS model can account for synergetic and redundant interactions among spectral bands of speech. It also accounts for effects of linguistic entropy and number of response alternatives on intelligibility scores without resorting to the empirically determined *ad hoc*transformations employed by AI models. The effects of linguistic entropy are modeled by an entropy-dependent central noise, which modifies the listener's identification sensitivity to the speech. The effect of the number of response alternatives on the test score is a direct consequence of using statistical decision theory. The SRS model also appears to predict how the effect of linguistic entropy varies with the filter condition and how linguistic entropy and language proficiency interact with the signal-to-noise ratio. Fits of the SRS model to data from the literature and consonant-discrimination data collected in our laboratory will be presented. [Work supported by NIH and SigmaXi.] [a]Current affiliation: GN ReSound Corporation.

10:15

4aPPa7. Speech task affects quality of articulation in Parkinsonian dysarthria. Diana Van Lancker, Lisa Tafuro (New York Univ., 719 Broadway, Rm. 200, New York, NY 10003), and Daniel Kempler (LAC and USC Medical Ctr., Los Angeles, CA 90033)

To examine the effect of speech task on articulation, this study assessed speech in a dysarthric patient across four production tasks: spontaneous speech, repetition, reading, and repeated singing, using the same phrases for all tasks. The study design ensured that all phrases were heard in all production tasks, but no listener heard any phrase twice. Data were obtained from listeners' identification and difficulty rating scores, perceptual judgments, and acoustic analyses. Items produced spontaneously were significantly less intelligible than the same phrases produced in the other three tasks (29% vs 79% correct identification). Spontaneous speech was also rated as more difficult to understand than other speech tasks. While relative intensity and word duration were not linked to intelligibility, dysfluencies and distorted articulatory/resonance patterns were found significantly more often in spontaneous than in the other tasks. These indications of variable speech production efficiency were found in a patient with Parkinson's disease, which is associated with basal ganglia impairment. These results can be related to recent models of normal basal ganglia function, which emphasize motor planning and monitoring. One explanation for the different articulation patterns reported here may be the availability of an external ''model'' which aids planning in the reading and repetition tasks.

10:30

4aPPa8. Backward-masking and syllable–rate thresholds have only minor associations with reading achievement, compared to several cognitive and linguistic measures. Gary R. Kidd, Charles S. Watson (Dept. of Speech and Hearing Sci., Indiana Univ., Bloomington, IN 47405), Andrya Lowther (School of Optometry, Indiana Univ., Bloomington, IN 47405), and Eddins A. David (State Univ. of New York–Buffalo, Buffalo, NY 14214)

Rate thresholds for syllable–sequences [the Indiana Test of Auditory Memory and Processing Rate—ITAMPR, Watson and Eddins, J. Acoust. Soc. Am. **105**, 1236 (1999)] and backward-masking thresholds, were obtained from 220 children. These measures were only weakly associated with reading achievement and with speech recognition. Other tests, however, accounted for roughly half of the variance in teacher-assigned reading grades. Results of 36 sensory, perceptual processing, linguistic, and cognitive tests and subtests administered to the population from which these children were obtained (approximately 98% of the 472 entering first-graders in Benton County Indiana from 1995 to 1997), were well-described by four factors. These factors and the systematic variance in reading achievement associated with them were: Factor 1 (31%): reading-related skills (phonological awareness, letter- and word-identification); Factor 2 (12%): visual cognition (visual perceptual abilities, spatial perception, visual memory); Factor 3 (9%): verbal cognition (language development, vocabulary, verbal concepts); and Factor 4 (1%): speech processing (the ability to understand speech, measured under difficult listening conditions). These data are inconsistent with the hypothesis that temporal processing deficits are causally related to reading impairments. [Work supported by NIH/NIDCD, Indiana Univ., and the Benton County (Indiana) School Corporation.]

THURSDAY MORNING, 7 JUNE 2001 RED LACQUER ROOM, 11:00 A.M. TO 12:05 P.M.

Session 4aPPb

Psychological and Physiological Acoustics and Archives and History: History of Psychological Acoustics

William A. Yost, Chair

Parmly Hearing Institute, Loyola University of Chicago, 6525 North Sheridan Road, Chicago, Illinois 60626

Chair's Introduction—11:00

Invited Paper

11:05

4aPPb1. Hearing research and the ASA. Ira J. Hirsh (Central Inst. for the Deaf and Washington Univ., 818 S. Euclid, St. Louis, MO 63110)

In the same year that the ASA was founded, its first President, Harvey Fletcher, had summarized research on *Speech and Hearing* (1929). The knowledge then about auditory sensitivity, discrimination of intensity and frequency, masking, auditory localization and associated binaural effects, loudness and annoyance and speech recognition is very humbling now, 72 years later. The vacuum-tube amplifier, soon to be replaced by the transistor, enabled auditory psychophysics (or psychological acoustics or, more briefly, psychoacoustics) to measure and control sounds, and extend detailed knowledge of hearing. Fletcher's group at Bell Telephone Laboratories consisted of people who were equally comfortable with the engineering aspects of sound making and control, and the psychophysical idea of putting concepts to test with human listeners. From the early 1930's onward, psychoacoustics has contributed much information about auditory processes and theoretical models of application to mathematical, physical and biological systems. The juxtaposition of physical and psychological sciences was just right for the Acoustical Society. Publications, discussions at technical meetings, and opportunities to serve the Society as well as the community at large were enriched by scientists of different training. And that tradition stood well as psychoacoustics reached across further to speech and music perception.

A NOTE ABOUT THE ASA HISTORY LECTURE SERIES

In 1997, the ASA Committee on Archives and History conceived a plan for a series of invited lectures on each of the technical areas of the Society which would memorialize the significant achievements and milestones of each of its technical committees during the first three quarters of the Society's first century.

With the cooperation of the technical committees, distinguished invdividuals are selected to review the history of their particular technical specialty and present a lecture which shows how that activity has developed and has contributed to the Society at large and to the broad field of acoustics as well.

The invited lecturers have been asked to prepare a written manuscript of their lectures which will be published in a commemorative book for the 75th Anniversary of the Society to be celebrated in 2004. The Archives and History Committee and the individual technical committees/group welcome comments and suggestions on both the History Lecture Series and on the proposed ASA Diamond Anniversary Book. Volunteers to assist the committees would be most welcome too. Contact Henry Bass, Chair, Committee on Archives and History, pabass@sunset.backbone.olemiss.edu

Session 4aSA

Structural Acoustics and Vibration and Physical Acoustics: Random Matrix Theory and Complexity I

Richard L. Weaver, Chair

Department of Theoretical and Applied Mechanics, University of Illinois, 216 Talbot Laboratory, 104 South Wright Street, Urbana, Illinois 61801

Chair's Introduction—8:00

Invited Papers

8:05

4aSA1. Universal spectral properties of complex structures and random matrix theory. Olivier Legrand (Laboratoire de Physique de la Matière Condensée, CNRS UMR 6622, Université de Nice-Sophia Antipolis, Parc Valrose, F-06108 Nice Cedex 2, France)

It is now widely recognized that universal properties of spectral fluctuations for waves in complex or disordered cavities can be accounted for by random matrix theory (RMT). These properties are manifested in short- and long-range correlations of the spectral response. In nondissipative systems, the wave evolution operator may be related to a Hermitian matrix. The lack of information about the system due to complexity or disorder is translated in RMT through statistical ensembles of such matrices (e.g., the Gaussian orthogonal ensemble) in which the only relevant features of the system are its modal density and its global symmetries, such as time-reversal invariance. Here, it is shown that, beyond quantum or electromagnetic systems, these concepts are relevant to acoustics and elastodynamics by presenting the few basics of RMT and a simple argument yielding Wigner's surmise for the distribution of frequency spacings between neighboring resonances, exemplifying *level repulsion*. Within the high-frequency geometrical limit of rays, a *semiclassical* argument due to Berry is sketched for the spectral long-range reduction of fluctuations named *spectral rigidity*. This is one of the most convincing arguments in favor of the conjecture which stipulates that wave systems which are chaotic in the limit of rays should follow RMT.

8:35

4aSA2. Ultrasonic resonances in the presence of ray chaos. Clive Ellegaard (Niels Bohr Inst., Blegdamsvej 17, 2100 Copenhagen Ø, Denmark)

The work described here was inspired by quantum chaos and will in large part be discussed in terms of models based on random matrix theory. We study wave mechanics of a system very different from quantum systems, namely ultrasonic resonances of solid objects. We measure the frequency spectra—the eigenvalues—and the standing wave patterns—the eigenfunctions. We are mostly looking at freely vibrating plates. The plates can have regular or chaotic shapes, where regular and chaotic are defined in terms of whether we in the ray limit have chaos or not. By going to the acoustic system we will be testing the predictions of RMT on systems governed by a much more complex wave equation than the Schrödinger equation. In plates the acoustic waves can be either flexural, in-plane compressional, or in-plane shear. These will be mixed through mode conversion at the boundaries and other perturbations so there will in general be complex mixtures in the standing waves. We show different experimental techniques for separating the different modes. We measure transitions from regularity to chaos and show the effects of specific symmetry breaking, some of these unique to the acoustic system.

9:05

4aSA3. Random matrix theory, random point process theory, and natural frequency statistics. Robin S. Langley (Dept. of Eng., Univ. of Cambridge, Trumpington St., Cambridge CB2 1PZ, UK, RSL21@eng.cam.ac.uk)

There has been much recent interest in the application of random matrix theory to the dynamics of uncertain engineering structures. In particular, it has been found experimentally and numerically that the results arising from the Gaussian orthogonal ensemble (GOE) are applicable to natural frequency spacing statistics. The occurrence of natural frequencies can also be viewed as a random point process, and although there is much in common between random point process theory and random matrix theory, the two subjects appear to have been developed independently. It is shown here that the two approaches employ very similar mathematical functions, albeit under a different name: For example, the distribution function (point process theory) corresponds to the n-point correlation function (random matrix theory), and the cumulant function corresponds to the n-level cluster function. By recognizing this similarity, it is possible to apply a number of established results in random point process theory to the statistics of the eigenvalues of a random matrix. This leads to new insights into the statistics of natural frequency spacings, and helps to explain why the Wigner surmise (which states that the spacings have a Rayleigh distribution) is applicable to a much wider class of matrix than the GOE.

9:35

4aSA4. Equipartition and mean square responses in large undamped structures. Richard L. Weaver (Theoretical and Appl. Mech., 104 S. Wright St., Univ. of Illinois, Urbana, IL 61801)

It is shown that the smoothed spectral energy density at one point in a large complex structure may be approximated in a simple manner as proportional to the frequency-smoothed admittance at the receiver and at the source, and inversely proportional to the global modal density. Thus the mean square response may be estimated with little computational burden, knowing only the local properties at source and receiver and the size of the system. The approximate predictions are compared with the results of direct numerical simulations and found to be accurate except in the presence of Anderson localization. [Work supported by NSF.]

10:05–10:20 Break

Contributed Papers

10:20

4aSA5. On the nature of random system matrices in structural dynamics. Sondipon Adhikari and Robin S. Langley (Eng. Dept., Cambridge Univ., Trumpington St., Cambridge CB2 1PZ, UK)

This work is concerned with the nature of random mass and stiffness matrices arising in linear dynamic systems due to inherent uncertainties in the system parameters. In an important recent paper, Soize [Prob. Eng. Mech. **15**, 277–294 (2000)] has developed an expression for the probability density function of such matrices by using the mean value of the matrices in conjunction with the entropy optimization principle. Although mathematically optimal given knowledge of only the mean values of the matrices, it is not entirely clear how well the results obtained will match the statistical properties of a physical system. This issue is investigated here by considering the structure of random matrices arising from Guyan-reduced dynamic models. It is shown that, under rather general conditions, the random parts of the mass and the stiffness matrices of the reduced system in the modal coordinates resemble the so-called ''Gaussian Orthogonal Ensemble (GOE).'' A single parameter σ (the standard deviation of the matrix diagonals) characterizes a GOE, and this offers the possibility of a very straightforward Monte Carlo simulation technique for the system matrices and response. A limited number of numerical results have been shown in favor of the proposed result.

10:35

4aSA6. Spectral power variances for random matrix systems at moderate modal overlap. Igor Rozhkov and Richard L. Weaver (Theoretical and Appl. Mech., Univ. of Illinois, Urbana, IL 61801)

Recent work has established statistical equivalence between the spectra of generic undamped structures and those of typical members of the Gaussian orthogonal ensemble (GOE) of random matrices. This suggests that a kind of universal behavior may obtain for the statistics of the power transmission coefficient in weakly damped generic structures. Unfortunately modal statistics are modified by nonproportional damping, and the GOE is not directly applicable. The power transmission coefficient is the square of elements of the matrix inverse to $[[H]-i[C]-E[I]]$ where $[H]$ is a GOE matrix (effectively the stiffness of the structure), $[C]$ is a (symmetric) damping matrix, E is frequency squared, and $[I]$ is the identity. In the case that $[C]$ is proportional to the identity (or more generally when C is diagonal in the natural basis), power transmission variances can be determined theoretically using the known properties of the GOE. Numerical simulation confirms that theoretical prediction. Further numerical simulation, in the case in which $[C]$ is nonproportional, agrees with laboratory measurements in demonstrating a power variance that is significantly lower than the naive theory predicts. Possible modifications of the theory are discussed, including Efetov's supersymmetry method and Mello's statistical S matrix.

10:50

4aSA7. Numerical evidence for the Wigner surmise. Alastair W. M. Brown and Robin S. Langley (CUED, Trumpington St., Cambridge CB2 1PZ, UK)

The ''Wigner surmise'' states that the probability density function of the eigenvalue spacings of a random system follows the Rayleigh distribution. Thus far the surmise has been proved mathematically only for systems described by the ''Gaussian orthogonal ensemble (GOE)'' of matrices. Random matrices associated with practical engineering structures do not have the form of the GOE, and yet there is strong evidence that the Wigner surmise may still be applicable. The aim of the present work is to determine the conditions under which the natural frequency spacings of an engineering structure will have a Rayleigh distribution. To this end, numerical simulations have been performed for a simply supported flat plate, a simply supported string, and a system composed of two spring-coupled plates. In each case the structure was made uncertain by the addition of randomly placed masses or springs. It was found that the validity of the Wigner surmise depends on three factors: (i) the level of randomness, (ii) the degree of symmetry, and (iii) for the coupled system, on the strength of coupling between the two subsystems. The relative importance of these factors has been quantified for the examples considered.

11:05

4aSA8. Level repulsion and spectral rigidity in reverberant microwave cavities. Jérôme Barthélemy, Olivier Legrand, and Fabrice Mortessagne (Laboratoire de Physique de la Matière Condensée, CNRS UMR 6622, Université de Nice-Sophia Antipolis, Parc Valrose, F-06108 Nice Cedex 2, France)

It is now widely acknowledged that spectral fluctuations of waves in complex systems are well described by random matrix theory. As a model system for *complex* or *disordered reverberant cavities*, a two-dimensional reverberant microwave cavity is investigated in a high-frequency regime where resonances are still below the large modal overlap regime. Modifying the cavity through the inclusion of scatterers makes the dynamics of rays chaotic or diffusive, thus rendering its spectral response strongly correlated over short and long frequency range. This can be illustrated by the distribution of frequency spacings between neighboring resonances which clearly exhibits the so-called *level repulsion*. Longer range correlations are demonstrated by considering an averaged time response of the cavity for times shorter than the inverse modal density. By scaling down the hyperfrequency range to audio frequency range and proper sampling, an auditive perception of the differences between the spectral responses of a regular (pure rectangular cavity) and a complex reverberant cavity is demonstrated.

11:20

4aSA9. Intensity maps of systems with complex eigenvalues. Xianhui Li and J. Gregory McDaniel (Dept. of Aerosp. and Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215)

Maps of structural and acoustic intensities have been widely used to understand the dynamics of complex systems with time-harmonic forcing. Identification of the dominant power paths helps designers introduce modifications that control them. For example, damping treatments can be placed to dissipate the maximum power and effectively reduce unwanted vibration. Similarly, high-impedance elements can be configured to block the power flow to sensitive system components. In typical applications, the applied force distributions are chosen to be representative of the expected in situ forces. In this presentation, an extension is described that allows one to construct an intensity map for one mode of a system with complex eigenvalues. This work is motivated by the need to map structural intensity for a squealing mode of an automotive brake system, but it would also be useful in understanding modes of highly damped systems. The proposed extension isolates such modes by choosing a time-harmonic force vector that is parallel to the complex-valued eigenvector, in which case a single mode is isolated. Numerical examples will illustrate the features of this map and contrast it to conventional maps. [Work supported by NSF.]

11:35

4aSA10. Electromechanical realization of impedance matrices. Pierre E. Dupont and Wenyuan Chen (Aerosp. and Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215)

Applications exist at both small and large length scales for which one wishes to design an electromechanical system that reproduces a desired impedance matrix. Such a matrix corresponds to one or more degrees of freedom at several attachment points. The design of MEMS filters for signal processing applications is a small length scale example. At a much larger scale, simplified models of dynamically complex machinery are useful for testing the vibration isolation properties of supporting structures. In the first example, the desired impedance matrix is usually defined mathematically while in the latter, a mobility matrix is obtained by experiment. The realization problem is to obtain an electromechanical system that approximates the desired impedance matrix. Constrained approximation is necessary since the literal realization of idealized filters and complex machinery is often precluded by modal complexity, affecting cost and difficulty of construction, and implementation issues, such as nonideal boundary conditions and the lack of ''sky hooks.'' A two-stage solution to the realization problem is presented that divides the impedance matrix into passive and active components. In the first stage, a reduced-order passive mechanical model is obtained. This model is then modified to incorporate the active behavior. [Work supported by ONR.]

THURSDAY MORNING, 7 JUNE 2001 MONROE ROOM, 8:30 A.M. TO 12:00 NOON

Session 4aSC

Speech Communication: Speech Production Potpourri (Poster Session)

Susan Shaiman, Chair

Department of Communication Science and Disorders, University of Pittsburgh, 4033 Forbes Tower, Pittsburgh, Pennsylvania 15260

Contributed Papers

All posters will be on display from 8:30 a.m. to 12:00 noon. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 8:30 a.m. to 10:15 a.m. and contributors of even-numbered papers will be at their posters from 10:15 a.m. to 12:00 noon.

4aSC1. English stress timing and dynamical control of speech timing. Robert Port, David Collins, Adam Leary, Deborah Burleson (Dept. of Linguist., Indiana Univ., Bloomington, IN 47405), and Mafuyu Kitahara (NTT Labs., Atsugi, Japan)

Studies of English speech timing have concluded that, despite earlier claims, English is neither stress-timed nor syllable-timed. This study directly measured context effects on timing depending on prosodic neighborhood, and compared the contribution of any stress-timing or syllable-timing tendency while speech cycling. Ss repeated a phrase to a metronome at varying rates. This complex metronome included high-pitched tones ('beeps') at roughly 1 s with a lower tone ('boop') occurring either 1/2 (2-beat) or 2/3 (3-beat) of the way between beeps. We hypothesize dynamical attractors for stressed syllables at simple harmonic fractions (1/2, 1/3, 2/3) of the phrase cycle. Various texts invite readings with two stresses ('Bite the back') or three stresses ('Bite Bill's back'), and some had perturbing unstressed syllables inserted between stressed ones ('Biting Bill's back'). Perturbing and target vowel onset times were fit with a multiple linear regression equation using time predictions assuming equal stress intervals or equal syllable intervals. The estimated slopes indicate the relative weight of stress versus syllable timing. The weights were greater for stress timing for all subjects, supporting a model of global speech timing where stressed syllable locations have attractors at harmonic fractions, as might be found in a bank of coupled oscillators.

4aSC2. Speech synthesis by mapping articulator movement patterns to a shape-based area function model of the vocal tract. Brad H. Story (Speech and Hearing Sci., Univ. of Arizona, P.O. Box 210071, Tucson, AZ 85721-0071, bstory@u.arizona.edu)

Orthogonal components of tongue and lip displacement data from two speakers in the Univ. of Wisconsin x-ray microbeam database (Westbury, UW–Madison, 1994) were determined with a principal components analysis (PCA). The PCA was performed on steady-state vowels and at the sentence level. For every time frame, the analysis consisted of fitting a cubic spline to four pellet points on the tongue and to modified positions for the incisor pellet and the upper and lower lip pellets. The modifications brought the pellet positions to the level of the airway. The PCA was performed over a collection of frames and showed that the first two (most significant in terms of variance) orthogonal components were similar in shape to those determined from a set of MRI-based vocal tract area functions [Story and Titze, J. Phonetics **26**, 223–260 (1998)]. Next, the coefficients determined from a sentence-level PCA of the microbeam data

were normalized and used as input to an area function model based also on orthogonal shaping components. The result is that articulatory movement patterns from data in the x-ray microbeam database can be transformed to movement patterns of a vocal tract area function to reproduce (synthesize) the original sentence. [Work supported by NIH R01-DC04789-01.]

4aSC3. Sentence-final lengthening in connected text. Caroline L. Smith and Lisa Hogan (Dept. of Linguist., Univ. of New Mexico, Albuquerque, NM 87131-1196, caroline@unm.edu)

Lengthening at the end of a prosodic domain is a robust pattern in English, with larger domains characterized by greater amounts of lengthening. Here we investigate whether this pattern extends above the sentence. Is there more sentence-final lengthening at the end of a paragraph than at the end of a paragraph-internal sentence? Ten recordings were made at weekly intervals of an American English speaker reading extracts from a computer manual, which was chosen because such writing has well-defined topic structure. At the same recording sessions, the speaker also read isolated sentences which were constructed so that each word that was sentence final in the connected text occurred sentence medially in one of the sentences, which were designed to match the presence or absence of a pitch accent on the word in the text. Sentence-final lengthening in the text was gauged by comparing a word's duration in the text to its mean duration across the ten repetitions of the isolated sentences. Very preliminary results suggest that there is more lengthening at the end of a paragraph than in the middle of a paragraph; headings may behave like part of the following paragraph. [Work supported by NSF.]

4aSC4. The influence of speaking rate on spirantization in motor speech disorders. Greg S. Turner (Dept. of Commun. Disord., Central Missouri State Univ., Martin 64, Warrensburg, MO 64093, turner@cmsul.cmsu1.edu)

Spirantization occurs often in the speech of individuals with dysarthria. Impairment of muscle movement can lead to an incomplete articulatory obstruction of the vocal tract during the closure period for the production of stop consonants. Air passes through the narrow constriction associated with the incomplete constriction, resulting in the production of fricative noise during the stop gap. This noise leads to the perception of imprecise articulation. It is hypothesized that the speaking rate at which an individual with dysarthria talks will influence the frequency of spirantization. The slower the speaking rate the more time is allowed for the speaker with dysarthria to achieve a tight constriction during the closure period for stop production, resulting in a reduction of spirantization. The opposite would occur for faster rates. For this experiment, individuals with dysarthria read the Farm Passage at three different speaking rates (hab, slow & fast). From this passage, the production of initial voiceless stop consonants (/p, t & k/) were analyzed for the presence of frication during the closure period. Statistical comparisons were made across speaking rate and place of production. group and individual results will be presented.

4aSC5. Variant frequency in American English flap production. David Patterson and Cynthia M. Connine (Psych. Dept., State Univ. of New York at Binghamton, Binghamton, NY 13902)

This study examined the dominance of flaps as the standard pronunciation of medial /t/ in American English compared to other allophonic variants, namely [t] or glottal stops. A large conversation speech database was utilized to generate statistics about the frequency of occurrence of medial flaps. Results confirmed the dominance of flapping in American English. Interesting effects of morphological complexity and lexical frequency were found to have an effect on the allophonic variant distribution patterns. Morphologically complex words (e.g., waiting) showed a redistribution of variants from flaps to [t] compared to dominant flapping in morphologically simple words (e.g., quota). A similar redistribution of variant production occurred for low-frequency words compared with high-frequency words. The pattern of results was maintained when the morphological analyses were conducted separately for low- and high-frequency words. A second analysis examining vowel length prior to medial /t/ an medial /d/ showed that the vowel preceding medial /d/ tended to be longer, even though both stops were pronounced as a flap. The relevance of the variability in flap occurrence was discussed in relation to rule based phonology and in reference to potential processes in recognizing spoken words.

4aSC6. The relationship between production of suprasegmental speech characteristics and lung volumes in conversational speech. Kate Bunton and Jill Petska (Inst. for Neurogenic Commun. Disord., Univ. of Arizona, P.O. Box 210071, Tucson, AZ 85721, bunton@u.arizona.edu)

''Monopitch'' and ''monoloudness'' are frequent descriptors of speech produced by people with dysarthria, these terms correspond to acoustic findings of flattened $F0$ contours, limited rms variability, and short phrases. Underlying reasons for these production characteristics, however, have not been investigated. The present study was designed to look at the relationship between use of lung volumes in conversational speech and the production of suprasegmental characteristics in healthy speakers versus those with Parkinson disease. Measures of the suprasegmental components of the speech signal included, $F0$ and rms variability, breath group duration, and number of syllables. Measures of starting and stopping lung volumes for each of the breath groups were also obtained. Results indicate a tendency for subjects to alternate a longer breath group, started at a higher lung volume, with exaggerated prosodic characteristics with a much shorter breath group. Although this pattern was seen in the disorder group, the length of their breath group, and $F0$ and rms variability were limited. The disorder group also tended to initiate utterances at lower lung volumes and continue speaking past resting expiratory level. [Work supported by NIDCD DC-01409.]

4aSC7. Coproduction in VCV disyllables produced by children and adults: A comparison of anticipatory and carryover effects. Carole Gelfer (Dept. of Commun. Disord., William Paterson Univ., Wayne, NJ 07470, gelferc@wpunj.edu) and Fredericka Bell-Berti (St. John's Univ., Jamaica, NY 11439)

Previously, we reported spectral data showing similar patterns of coproduction for children and adults in schwa preceding a stop + vowel [C. E. Gelfer and F. Bell-Berti, J. Acoust. Soc. Am. **107**, 2855 (2000)]. The absence of a clear developmental pattern differs from some previous reports of children's speech gestures [e.g., Nittrouer *et al.*, J. Speech Hear. Res. **39**, 379–389 (1996)]. For both groups in our study, the alveolar place blocked the effects of the following vowel, while vowel effects were observed in the context of the bilabial and velar stops. These results also suggested that the gestures associated with a segment are an integrated unit: The lip gesture associated with the vowel seems to be delayed until the lingual gesture associated with the alveolar closure occurs. This study continues our investigation of these patterns and compares the productions of children 4 to 8 years old to those of adults. Here we study carryover as well as anticipatory effects, using VC+schwa and schwa+CV utterances, where V is /i/ or /u/ and C is /p,b,t,d,k,g/. $F2$ measurements taken at various intervals during schwa will be used as an index of these effects and the presence or absence of developmental trends.

4aSC8. Extracting tongue muscle contraction patterns from tagged cine MRI. Maureen Stone (Univ. of Maryland Med. School, 16 S. Eutaw St., Rm. 500, Baltimore, MD 21201), Danielle Dick, Andrew S. Douglas, Guy Shechter (Johns Hopkins Univ., Baltimore, MD 21218), Cengizhan Ozturk (Bogazici Univ., Istanbul, Turkey), and Michael Guttman (Natl. Inst. of Health, NHLBI, Bethesda, MD 20892)

This study presents mechanically modeled 3-D volumetric strains for the tongue during speech for the syllable ''sha.'' Multiplanar tagged cine MRI (tMRI) provided input data for a B-spline, geometry-independent, cardiac tag tracking method, devised for the heart, which has been adapted for the tongue. Three sets of tMRI images with orthogonal tag planes were collected in 10 axial and 5 sagittal slices for 24 consecutive time phases. The sagittal slices were recorded twice, once each with a series of horizontal and vertical tag planes. The axial slices were recorded once with lengthwise (anterior-to-posterior) tag planes. These tag planes reflect deformations in the SI, AP, and RL directions, respectively. Within the tongue we tracked 3-D motion, calculated 3-D strains in each image plane, and reconstructed 3-D deformation for the entire volume. From the model muscle contraction patterns are inferred for genioglossus anterior, verticalis, and transverse. To infer muscle contraction we determined the lines of action for each muscle, and tracked their linear strains in the appropriate planes for all 24 time phases. Results showed that muscle compressions are consistent with expected muscle contractions, and we were able to distinguish between transverse and verticalis activity. [Work supported by NIH/NIDCD Grant No. DC01758.]

4aSC9. Stability of intra- and inter-articulatory timing for contiguous jaw cycles. Susan Shaiman (Dept. of Commun. Sci. and Disord., Univ. of Pittsburgh, 4033 Forbes Tower, Pittsburgh, PA 15260, shaiman@shrs.pitt.edu)

The current study compared how contiguous cycles of jaw movement, composed of different phonetic contexts, were impacted by global changes across the entire utterance and by more local, phonetic changes. Specifically, articulatory kinematics were examined to determine if patterns of intra- and inter-articulatory timing across manipulations of speaking rate were maintained for both jaw cycles, while being reorganized for phonetic changes (i.e., in coda composition) specific to the second jaw cycle. Five normal speakers repeated the syllables /paep/, /paeps/, and /paepst/, embedded in the carrier phrase, ''Now say —- again,'' using slow, normal, and fast speaking rates. Changes in speaking rate impacted the utterance globally, with both jaw cycles evidencing similar patterns of kinematic change. Conversely, changes in coda composition for the second jaw cycle resulted in local changes to that cycle only, without impacting the first cycle. Individual speakers demonstrated distinct, but systematic patterns of intra- and interarticulatory timing as a function of coda composition and speaking rate. While functional groupings of articulators may be broken apart and reconfigured for subsequent movements, global suprasegmental changes result in timing patterns which are consistent across these reorganized functional groupings. [Work supported by CRDF–University of Pittsburgh and NSERC.]

4aSC10. Schwas with and without active gestural control. Iris Smorodinsky (Dept. of French and Italian, Univ. of California, Santa Barbara, CA 93106; Yale Univ.; and Haskins Labs., smorodin@humanitas.ucsb.edu)

This study investigates whether some epenthetic vowels are targetless, that is, whether they can arise from the timing of the surrounding consonants [C. P. Browman and L. Goldstein, Papers in laboratory phonology II: Gesture, segment, prosody, 26–56 (1992)]. Specifically, the difference in targetlessness between past tense and lexical schwas in American English is examined. Articulatory data were collected from three speakers of American English using an electromagnetic midsagittal articulometer. The stimuli included phrases with past tense and lexical schwas embedded in a common environment: for example, ''If cheated even once'' (past tense schwa) and ''If Cheetah'd even known'' (lexical schwa). If the tongue body is assumed to be controlled continuously by the targetful vowels, the tongue body position ''during a schwa'' should not differ significantly from the tongue body position during the preceding vowel in the epenthetic schwa tokens while it would in the lexical schwa tokens. In addition, if the lexical schwas have tongue body gestures associated with them but the past tense schwas do not, a significant interaction between schwa type and vowel context is expected. These results would provide evidence of a difference in targetlessness between the past tense and lexical schwas in American English. [Work supported by NIH.]

4aSC11. The phonetics of stress clash in spontaneous speech. Lesley M. Carmichael (Dept. of Linguist., P.O. Box 354340, Univ. of Washington, Seattle, WA 98195-4340, lesley@u.washington.edu)

This study examines the rhythm rule (RR) in spontaneous speech to determine its acoustic correlates and phonetic robustness in conversational intonation. RR in English is a phonological process that accommodates stress clash. It has been primarily investigated in controlled speech. Both analyses of RR were considered in this study: (1) accent reversal—the relative prominence of key syllables is reversed, (2) accent deletion—the relative prominence of key syllables is neutralized. Two hypotheses were investigated: (1) RR is always indicated by measurable acoustic properties. (2) RR is subject to intonational factors and depends upon the attraction of a pitch accent for realization. Duration, f0, and amplitude were measured on relevant syllables. ToBI labeling was used to indicate pitch accents and phrase boundaries. Duration was the most consistent acoustic cue to stress clash resolution. When the target phrase attracted a pitch accent, duration indicated stress clash resolution by accent reversal. When the target phrase did not attract a pitch accent, f0 and duration indicated accent deletion. Crucially, when the phrases did not attract pitch accents (were not intonationally prominent), stress clash was resolved. This finding provides support for hypothesis (1): RR is pervasive in the acoustic signal independently of intonational factors.

4aSC12. Vocal tract length development: MRI procedures. Houri K. Vorperian (Waisman Ctr., Univ. of Wisconsin–Madison, 1500 Highland Ave., Madison, WI 53705), Cliff M. Kalina, Ray D. Kent, Brian S. Yandell, and Lindell R. Gentry (Univ. of Wisconsin–Madison, Madison, WI)

From infancy to adulthood, vocal tract length increases by about twofold. The purpose of this study is to assess the developmental changes in the various hard and soft tissue structures in the vicinity of the vocal tract that contribute toward its length. Magnetic resonance images (MRI) from children (birth to 6 years) and adults were used since MRI provides detailed visualization of the soft tissues in the oral and pharyngeal regions along with adequate visualizations of related bony and cartilaginous structures. Using previously established measurement procedures (Vorperian *et al.*, 1999), the following structures were measured: lip thickness, hard and soft palate length, tongue length, oro- and naso-pharyngeal length, mandibular length, and position of the hyoid bone and larynx in relation to the nasal spine. Findings will be discussed in terms of: (a) the relative contribution of the various structures toward vocal tract length and how the extent of contribution changes with age; and (b) the relative and relational growth of the different structures. Findings provide normative data on the various vocal tract structures measured; also, they contribute toward understanding the anatomic changes that may be a substrate to speech emergence and development. [Work supported by NIDCD Grant No. R03-DC 4362.]

4aSC13. Articulatory and acoustic relations during variations in speech rate and loudness. Jordan R. Green and Gary Weismer (Dept. of Communicative Disord., Univ. of Wisconsin–Madison, 1975 Willow Dr., Madison, WI 53706, jordangreen@facstaff.wisc.edu)

Rate and loudness manipulations are standard treatment strategies used to recover speech intelligibility. Although numerous studies have characterized the articulatory changes that result from these manipulations, their influence on segmental aspects of the acoustic signal are not well understood. This preliminary investigation examines acoustic, electromyographic (EMG), and kinematic changes associated with elicited modulations of rate and loudness in typically speaking individuals. Participants were instructed to repeat a sentence five times while incrementally increasing or decreasing either their speaking rate or loudness with each repetition. Acoustic measures included total utterance durations, segment durations, vowel formant frequencies, and trajectories. Displacements of the upper lip, lower lip, and jaw were captured in three dimensions using a computerized video movement-tracking system. Surface EMG signals were recorded from the right masseter, right temporalis, and submental group. Data analyses were directed toward describing the degree of correspondence among parametric changes in muscle activity, articulatory kinematics, and segmental features of the acoustic signal that result from nearly continuous changes of rate and loudness. [Work supported by NIH-NIDCD.]

4aSC14. Temporal properties of electromyographic patterns in children's and adults' speech. Mary K. Kenney (Laryngeal and Speech Section, NINDS, Bldg. 10, Rm. 5D38, 10 Center Dr., MSC 1416, Bethesda, MD 20892-1416), Bruce L. Smith, and Charles Larson (Northwestern Univ., Evanston, IL 60208)

Electromyographic patterns of children's and adults' speech production were analyzed to examine temporal relationships among pairs of muscles commonly viewed as functioning either as ''agonists'' or ''antagonists.'' It was hypothesized that adults' greater degree of speech motor control would be evidenced by stronger positive correlations between agonist muscle pairs (e.g., obicularis oris inferior and obicularis oris superior or anterior belly of digastric and depressor labii inferior). In addition, it was expected that greater negative correlations would occur for adults between antagonist pairs (e.g., obicularis oris inferior and depressor labii inferior). A group of adults and three groups of children (5-, 8-, and 11-years old; eight subjects/group) produced multiple repetitions of the phrase, ''Say pap again.'' EMG activity was measured relative to the two labial consonants in the acoustic waveform, and cross correlations were computed between various muscle pairs. In general, the predicted results were obtained, although negative correlations between antagonist pairs often were not observed. Significant, age-related differences occurred primarily between the adults and the 5-years old, and occasionally between the adults and the 8-years old. The findings suggest that adults tend to have more discrete separation between agonist and antagonist muscle groups in speech production than young children.

4aSC15. Developmental changes in the production of obstruents. I. An acoustic analysis of voiceless obstruents. Shawn L. Nissen, Elaina Frieda, Julie McGory, Robert A. Fox (Dept. of Speech and Hearing Sci., The Ohio State Univ., Columbus, OH 43210-1002), Kelly Friedman, and Kim Rosenbauer (The Ohio State Univ., Columbus, OH 43210-1002)

This is one of two papers measuring the acoustic structure of prevocalic obstruents in children and adults, examining age-related differences. This study looks at voiceless consonants. The data will be drawn in part from a large scale acoustic database developed in collaboration with ATR and include word initial voiceless obstruents from real words (in a range of phonetic contexts) as well as a series of syllable initial intervocalic obstruents drawn from nonwords (in the constrained context [a _a]). Of particular interest are the intervocalic consonants, as there is a relatively limited amount of knowledge on the acoustic nature of obstruents produced by children in this position. Utterances elicited from speakers ranging in age from 3–70 years were recorded online in a quiet room environment using high quality microphones and direct analog-to-digital conversion to computer disk. Acoustic measurements collected include the spectral moments [K. Forrest *et al.*, J. Acoust. Soc. Am. **84**, 115–123 (1988)], formant transitions, VOT, closure duration (for stops), noise duration (for fricatives), and preceding vowel duration. Acoustic differences obtained will be discussed as a function of age group, gender, and phonetic context. [Work supported by an INRS Award from NIH and research funding from ATR (Fox, P.I.).]

4aSC16. Developmental changes in the production of obstruents. II. An acoustic analysis of voiced obstruents. Elaina Frieda, Shawn L. Nissen, Julie McGory, Robert A. Fox (Dept. of Speech and Hearing Sci., The Ohio State Univ., Columbus, OH 43210-1002), Kim Rosenbauer, and Kelly Friedman (The Ohio State Univ., Columbus, OH 43210-1002)

This is one of two papers measuring the acoustic structure of prevocalic obstruents in children and adults, examining age-related differences. This study looks at voiced consonants. The data will be drawn in part from a large scale acoustic database developed in collaboration with ATR and include word initial voiceless obstruents from real words (in a range of phonetic contexts) as well as a series of syllable initial intervocalic obstruents drawn from nonwords (in the constrained context [a _a]). Of particular interest are the intervocalic consonants, as there is a relatively limited amount of knowledge on the acoustic nature of obstruents produced by children in this position. Utterances elicited from speakers ranging in age from 3–70 years were recorded online in a quiet room environment using high quality microphones and direct analog-to-digital conversion to computer disk. Acoustic measurements collected include the spectral moments, formant transitions, VOT, percentage of voicing during consonant closure or approximation, closure duration (for stops), noise duration (for fricatives), and preceding vowel duration. Acoustic differences obtained will be discussed as a function of age group, gender, and phonetic context. [Work supported by an INRS Award from NIH and research funding from ATR (Fox, P.I.).]

4aSC17. Eliciting reduced, citation, and hyperarticulated speech in the laboratory: Acoustic and perceptual analyses. James D. Harnsberger, David B. Pisoni (Speech Res. Lab., Dept. of Psych., Indiana Univ., Bloomington, IN 47405, jharnsbe@indiana.edu), and Richard Wright (Univ. of Washington, Seattle, WA 98195)

A method was developed to elicit sentences in three speech styles (reduced, citation, hyperarticulated) using controlled materials. The reduced style was elicited by having subjects read sentences while, as a distractor task, recalling a digit sequence from short-term memory. In prior work, the digit sequence length was calibrated to the individual subject's digit span. Using this procedure, a reduced style was elicited from half of the subjects. In this study, the sequence length was set to one additional digit beyond the individual subject's digit span, in the hope that the more difficult distractor task would consistently elicit a reduced style. The citation style was elicited by having subjects read sentences from a list. The hyperarticulated style was elicited by prompting subjects to reread some sentences again more carefully. Ten subjects were recorded in this experiment. The resulting sentences were acoustically analyzed in terms of sentence duration, keyword duration, and vowel space size. Sentences differing only in style were also presented in a paired comparison

test to 35 listeners. The results of both the acoustic analysis and the perception test showed the method was successful in consistently eliciting reduced sentences from 8 subjects, and citation and hyperarticulated sentences from all 10 subjects.

4aSC18. Automated head and transducer support (AHATS) system for ultrasound data collection. Andrew Jon Lundberg, Maureen Stone (Univ. of Maryland School of Medicine, 16 S. Eutaw St., Ste. 500, Baltimore, MD 21201), and Edward P. Davis (Paratek Microwave, Inc., Columbia, MD 21045)

Ultrasound has been shown to be a useful tool for measuring the vocal tract during speech. This is largely because ultrasound is noninvasive, safe, relatively inexpensive, and can capture 2-D images at video frame rates. In previous work, a head and transducer support (HATS) system was developed which allowed stable positioning of the subject and the ultrasound transducer, and led to improved measurement accuracy. We have now automated the system (AHATS), replacing the manually positioned transducer holder with a computer controlled robotic arm that supports automatic positioning and logging of the transducer holder. Computer controlled positioning of the data collection device increases the speed at which we can collect ultrasound data sets with multiple image planes. It also increases our confidence in the accuracy of combined 3-D data sets (composed of multiple 2-D data collections within the 3-D space), and facilitates efficient 3-D reconstructions which require data collection positions to be optimized for individual subjects. [Supported by NIH-NIDCD DC01758 and Northrop Grumman, Inc.]

4aSC19. On the error distributions of speech production data. H. Betty Kollia (Dept. of Commun. Disord., William Paterson Univ., 300 Pompton Rd., Wayne, NJ 07470) and Jay Jorgenson (CCNY, New York, NY 10031 and CUNY Grad. Ctr.)

In previous works [J. Acoust. Soc. Am. **102**, 3164–3165(A) (1997) and **105**, 2(A), 1354(A) (1999)] the following problem was presented. Assuming that two-dimensional data (such as that obtained in most kinematic speech studies) are approximately linear with error terms about the mean coming from a quadratic—rather than a normal—distribution, then the regression analysis is best performed using a maximum likelihood stochastic linear regression methodology, instead of a least squares methodology. It was shown that one could expect, among other results, sharper confidence intervals about the mean. It remains to be determined if the distribution of error terms about the mean (for speech production data) is better described by a quadratic than by a normal distribution. Indeed, using the classical Kolmogorov–Smirnov test, it is shown that in many cases one can assume that the error terms follow a quadratic distribution and that at any reasonable confidence level one can reject the hypothesis that error terms follow a normal distribution. The result of this test substantiates the theoretical premise assumed in the earlier work.

4aSC20. Dialectal differences in vowel articulations in pre-consonantal contexts. Alice Turk (Univ. of Edinburgh, Edinburgh, UK), Jim Scobbie, and Nigel Hewlett (Queen Margaret Univ. College, Edinburgh, UK)

Many varieties of English show the well-known pattern of longer vowels before voiced consonants, at least in monosyllabic words. Scottish English, however, shows a different pattern of adjacent-consonant conditioned vowel duration for the vowels /i,u,ai/: In this variety, short vowels are found before tautomorphemic voiceless stops, voiced oral and nasal stops, and voiceless fricatives. Long vowels are found before voiced fricatives and morpheme boundaries. For example, ''bead,'' ''bean,'' and ''beat'' and ''peace'' all contain short vowels, while ''please,'' ''free,'' and ''freed'' contain long vowels. This contextually conditioned durational pattern has been described by the Scottish vowel length rule (SVLR). We conducted an EMA study of the SVLR to examine the articulatory strategies speakers use to produce these well-documented acoustic durational patterns. We are particularly interested in (1) whether there is a single mechanism underlying all long duration vowels, i.e., those conditioned by tautomorphemic /z/ as compared to tautosyllabic but heteromorphemic /z/ and /d/, and (2) whether the strategy Scottish English speakers use to lengthen vowels in the SVLR conditioning environments is the same as that used by speakers of other varieties to produce longer vowels before voiced obstruents. [Work supported by ESRC.]

4aSC21. Lingual awareness and the description of isolated syllables of English: A distinctive features analysis. Patricia Lohman (Dept. of Speech and Hear. Sci., Ohio Univ., 218 Lindley Hall, Athens, OH 45701, p1206190@ohio.edu) and Donald Fucci (Ohio Univ., Athens, OH 45701)

The purpose of this investigation was to determine whether normal adult speakers of English could accurately described tongue position within the oral cavity during the production of isolated English syllables. The phonemes were analyzed according to their distinctive features. The effects of training on this task were also examined. Sixty students, 30 majors in speech-language pathology (M age=22.7) and 30 non-majors (M age=18.8) participated in this study. The procedure involved asking participants to (a) imitate the production of a syllable, (b) respond to the question, ''Where was your tongue?'' and (c) answer four multiple-choice questions regarding tongue height, position, and contact with other structures within the oral cavity. Results indicated that the trained group had higher test scores on all phonemes ($p<0.05$) with the exception of **r** and **sh** ($p=>0.05$). The Trained group had the lowest scores on the phonemes with the greatest number of features (**l**=6 features; **r**,**sh**=5 features). The Untrained group had the most difficulty describing the phonemes **k** (3 features), **s** (5 features), and **l** (6 features). The results indicated **training**, as well as **type** and **number** of features significantly affected the task.

4aSC22. Fundamental frequency change as a result of repeated corrections. Caroline M. Menezes and Osamu Fujimura (Dept. of Speech & Hearing, The Ohio State Univ., Columbus, OH 43201-1002)

Articulatory and intonational patterns were studied, using a simulated discourse-like paradigm, in which subjects were asked to make repeated corrections of a three-digit sequence street address consisting of fives and nines and Pine. Articulatory and acoustic signals were collected using the x-ray microbeam system for four native American English speakers at the University of Wisconsin. Based on the C/D model, a linear pulse train representing the rhythmic organization of an utterance can be inferred by analyzing jaw movements as syllable magnitude over time. Fundamental frequency contours were examined to study the effects of phrasing as a result of contrastive emphasis (digit corrected is assumed to have contrastive emphasis) and repeated corrections in the three-digit sequence. These patterns were then studied in relation to the jaw movement patterns. The data show that subjects change speech strategies when they need to make a correction or repeated corrections [Erickson *et al.* (1992)]. The word in particular that is corrected is made prominent from the rest of the utterance, in all repeated corrections. Effects of speaker variability will also be discussed. [This work has been supported in part by NSF (BCS-9977018) and ATR/MIC, Japan.]

Session 4aUW

Underwater Acoustics and Signal Processing in Acoustics: Underwater Acoustic Communications I: Algorithm Design and Analysis for Realistic Propagation Conditions

James C. Preisig, Chair

Department of Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution, M/S 11, Bigelow 210, Woods Hole, Massachusetts 02543-1053

Chair's Introduction—8:00

Invited Papers

8:05

4aUW1. Channel characterization for high-frequency acoustic communications. Michael B. Porter (Sci. Applications Intl. Corp., 888 Prospect St., Ste. 201, La Jolla, CA 92037), Vincent K. McDonald, Paul A. Baxley, and Joseph A. Rice (Space and Naval Warfare Systems Ctr., San Diego, CA 92152-5001)

Undersea internets or seawebs are currently being developed and tested for a variety of applications in which some kind of device (e.g., an ADCP, hydrophone, vertical line array, autonomous undersea vehicle) needs to pass information to another (e.g., radio buoy, surface ship). The underlying physical layer is often an acoustic modem. However, the ocean medium, though it can carry acoustic waves with remarkable fidelity to long distances, is also somewhat unreliable. Storms may cause network outages by both driving up the ambient noise and corrupting the clarity of the acoustic mirror formed by the air–sea interface. Meanwhile, as the SONAR community is well aware, the refractive effects of the ocean can lead to sweet spots and dead zones (shadows). The required sound output of a modem varies accordingly. We have been conducting a program (SignalEx) to: (1) better understand propagation in the high-frequency range (around 10 kHz), currently of greatest interest for modems; (2) understand how the environment degrades different modulation schemes; and (3) develop more robust schemes or select the appropriate scheme for the given conditions. This talk will summarize results from recent SignalEx sea tests.

8:25

4aUW2. Characterization of very shallow water time-varying acoustic channels. David Brady (ECE Dept., Northeastern Univ., Boston, MA 02115)

An accurate and tractable model for communication channel dynamics is important for the design of adaptive transceivers. This channel model should account for hardware-related dynamics (oscillator drift) as well as environmental effects (platform and scatterer motion). Typically, the parameters of a channel model are estimated using a time series of noisy channel outputs, due to a known time-series excitation. Additionally, when ancillary environmental measurements are available (wind, current, water-column history), the channel parameter estimators should also incorporate this information. An investigation of channel dependence on environmental measurements is the main focus of this talk. In this work the environmental dependence of a multipath acoustic communication channel on environmental measurements is investigated for a very shallow water environment. Conjectures about this dependence are confirmed by a data from multiday experiment in the New England area. Observations for this experiment include noisy channel outputs, wind, current, and water-column measurements.

Contributed Papers

8:45

4aUW3. Environmental factors that affect acoustic propagation and underwater communications in the surf zone. Grant B. Deane (Scripps Inst. of Oceanogr., La Jolla, CA 92093-0238, grant@mpl.ucsd.edu)

Acoustic signals propagating through the surf zone are strongly influenced by scattering from surface gravity waves, reflection from the seafloor and absorption and scattering by the clouds of bubbles entrained by breaking surf. Shoaling gravity waves create focal regions in surface-reflected wavefronts, and introduce systematic variations in wavefront travel times. Both of these surface effects are important to the design and operation of very shallow water communication systems. To study them, an experiment was conducted 40 m north of Scripps Pier in November of 2000. Single cycle acoustic pulses centered at 10 kHz and phase modulated pseudorandom sequences were repetitively transmitted over a 40 m path to a 3-element vertical hydrophone array in 5 m deep water. Environmental characterization was provided by a 10-element pressure array mounted along the propagation path to measure the shoaling gravity wave field. The results of this experiment and additional experimental data collected at Scripps Pier during the Fall of 2000 over several regimes of environmental conditions (the SZATE experiment) will be discussed. [Work supported by ONR.]

9:00

4aUW4. Modeling and evaluation of coherent equalizer performance in the surf zone environment. James C. Preisig (Dept. of Appl. Ocean Phys. and Eng., Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

The environmental conditions encountered in the surf zone give rise to acoustic channel characteristics which can vary significantly. A state space model is developed which describes these characteristics in sufficient accuracy for evaluating their effect on the performance of a coherent acoustic communications system in the surf zone environment. Using techniques developed by the author and presented at the two prior ASA meetings, the impact of the acoustic channel characteristics on coherent communications system performance is analyzed. The theoretical results

are compared to experimental data collected at the Scripps Pier during the Fall of 2000 over several regimes of environmental conditions. [Work supported by ONR.]

9:15

4aUW5. Analysis of acoustic telemetry signals in the surf. John S. Stroud, Kerry W. Commander (Coastal Systems Station, Code R21, 6703 W. Hwy 98, Panama City, FL 32407-7001, stroudjs@ncsc.navy.mil), Robert J. McDonald, and JoEllen Wilbur (Coastal Systems Station, Panama City, FL 32407-7001)

The 2000 Surf Zone Acoustic Telemetry Experiment (SZATE) was conducted off of the Scripps Institute of Oceanography pier in La Jolla, CA. In this experiment, measurements were conducted in a very shallow-water/surf zone (VSW/SZ) region to investigate acoustic propagation and sound channel stability in the coastal area. These measurements utilized two receivers and a broadband source projecting shoreward toward the receivers. The data to be presented were acquired by a computer-controlled system designed and operated by Grant Deane of Scripps Institution of Oceanography. This system provided for 12 18-min data acquisition periods per day. The probative transmissions to be discussed consisted of a variety of linear frequency-modulated (LFM) and binary phase shift keyed (BPSK), as well as 4-ary frequency shift keyed (FSK) signals. Each of the types of signals was used during each data acquisition period, providing a record of propagation throughout the tidal cycle. Results of this acoustic measurement to be presented include the mean, variance, and probability density function of the integrated energy as well as the delay spread (duration) of the complex baseband impulse response of the channel. [Work supported by ONR.]

9:30–9:45 Break

9:45

4aUW6. Statistical modeling of very shallow water/surf zone underwater communications channels. Dale Green (Benthos, Inc., 49 Edgerton Dr., North Falmouth, MA 02556)

The performance of a phase-coherent underwater communications system in a nonstationary environment is dependent in large measure on the temporal coherence of each of the principal components (paths) over which the signal arrives at the receiver. It was found that even though a path may bring considerable energy to the receiver, if it is a rapidly scintillating arrival (i.e., short temporal coherence), it can degrade the performance of a channel equalizer in the same way that a noise burst at the same delay and power would do. A method is described for estimating the temporal and inter-lag coherence properties of the time-varying channel impulse response. The method is applied to an extensive database reflecting a recent experiment in the VSW/SZ channel at the Scripps' pier in California. A statistics-based channel model is described, which, when used in a real-time simulator, provides validated analog realizations of signals of opportunity passing through the channel. [Work supported by ONR, Code 32 OE, under the direction of Dr. Tom Swean.]

10:00

4aUW7. Statistical characteristics of sound channel in shallow water. Geng Chen, Yan Chen, and Jian Lan Zhang (Inst. of Acoust., Chinese Acad., Beijing 100080, PROC, chg@oceana.ioa.ac.cn)

In shallow water under the influence of sea bottom and complicated construction of sound speed profile on sound signal propagation, the system characteristics of channel appear strong time-frequency varied characteristics. In this paper several very efficient measuring methods for obtaining the channel characteristics are developed. For example, a method called pulse-to-pulse correlation is used for measuring long time time-variation characteristics of the sound channel, a linear frequency modulated signal, and a pseudo-random coded signal are used for measuring channel system functions as a response function, transfer function, and scattering function. Among these functions, the scattering function is more important from the point of view of the channel match in under water acoustic signal processing. In this paper, the statistical study of measured scattering functions shows that shallow water sound channel meets the requirement of WSSUS (wide sense stationary uncorrelation scattering) channel. In the case of multipath channel, if the difference of time delays among paths is over the time duration of the code chip, an intersymbol interference will occur. To overcome the ISI a quite simple coherent accumulation of different arriving paths is realized. In the case of shallow water acoustic communication the coherent accumulation shows its efficiency obviously.

10:15

4aUW8. Intensity variations of high-frequency broadband acoustic scatter from sea surface in coastal environment. Luc Lenain, Mohsen Badiey (College of Marine Studies, Univ. of Delaware, Newark, DE 19716), and Stephen E. Forsythe (Naval Undersea Warfare Ctr., Newport, RI 02341)

Data from a broadband propagation experiment for mid-to-high-frequency (0.6–18 kHz) in the very shallow water (15 m) of Delaware Bay are studied over a wide range of environmental conditions. Physical parameters such as temperature and salinity as well as hydrodynamic parameters such as surface waves, tide, and current were measured simultaneously with acoustic transmissions. Three hydrophones were utilized to record the received signals in the source/receiver range of 389 m. A beamforming technique is used to allow examination of the arrival angle as a function of arrival time. Statistical correction is applied to normalize the arrival signals and to correct for sidelobes. Four ray paths having only one interaction and scattered by the sea surface into micro-multipaths are considered for analysis. The focus of this study is on the evaluation of these signals as a function of sea surface roughness during several tide cycles. To analyze the data, PE calculations are conducted and parameter studies for different wind speeds are performed using the measured sea surface wave spectra. [Work supported by ONR and Sea Grant.]

10:30

4aUW9. A comparison of normal mode and parabolic equation range-dependent propagation models as tools for acoustic communication in shallow water. Natalia A. Sidorovskaia (Phys. Dept., Univ. of Louisiana–Lafayette, UL Box 44210, Lafayette, LA 70504-4210, nsidorovskaia@louisiana.edu), Robert L. Field (Naval Reseach Lab., Stennis Space Center, MS 39529), Cheryl L. Sephus, George E. Ioup, and Juliette W. Ioup (Univ. of New Orleans, New Orleans, LA 70148)

Accurate prediction of the effects of a shallow-water propagation channel on acoustic signal distortion would aid in the optimal design of shallow-water communication systems. It has been shown that the correlation between modeled (based on a parabolic equation approach) and measured ocean responses decreases significantly with increasing propagation range because of the mismatch in relative amplitudes and time delays between the modeled and measured bottom-interacting arrivals. This raises the question whether the discrepancies are related to the particular modeling algorithm or to insufficient information about the propagation channel. A new normal mode range-dependent algorithm is used to simulate the acoustic response for the source–receiver configuration used in an experiment in the Atlantic Ocean at the Southern end of the Blake Plateau in which a 25–150 Hz linear, frequency-modulated source signal

is received on a 15-element vertical array. The normalized correlation coefficients between source and receiver signals as a measure of signal distortion are computed as functions of receiver depth and propagation range and compared for both modeling algorithms. The influence of the inaccuracy of the bottom parameters, particularly the compressional to shear wave conversion in determining high distortion areas, is also discussed.

10:45

4aUW10. Underwater acoustic channel simulator for broadband communication systems. Byung-Chul Kim and I.-Tai Lu (Dept. of Elec. and Computer Eng., Polytechnic Univ., 901 Rte. 110, Farmingdale, NY 11735, itailu@poly.edu)

In underwater acoustic (UWA) communications, a reliable numerical simulator is an economical and effective tool for system design and development or for performance prediction and assessment. Some useful channel simulators have been developed in past decades. However, they are only suitable for narrow-band systems, where the channel characteristics are assumed to be the same within the whole signal bandwidth. In order to meet the ever-growing demands of higher and higher data rates for modern UWA communications, the need of broadband simulators is acute. In this paper, a broadband UWA channel simulator is developed, which models UWA channel as a time varying filter with added colored noises. Special emphasis has been given to Doppler effects caused by motion of sensors, platforms and fluctuation of reflecting boundaries. The time varying filter is composed of variable delay generators and amplitude modulators to simulate the Doppler effect for the broadband signal in the UWA channel. By generating a time-varying delay in addition to amplitude variation for each eigenray path, compression and expansion effects of the waveforms and frequency spread are effectively realized. This simulator is employed for designing an orthogonal frequency division multiplexing (OFDM) system in a shallow-water channel.

11:00

4aUW11. The effect of diffused arrivals on the performance of phase coherent underwater acoustic communications. T. C. Yang, Kwang Yoo (Naval Res. Lab., Washington, DC 20375), and Azmi Al-Kurd (Patton Electronics, Inc., Gaithersburg, MD 20879)

The delayed multipath arrivals cause intersymbol interference (ISI) in underwater acoustic communications, which can sometimes extend over tens to hundreds of symbols. The performance of the decision feedback equalizer (DFE) requires a reasonably accurate estimate of the number of tap coefficients as too few taps fail to remove the ISI and too many taps increases the mean square error (the processor noise). An initial estimate of the channel impulse response is often used to determine the number and positions of the taps. This decision is clear-cut for well-defined (usually high-amplitude) arrivals but ambiguous for diffused arrivals that are weak and undistinguished. It should be noted that diffused arrivals can be a significant component of the signal, resulting usually from sound scattering from fluctuating surface/bottom and random media. Its amplitude and complexity are compounded and enhanced by platform motion. This work illustrates the effect of diffused arrivals on the performance of the equalizer by introducing diffused arrivals in a simple model and proceeds to show the effect of wind speed on the bit error rate in realistic ocean channels. [Work supported by ONR.]

11:15

4aUW12. The effect of Doppler shift/spread on the performance of the phased locked loop for phase coherent underwater acoustic communications. T. C. Yang (Naval Res. Lab., Washington, DC 20375)

Phase coherent acoustic communication systems, which employ joint adaptive channel equalization and phase-locking synchronization, have been shown to yield a much higher symbol rate (e.g., 1–2 kb/s at 2–5 kHz) by successfully tracking and updating time varying acoustic channels dominated by multipath components. While the decision feedback equalizer (DSE) can in principle adapt to the change of both the signal amplitude and phase, the adaptation rate (requiring typically 20–30 symbols) is in practice too slow compared with the phase rate, hence an additional phase locked loop (PLL) is required. However, the PLL has its limitations. It is found in data analysis that the failure of the PLL is responsible for the high bit error rate (BER) for many situations. This paper illustrates the effect of the phase shift/fluctuations (emulating the Doppler shift/spread) on the performance of PLL and BER, keeping the channel impulse response fixed in time. It is shown with at-sea data that the BER is significantly reduced by removing the phase wander after precise Doppler estimation. The residue phase is apparently below the tolerance threshold as determined by simulations. [Work supported by ONR.]

11:30

4aUW13. Temporal coherence: A critical function for phase coherent underwater acoustic communications. T. C. Yang (Naval Res. Lab., Washington, DC 20375) and Azmi Al-Kurd (Patton Electronics, Inc., Gaithersburg, MD 20879)

The temporal evolution of an acoustic channel can be characterized by its temporal correlation function from which temporal coherence (the value) and temporal correlation time (at coherence equal to, say, 0.5) can be deduced. Temporal coherence and correlation time have been experimentally measured in the past when the coherence time is much longer than the pulse duration. This is the intra-packet coherence. As frequency increases to tens of kHz or above, the coherence time can be shorter than the packet length. Under such circumstances, the channel coherence can impede the equalizer performance. In the extreme case when the coherence time is shorter than the multipath delay time, the later arrivals may travel through a channel that is different from the earlier arrivals. This paper presents the simulation of communications signals through a time-evolving channel including random media due to internal and turbulence for which temporal correlation time can be predicted for stationary source/receivers. For moving source/receivers, the temporal correlation time (the inverse of the Doppler spread) can be significantly shorter. The performance of the equalizer is found limited by the coherence value. Some at-sea data are presented that support the simulation predictions. [Work supported by ONR.]

Session 4pAA

Architectural Acoustics, Engineering Acoustics and Noise: Measurements and Predictions of Building Materials and Systems

Brandon Tinianov, Chair
Acoustical Laboratory, Johns Manville Technical Center, 10100 West Ute Avenue, Littleton, Colorado 80127

Chair's Introduction—1:00

Contributed Papers

1:05

4pAA1. An experimental study for multilayered building elements with a sub-study on the effect of finite-size test specimens. Selma Kurra (Istanbul Tech. Univ., Physical Environ. Control Unit, Faculty of Architecture, Taksim 80191 Istanbul, Turkey) and David Arditi (Illinois Inst. of Technol., Chicago, IL 60616)

Sound transmission through multilayered building elements was investigated in this experimental study in which sound transmission losses were measured in 28 wall samples with various layer configurations including gypsum board, steel plates, vinyl damping material, glasswool, and airgap. ASTM E90-90 standard was used in the experiment that was conducted at the Riverbank Acoustical Laboratories within a study supported by a grant from IIT/IITRI. The effects of some physical parameters were investigated and comparisons were made in terms of frequency-dependent TL, Rw, and STC values. Test specimens of two sizes (10 m2 and 2.9 m2), were used to derive a size-effect correction which could be used in comparison of measured data and those to be calculated for infinite elements. It was found that the sound insulation values for small-size elements were higher at low frequencies than those of large-size elements, whereas they were lower at mid and high frequencies for both single- and multilayered elements. Statistical analysis indicated that the difference was significant and more emphasized for double gypsum walls with airgaps. Based on the experimental data, variations of transmission loss nd insulation values according to gap width and layer configuration are presented in comparative charts and the results are discussed.

1:20

4pAA2. The sound insulation of gypsum board on resilient supports. John Bradley and Jennifer Birta (Natl. Res. Council, Montreal Rd., Ottawa K1A 0R6, Canada)

This work reports on the development of a simple model to explain the effects of adding resilient channels to a rigid double-leaf wall or floor system. The surface layers mounted on resilient channels are treated as a vibration isolator with a fundamental resonance frequency determined by the combination of the stiffness of the resilient channels and the stiffness of the air cavity along with the masses of the surface layers. The effective stiffness of various common forms of resilient channels is determined to be approximately equivalent to that of a 160-mm air cavity. The damping of constructions including resilient channels and cavities with sound-absorbing material is found to vary with cavity depth. The model does not attempt to explain high-frequency transmission through the resilient supports but does permit the optimization of the low-frequency performance of double-leaf constructions with resilient channels. It also demonstrates that the practical range of improvements is limited.

1:35

4pAA3. A reduction effect of the floor impact sound after installation of the sound insulation for SMT factory. Jaehee Jang (R&D Ctr., SKEC, 192-18 Kwanhun-Dong, Chongro-Gu, Seoul 110-300, South Korea, jhjang@skec.co.kr)

This study aims to examine some kinds of plans to solve the ploblem of noise and vibration transmitted to adjacent rooms, especially to the downstairs, when using SMT (surface mount technology) equipment. To achive this, a kind of system for preventing noise and vibration was applied, and the effects of sound insulation performance for that were tested and analyzed. To reduce floor impact noise, a floating floor system using a plastic pallet and rubber was applied. The applied plastic pallet is 30 cm wide, 60 cm deep, and 6 cm high and designed to array on a concrete slab. Above this plastic pallet, a 14-cm-thick concrete slab was constructed. After installation of this system the floor impact sound was reduced about 7 dB(A). For more reduction, a ceiling with a 5-cm-thick glass wool and twofold of 0.9-cm-thick plywood was constructed at the directly below downstairs. According to this, 4 dB(A) was additionally reduced.

1:50

4pAA4. Effect of vibration of porous absorbent layers in the cavity of a multiple-leaf structure on its acoustic properties. Kimihiro Sakagami (Environ. Acoust. Lab., Kobe Univ., Rokko, Nada, 657-8501 Kobe, Japan), Motoki Yairi (Kobe Univ., 657-8501 Kobe, Japan), Masayuki Morimoto (Kobe Univ., Rokko, Nada, 657-8501 Kobe, Japan), Atsuo Minemura, and Kei Andow (Kajima Tech. Res. Inst., Tokyo, Japan)

Porous absorbent layers are often used in the cavity of multiple-leaf acoustic panels in buildings to reduce the sound power radiated by the panels or to enhance the panel absorption. It is known that bulk vibration of the porous layer can result in the reduced transmission loss of the multiple-leaf acoustic panel. This work presents a simple method to account for the effect of the bulk vibration of the layer on its acoustic resistance. In this method the *''modified''* flow resistance of the layer is defined as the real part of the parallel impedance in the electro-acoustical equivalent model. The model relates the impedance to the flow resistivity and the density of the porous layer. The *''modified''* flow resistance is then used to predict the propagation constant and the characteristic impedance of the material using the appropriate model for sound propagation in rigid-frame porous materials. As an application, the effect of a porous layer in the cavity of double-leaf panels on the radiated sound power is

studied. Calculated examples show that the acoustic efficiency of the layer decreases at low frequencies. The reduction in the acoustic efficiency is more pronounced in the case of porous materials with high-flow resistivity.

2:05–2:20 Break

2:20

4pAA5. Experimental structure factor in stratified rockwool samples. M. A. Picard Lopez, P. E. Solana Quiros, and J. V. Arizo Serrulla (Univ. Politecnica de Valencia, E.T.S. Ing. Industriales, E.T.S. Ing. Caminos Canales y Puertos, Camino de Vera S/N, 46022 Valencia, Spain)

In the study of porous materials, the structure factor allows us to relate the phenomenological approach and the microestructural. In these approaches the structure is supposedly more rigid than that of the contained air in its pores. In the literature it is demonstrated that remarkable efforts have been made to modeling the shape of the pore whose section in a first approach was considered cylindrical. But when the shape is studied on the microscope stratified commercial samples, one can observe that as the density diminishes it is practically impossible to assign a shape to the pore. For these situations, it is not easy to assign a value to the structure factor by means of geometric considerations in commercial samples. In this work an acoustic evaluation method has been developed in the phenomenological approach that can be applied to real samples. The existing formulation has been revised in several suppositions. The function of the obtained results can be considered as a parameter that helps toward the design and improvement of these materials and the information is complete with other parameters, such as flow resistivity.

2:35

4pAA6. Nonlinear behavior of poroelastic absorbing materials. Olga Umnova, Keith Attenborough, and Alan Cummings (Dept. of Eng., The Univ. of Hull, Hull HU6 7RX, UK)

A theory for the non-linear acoustic behavior of porous materials with an elastic frame composed of incompressible grains or fibers has been derived. Johnson's expression for complex flow resistivity [D. L. Johnson *et al.*, J. Fluid. Mech. **176**, 379–402 (1987)] has been combined with the dependence of dc flow resistivity on the amplitude of particle velocity (Forchheimer's nonlinearity). The frame compressibility has been assumed to be linear. The system of coupled nonlinear equations for slow and fast planar compressional modes propagating in the material has been solved approximately by method of slow varying amplitudes. As a result of calculations the dependence of surface impedance on the amplitude of sound wave has been obtained for either fixed or moving surfaces. For all of the parameter values considered so far, it is predicted that the amplitude of the reflection coefficient increases with the incident sound intensity. This approach has been extended to allow for waves reflected from rigid backing of a layer of material. The influence of nonlinearity on the impedance of a sample of finite thickness has been estimated. [Work supported by USARSDG (UK), Contract No. R7D 8901-EN-01.]

2:50

4pAA7. Source dependency on anechoic chamber validation. Kenneth A. Cunefare, Van Biesel, Anne-Marie Albanese, Lisa Chang (G. W. Woodruff School of Mech. Eng., The Georgia Inst. of Technol., Atlanta, GA 30332-0405), and Mark Holdhusen (The Georgia Inst. of Technol., Atlanta, GA 30332-0405)

The ANSI S12.35-1990 and similar ISO standard establish procedures for measuring sound power in anechoic and hemi-anechoic chambers. In addition, the standards outline procedures for use in the qualification of the test chambers that are used for such measurements. The basic qualification procedure requires a draw-away test from a sound source. Different sources are recommended to generate test signals over low-, medium-, and high-frequency ranges. These ranges overlap at crossover frequencies, where two different sources may be used to test the chamber. For example, the mid-range source is constructed by bolting together two 10-cm speakers face-to-face. The recommended high-frequency source is a baffled speaker radiating through a narrow cylindrical tube. The work presented here investigates how the qualification tests at mid and high frequencies are impacted by the use of sound sources fabricated to the recommended configurations, but using speakers from different manufacturers, and, for the high-frequency source, constructed with different diameters and lengths of tube.

3:05

4pAA8. Intensity as a quality assessment method for anechoic chambers. Francis J. Babineau and Brandon D. Tinianov (Johns Manville Tech. Ctr., 10100 W. Ute Ave., Littleton, CO 80127)

Data regarding the Johns Manville hemi-anechoic chamber qualification according to ISO 3745 Annex A and ANSI S12.35 Part 9 were presented at ASA/NoiseCon in December 2000. In addition to techniques based on these standard test methods, an alternate qualification technique was devised using a pressure–pressure intensity probe. The basis of the technique stems from the fact that the sound pressure and intensity should be nearly equal in a free-field environment. The difference between the sound pressure and sound intensity at a point is known as the pressure-intensity index (p-i index), a common ISO measure of source strength. P-i index measurements were made along several measurement paths and at discrete points at the predicted boundaries of the room. Results of this nonstandard test were compared to those of traditional qualification techniques. Although the method uses a possibly uncommon instrument, the results suggest a greater tolerance for various sound sources and allow for quick computation of chamber qualification.

Session 4pAB

Animal Bioacoustics: Animal Responses to Sound

Robert J. Dooling, Chair
Psychology Department, University of Maryland, College Park, Maryland 20742

Contributed Papers

1:15

4pAB1. Directional responses of blueback herring to ultrasonic pulses. Carl R. Schilt, Charlie Escher (Mevatec Corp., 458 Evergreen Dr., North Bonneville, WA 98639, schilt@saw.net), and John M. Nestler (U.S. Army Res. and Dev. Ctr., Vicksburg, MS 39180-6199)

Adult blueback herring (*Alosa aestivalis*) have been found to be directionally responsive to very high (118 kHz fundamental) pulsed sounds. The experiments were conducted in a pen containing filtered fresh water. Fish responses were videotaped. Fish usually only flinched in response to the onset of continuous wave stimuli. They usually moved immediately away from the source of short repetitive impulse sounds (duration from around 20 cycles to about four ms and less than 174 dB *re* 1 mPa at the nearest fish). Although a variety of pulse repetition rates were tested, no clear effect of that variable on fish response was apparent within the limits set by the experimental design. Sounds of similar frequency (120 kHz fundamental, 4 ms pulse duration, 5 Hz repetition rate) but over 200 dB *re* 1 mPa at the nearest fish also produced a behavioral response which sometimes resulted in the fish moving quickly away from the sound source but always included a dramatic and immediate coalescing into a tight school, as is often seen in the presence of a predation threat. Possible implications for a putative coevolution between Alosine shads and marine mammel predators as well as for ultrasound source localization are discussed.

1:30

4pAB2. Computational study of cat eardrum mechanics. Jonathan P. Fay, Sunil Puria, and Charles R. Steele (Stanford Univ., 262 Durand Bldg., Stanford, CA 94305, puria@stanford.edu)

The function of the eardrum and the middle ear is to resolve the acoustic impedance mismatch between the air of the outside world and the fluid of the inner ear. Without an impedance matching device, very little acoustic energy would be absorbed by the inner ear and hearing would be severely limited. It remains a mystery how the eardrum accomplishes this over the audible frequency range. A computer simulation of the cat eardrum was constructed. The vibrations of the eardrum were fully coupled to the acoustics of the ear canal and the dynamics of the middle ear bones. The eardrum's fibrous microstructure was taken into account. This model appears to be valid up to 20 kHz. This represents a significant improvement over previous modeling efforts. With this model, the following questions surrounding the evolution and design of the eardrum have been addressed: (1) Why does the eardrum have its distinctive conical and toroidal shape? (2) What is the significance of its highly organized fibrous structure? (3) What is the advantage of its angular placement in the ear canal? and (4) What is the benefit of complex vibrational modes on the eardrum? [Work supported in part by research grant R29 DC03805 from NIDCD.]

1:45

4pAB3. Masking by harmonic complexes with different phase spectra in zebra finches. Amanda M. Lauer, Robert J. Dooling (Dept. of Psych., Univ. of Maryland, College Park, MD 20742, alauer@psyc.umd.edu), Marjorie R. Leek, and Jennifer J. Lentz (Army Audiol. and Speech Ctr., Walter Reed Army Medical Ctr., Washington, DC 20307)

Temporal waveform characteristics differentially affect masking by harmonic complexes in birds and humans. In birds, complexes with relatively flat temporal envelopes are more effective maskers than complexes with highly modulated temporal envelopes. Flat temporal waveforms constructed using positive and negative Schroeder-phase algorithms result in large masking differences in humans, but little masking difference in birds. This effect may be related to differences in the cochlear phase characteristics in the two species. A psychophysical estimate of the phase characteristic has been reported in humans by finding the phase spectrum of a harmonic masker that produces the least masking. Here we estimated the phase characteristic for zebra finches using operant conditioning methods to measure masked thresholds for tones embedded in harmonic complexes with phases selected according to scaled modifications of the Schroeder algorithm. There was little difference between birds and humans when the scalar was between 0.0 and -1.0 (i.e., negative Schroeder phases), but humans showed less masking than birds when the scalar was positive. These results may reflect a more linear phase characteristic along the avian cochlea. [Work supported by NIH R01 DC00198 and NSRA DC00046.]

2:00

4pAB4. The precedence effect in budgerigars (*Melopsittacus undulatus*). Micheal L. Dent and Robert J. Dooling (Dept. of Psych., Univ. of Maryland, College Park, MD 20742, mdent@psyc.umd.edu)

When localizing sounds in a reverberant environment, the precedence effect allows us to localize direct sounds accurately while disregarding the many echoes in that environment. The precedence effect has been well studied behaviorally in humans and behaviorally and physiologically in other animals. Until now, however, precedence effect experiments in birds have been limited to barn owls. These birds have highly specialized auditory systems; nothing is really known about how unspecialized birds deal with competing sound sources. Budgerigars (*Melopsittacus undulatus*) are excellent subjects for this type of investigation since so much is known about their hearing. These birds have unremarkable sound localization abilities, but exhibit free-field binaural unmasking at amounts similar to those found in humans with much larger heads. We examined whether budgerigars exhibit the precedence effect, and if the phases of the precedence effect were similar to those found in humans and other animals. Psychoacoustic methods were used to measure the discrimination performance of click pairs from different locations separated by a short delay. The results suggest that budgerigars exhibit the precedence effect and that the time-courses are similar to those found in humans and other animals. [Work supported by NIH MH12698 to MLD and DC00198 to RJD.]

2:15

4pAB5. Duplex auditory distance assessment in a small passerine bird and the relationship with sound attenuation and vocalization production. Brian S. Nelson (Dept. of Biol., Indiana Univ., Jordan Hall 142, 1001 E. Third St., Bloomington, IN 47405-3700, brsnelso@indiana.edu)

Eastern towhees, *Pipilo erythrophthalmus* (Emberizidae, Passeriformes, Aves) estimate how loudly conspecific vocalizations are produced using correlated spectral and temporal variables and often misjudge speaker distance in the field when vocalization source sound-pressure level (SPL) is varied experimentally. I demonstrate that towhees often misjudge speaker distance when played SPL altered stimuli produced with sound frequencies below 3.5 kHz but rarely misjudge speaker distance when played SPL altered stimuli produced with sound frequencies above

3.5 kHz. I describe the homogeneous physical-acoustic structure of towhee Florida scrub habitat and demonstrate that sound frequencies below 3.5 kHz attenuate more reliably over distance in comparison with sound frequencies above ~3.5 kHz (i.e., sound frequencies below 3.5 kHz propagate with less variation in attenuation across sites sampled, several speaker and microphone elevations and between days). Lastly, I describe acoustical measurements obtained from calls produced by 52 caged subjects in the field, which suggest that sound frequencies to each side of 3.5 kHz are produced independently or are otherwise emphasized differently during production. These experiments suggest that auditory distance perception in the towhee might be described as ''duplex'' and that variation in sound propagation might often mediate a close relationship between vocalization production and auditory perception.

2:30

4pAB6. The underwater noise of vessels in the Hervey Bay (Queensland) whale watch fleet and its impact on humpback whales. Robert D. McCauley (Ctr. for Marine Sci. Technol., Curtin Univ., Kent St. Bentley, Perth, Western Austalia) and Douglas H. Cato (Defence Sci. and Technol. Organisation, Pyrmont NSW, Australia)

In 1994 the underwater noise of 19 vessels involved in whale watching was measured. Vessels ranged from 1.5–70 tons and included yachts, runabouts and high-speed and displacement mono and multihulls. Except for one water-jet trimaran, all vessels were propeller driven. Unlike the directional patterns reported for merchant shipping, each vessel projected lobes of sound fore and aft with lower levels abeam. In the high speed planing vessels this was exacerbated by the deep propellers exposed forward, and the vessel squat while on the plane. All vessels displayed a linear relationship with broadband noise level and the logarithm of speed. Although unique for each vessel, as a rule of thumb doubling the speed doubled the detection range. The response of whales to vessel noise was as much a function of the rate of change of noise as its steady level. Rapid increases in noise produced more responses. Vessels which, by their design, required constant maneuvering to maintain station produced greater adverse responses from whales. Some design criteria important in reducing noise impacts from whale watch vessels include shielding of the propellers in the forward direction, windage in relation to draft, slow speed steerage, machinery noise reduction and passenger viewing access.

2:45

4pAB7. The ultrasonic receiving system of a Pacific white-sided dolphin: Part 1. Experimental results. Shigeru Takagi (Nagasaki Res. and Development Ctr., Mitsubishi Heavy Industries, Ltd., 5-717-1, Fukahori-machi, Nagasaki 851-0392, Japan), Akira Takemura, Takashi Koido (Nagasaki Univ., Nagasaki 852-8521, Japan), and Kazuhiro Yoshizumi (Nagasaki Res. and Development Ctr., Mitsubishi Heavy Industries, Ltd., Nagasaki 851-0392, Japan)

To investigate the ultrasonic receiving system of a dolphin's sonar, three experiments were carried out using dead dolphins' heads. The first experiment was an anatomical investigation by magnetic resonance imaging followed by dissection. The second experiment was measurements of the receiving sensitivity distribution on a dolphin's head surface to locate receiving areas with high sensitivity. The third experiment was muscle excision to determine which organs contribute to the ultrasonic sound wave propagation from the highly sensitive areas to the tympanic bone. These experimental results show that the receiving sensitivity is highest at the middle points of both the left and right mandibula, at which points the left and right digastrici terminate, respectively. The highly sensitive areas lie along lines just inside of the lower mandibula, where both digastrici are located. Furthermore, not only are the highly sensitive areas along the right mandiblum connected acoustically to the right tympanic bone but those on the left side are also cross-connected by muscles on the lower jaw and vice versa. Therefore, a dolphin's ultrasonic receiving system may be modeled by a multi-input/two-output system in which each right or left ear, respectively, is a multi-input/single-output system.

3:00

4pAB8. The ultrasonic receiving system of a Pacific white-sided dolphin: Part 2. Multi-input/two-output system. Shigeru Takagi (Nagasaki Res. and Development Ctr., Mitsubishi Heavy Industries, Ltd., 5-717-1, Fukahori-machi, Nagasaki 851-0392, Japan), Akira Takemura, Takashi Koido (Faculty of Fisheries, Nagasaki Univ., Nagasaki 852-8521, Japan), and Kazuhiro Yoshizumi (Nagasaki Res. and Development Ctr., Mitsubishi Heavy Industries, Ltd., Nagasaki 851-0392, Japan)

According to anatomical investigations, measurements of the receiving sensitivity distribution and investigation of ultrasonic sound wave propagation paths, each ear may be constructed as a multi-input/single-output system in the head. These two multi-input/single-output systems may be considered to act together as one multi-input/two-output system. Here, the input points are highly sensitive areas lying just inside of the lower mandibula. The two output points are the tympanic bones in each mandibula. All input points are connected acoustically to both tympanic bones by the the digastrici, mylohyoidei and sternohyoid muscles. Mathematical simulations were carried out using this multi-input/single-output system, applying impulse response functions previously measured from the right tympanic bone to each of the measuring points on the dolphin's head surface. From the results of these simulations, it is considered that the highly sensitive areas located in lines along both mandibula act as line arrays, so the beamwidth of the receiving system can be narrow. Furthermore, the spectrum of the echo signal is changed by the target direction, and human beings can recognize this difference when frequency components of the echo signal are shifted down to the range of human hearing.

3:15

4pAB9. Underwater masked thresholds of the West Indian manatee. Edmund R. Gerstein, Laura A. Gerstein, Steven E. Forsythe, Joseph E. Blue (Leviathan Legacy, Inc., 1318 S.W. 14th St., Boca Raton, FL 33486, gerstein2@aol.com), and Steven E. Forsythe (Naval Undersea Warfare Ctr., Newport, RI 32841)

Underwater acoustic masked thresholds at 0.5, 1.6, 3, 6, 12, 18, and 26 kHz were obtained from two test-sophisticated West Indian manatees using a forced-choice, two-alternative, paradigm and an up–down staircase psychometric method. Pure tones were presented under two different stimulus conditions: (1) as pulsed signals; and (2) as nonpulsed signals. Under both stimulus conditions the tones were presented against a 1/3 octave continuous white-noise masker. Three different masking intensities were used to simulate moderate ambient noise levels recorded from manatee habitats. Masked thresholds across frequencies increased in a linear fashion with increased masking intensity. Critical ratios for both stimulus conditions increased with higher frequencies, however, thresholds for pulsed signals were significantly lower than nonpulsed signals suggesting some attenuation or higher-order inhibitory process affected the perception of nonpulsed tones. Comparisons of critical ratios with other marine mammals suggest manatees have acute filtering abilities for detecting pulsed tones under continuous noise conditions. While manatees do not exhibit a rich vocal repertoire to account for acute filtering abilities, they are passive listeners which may be adapted to selectively filter out continuous noise in favor of detecting biologically significant sounds like their pulsed 200–500 ms species-specific calls.

Session 4pBB

Biomedical Ultrasound/Bioresponse to Vibration: Bioeffects and High Intensity Ultrasound

William D. O'Brien, Chair

Department of Electrical and Computer Engineering, University of Illinois, 405 North Mathews, Urbana, Illinois 61801

Contributed Papers

1:30

4pBB1. Contrast agent-induced cardiac arrhythmias in rats: Exposimetry and clinical findings. Sarah A. Hartleben, William D. O'Brien, Jr., Leon A. Frizzell (Dept. of Elec. and Computer Eng., Univ. of Illinois, 405 N. Mathews, Urbana, IL 61801), and James F. Zachary (Univ. of Illinois, Urbana, IL 61802, zacharyj@uiuc.edu)

Other than minor contraindications, no known meaningful health risks have been reported following the use of microbubble ultrasound echogenic contrast imaging agents in the United States. Recently, a report from the Netherlands has documented premature ventricular contractions during triggered second harmonic imaging of a contrast agent for myocardial perfusion in healthy male volunteers. Because of this report, we designed a study to characterize electrocardiogram conduction abnormalities in rat hearts exposed to pulsed ultrasound following injection of a contrast agent. We used a focused $f/1$ 3.1-MHz, transducer that had in situ (at the heart) peak rarefactional and compressional pressures of 15.9 MPa and 36.1 MPa, respectively, 1.7-kHz PRF, 1.2-μs pulse duration, and MI of 5.8. Rats had no meaningful conduction abnormalities when exposed to ultrasound alone or contrast agent alone but developed premature atrial and ventricular complexes and polymorphic ventricular tachycardia when a contrast agent was administered intravenously and the heart exposed to pulsed ultrasound. When ultrasound exposure was concluded, cardiac arrhythmias ceased but reoccurred when ultrasound exposure was resumed. These results suggest that a microbubble contrast agent through its biomechanical interactions with pulsed ultrasound has the potential to induce conduction abnormalities leading to potentially hazardous cardiac arrhythmias. [Work supported by NIH Grant No. HL58218.]

1:45

4pBB2. Contrast agent-induced cardiac arrhythmias in rats: Lesions and mechanisms of injury. James F. Zachary (Dept. of Veterinary Pathobiology, Univ. of Illinois, 2001 S. Lincoln Ave., Urbana, IL 61802, zacharyj@uiuc.edu), Sarah A. Hartleben, Leon A. Frizzell, and William D. O'Brien, Jr. (Univ. of Illinois, Urbana, IL 61801)

Other than minor contraindications, no known meaningful health risks have been reported during the use of microbubble ultrasound echogenic contrast imaging agents in the United States. At this meeting, we have reported on electrocardiogram conduction abnormalities in rat hearts exposed to pulsed ultrasound following intravenous injection of a contrast agent. Rats had no meaningful conduction abnormalities when exposed to ultrasound alone or contrast agent alone but developed premature atrial and ventricular complexes and polymorphic ventricular tachycardia when a contrast agent was administered and the heart was exposed to pulsed ultrasound. When ultrasound exposure was concluded, cardiac arrhythmias ceased but reoccurred when ultrasound exposure was resumed. Acute multifocal coagulative necrosis of cardiac rhabdomyocytes was observed microscopically in myocardium of rats exposed to ultrasound following injection of a contrast agent. Because cardiac arrhythmias were induced only when a contrast agent interacted with ultrasound during exposure, the presence of myocardial lesions alone did not appear sufficient to serve as foci for ectopic electrical activity. These results suggest that a microbubble contrast agent through its biomechanical interactions with pulsed ultrasound has the potential to induce conduction abnormalities leading to potentially hazardous cardiac arrhythmias. [Work supported by NIH Grant No. HL58218.]

2:00

4pBB3. Potential for contrast agent-induced endothelial cell injury in rabbit arteries exposed to pulsed ultrasound. James F. Zachary (Dept. of Veterinary Pathobiology, Univ. of Illinois, 2001 S. Lincoln Ave., Urbana, IL 61802, zacharyj@uiuc.edu), Leon A. Frizzell, and William D. O'Brien, Jr. (Univ. of Illinois, Urbana, IL 61801)

No known meaningful health risks have been reported following the use of microbubble ultrasound echogenic contrast imaging agents in the United States. We designed a preliminary study to examine the potential for injury to endothelial cells in vivo following the interaction of intravascular contrast agent with ultrasound. We used a focused $f/1$ 3.1-MHz, transducer that had in situ (at the skin) peak rarefactional and compressional pressures of 20 MPa and 46 MPa, respectively, 1.0-kHz PRF, 1.2-μs pulse duration, 180-s exposure duration, and MI of 5.8. The central auricular artery of a rabbit ear was exposed and then contrast agent administered intravenously in the lateral vein of the contralateral ear. The central auricular artery was again exposed to ultrasound at additional sites. No significant lesions were observed in arteries of rabbits evaluated immediately following exposure. When evaluated 24 or more hours after exposure, necrosis of vascular endothelial cells and subjacent smooth muscle cells, early reparative responses, leukocyte margination, and thrombosis were observed. Because of the wide use of echocardiography and contrast agents, the described vascular phenomena need additional evaluation for short and long term effects related to thrombosis and atherosclerosis in at risk patients. [Supported by NIH Grant No. HL58218.]

2:15

4pBB4. Ultrasound assisted clot lysis for stroke therapy. George J. Shaw (Dept. of Emergency Medicine, Univ. of Cincinnati, 234 Goodman St., Cincinnati, OH 45267-0761), Nancy L. Hahn, Kenneth R. Wagner, Daniel S. Kanter, and Christy K. Holland (Univ. of Cincinnati, Cincinnati, OH 45267-0761)

Thrombolytics such as tissue plasminogen activator (tPA) have advanced the treatment of ischemic stroke. To aid in the development of a transcranial ultrasound thrombolysis system (TUTS), the synergistic thrombolytic effect of tissue plasminogen activator (tPA) and 1-MHz ultrasound was assessed *in vitro* in a porcine clot model. Thirty clots were made by incubating 2-ml aliquots of citrated whole pig blood with 2 μl of bovine thrombin (1 NIH Unit/μl) at 37 °C for 3 h, refrigerated overnight, blotted, and weighed. Clots were placed in an acoustically transparent latex sample holder filled with porcine fresh frozen plasma and tPA at a clinically relevant concentration of 0.0126 mg/ml. The clots were either sham exposed to ultrasound (0.0-atm amplitude) or exposed to 1-MHz pulsed ultrasound over a range of duty cycles (10%–100%) at two amplitudes (1.9 and 11.4 atm) for 30 min in a 37 °C water bath, and were weighed to determine the efficacy of thrombolysis. Clots that received continuous wave ultrasound treatment exhibited significantly more thrombolysis than those that were sham exposed ($p=0.001$, students t-test). The TUTS technology promises to improve the effectiveness of thrombolytic

drugs so that ultimately lower concentrations of drugs can be used for faster recanalization of blocked arteries. [Work supported by The Neuroscience Institute, University of Cincinnati Medical Center.]

2:30

4pBB5. Ultrasound enhanced bone formation during mandibular osteodistraction in rabbits. Tarek H. El-Bialy, Richard L. Magin, Carla A. Evans, A. Moneim Zaki (Univ. of Illinois–Chicago, Chicago, IL 60607), Leon A. Frizzell (Univ. of Illinois at Urbana–Champaign, Urbana, IL 61801), and Thomas J. Royston (Univ. of Illinois–Chicago, Chicago, IL 60607)

Distraction osteogenesis is used in orthopedics to lengthen bones, by cutting or breaking the bone and gradually separating the two pieces as new bone fills the intervening space. This study evaluated whether pulsed ultrasound can be used to shorten the long post-distraction period associated with mandibular osteodistraction. Twenty-one rabbits were divided into three groups of seven. The distraction started 72 h after surgically severing both sides of the mandible at a rate of 3 mm/day. Group 1 received ultrasound for 20 min on both sides of the mandible every other day (alternating sides). Group 2 received ultrasound on one side of the mandible every day for 20 min. Group 3 did not receive any ultrasound. Bone formation at the distraction site was assessed by *in vivo* photodensitometry on head radiographs, an *in vivo* (nondestructive) vibratory coherence test across the distraction site, a post-mortem, *ex vivo* (destructive) three-point bending mechanical test, and by post-mortem, *ex vivo* (destructive) histological examination. Statistical analyses performed using analysis of variance (ANVOVA) tests revealed that ultrasound enhanced bone formation at the distraction site as evidenced by the increase in new bone photodensity, mechanical properties, and histological maturation, especially when ultrasound was applied daily.

2:45

4pBB6. Acoustic vaporization of single droplets. Oliver D. Kripfgans, J. Brian Fowlkes, and Paul L. Carson (Dept. of Radiol., Univ. of Michigan Health Systems, Ann Arbor, MI 48109)

It has been observed in the past that droplets in emulsions can be vaporized into gas bubbles by the application of acoustic tone bursts. This paper will show the vaporization of single albumin stabilized droplets as they are exposed to single acoustic tone bursts. An optical microscope was used to monitor single droplets in a flow tube (200 μm diameter). The flow tube was mounted in a small water tank with degassed and heated water (37 °C). A single-element focused transducer with a center frequency of 10 MHz was aimed at the intersect of the tube and the optical beam. A highly dilute droplet emulsion was injected into the flow tube and a special syringe setup was used to move the fluid column in the flow tube back and forth to position individual droplets in the volume of interest. A single sinusoidal tone burst (33 cycles at 10 MHz) was sufficient to transition single droplets in the range from 5 to 15 μm diameter. The formation of the corresponding gas bubbles was monitored and measured in the range of 30 to 90 μm. [Work supported by PHS Grant No. R01HL54201 and US Army Grant No. DAMD17-00-1-0344.]

3:00–3:15 Break

3:15

4pBB7. Role of high intensity focused ultrasound induced cavitation on platelet aggregation. Sandra L. Poliachik, Ryan J. Ollos, Lawrence A. Crum (Ctr. for Industrial and Medical Ultrasound, Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105), and Wayne L. Chandler (Dept. of Lab. Medicine, Univ. of Washington, Seattle, WA 98195)

Our previous studies showed that high-intensity focused ultrasound (HIFU) is capable of producing ''primary acoustic hemostasis'' in the form of platelet activation, aggregation, and adhesion to a collagen-coated surface. In current studies, 1.1 MHz CW HIFU was used to investigate the role of cavitation as a mechanism for platelet aggregation in samples of platelet rich plasma. A 5 MHz passive cavitation detector was used to monitor cavitation activity, and laser aggregometry was used to measure platelet aggregation. Using spatial average intensities from 0 to 4000 W/cm^2, the effects of HIFU induced cavitation on platelet aggregation were investigated by enhancing cavitation activity through use of ultrasound contrast agents, and by limiting cavitation activity through use of an overpressure system. Our results show that increased cavitation activity lowers the intensity threshold to produce platelet aggregation, and decreased cavitation activity in the overpressure system raises the intensity threshold for platelet aggregation. [Research supported by DARPA.]

3:30

4pBB8. High-intensity focused ultrasound for noninvasive disease treatment: Theoretical predictions and experimental results. Francesco P. Curra, Steven G. Kargl, Cyril Lafon, and Lawrence A. Crum (Ctr. for Industrial and Medical Ultrasound, Appl. Phys. Lab., Univ. of Washington, Seattle, WA 98105)

High-intensity focused ultrasound (HIFU) is becoming a widely accepted and ''clean'' modality to induce noninvasive coagulative necrosis of biological tissue for both cancer treatment and hemostasis. Theoretical predictions via the MEDUSA (medical ultrasound algorithm) computer model of ultrasonic fields, temperature responses, and lesion dynamics are simulated for turkey breast. The model accounts for nonlinear sound propagation in inhomogeneous media, arbitrary frequency power law for acoustic attenuation, and temperature and lesion time histories. Generation of gas bubbles within the tissue may also be considered. Results are presented in terms of a comparison study with *in vitro* experiments on common turkey breast. Attention is mainly focused on temperature and lesion evolutions; in particular, induced lesion boundaries and collateral damage to surrounding areas.

3:45

4pBB9. Laparoscopic partial kidney ablation using high-intensity focused ultrasound. Jahangir Tavakkoli, Victor V. Rao, Ralf Seip, Narendra T. Sanghvi (Focus Surgery, Inc., 3940 Pendleton Way, Indianapolis, IN 46226, jahan@focus-surgery.com), Eric Barret, and Arieh L. Shalhav (Indiana Univ. School of Medicine, Indianapolis, IN 46202)

A laparoscopic ultrasound probe was developed for image-guided high-intensity focused ultrasound (HIFU) ablation of renal tissue. The probe expands the application of a commercially available HIFU system (Sonablate^{TM}) to laparascopic operations. It consists of a 4-MHz focused rectangular HIFU piezoceramic transducer (30×10 mm, FL=30 mm) confocally coupled with a circular central element (8 mm in diameter) used for real-time pulse-echo imaging during treatment. Precise mechanical movements of the transducer enable it to treat targeted tissue volumes. The probe was fully characterized by measuring its electrical impedance, acoustic field, and total acoustic power output. Moreover, the acoustic field was simulated using an analytical model. The laparoscopic probe was able to generate acoustic power levels up to 35 W that correspond to focal intensities over 2200 W/cm^2 in tissue, sufficient for tissue ablation through rapid temperature rise (>90 °C) and cavitation mechanisms. The probe was tested in an *in vivo* animal study in which 15 pigs were treated through a sterile laparoscopic procedure. A lesion with dimensions of 23 ×17×11 mm ±1 mm was successfully generated in the left kidney of each pig during a treatment time of less than 30 min. Histology results demonstrated homogeneous well-delineated lesions generated in the target region.

4:00

4pBB10. Design of focused ultrasound phased arrays for prostate treatment. Joseph S. Tan, Leon A. Frizzell (Beckman Inst., Univ. of Illinois, 405 N. Mathews Ave., Urbana, IL 61801), Narendra T. Sanghvi, and Ralf Seip (Focus Surgery, Inc., Indianapolis, IN 46226)

The effect of size and shape of phased arrays for transrectal treatment of the prostate was examined theoretically. The ultrasonic fields were calculated using the point radiator method, including the effects of tissue attenuation. The *G* parameter, the ratio of intensity at the desired focus to

the maximum intensity of any unwanted lobes, was used to compare the performance of various designs. Studies were conducted on spherical segment, cylindrical, and curved cylindrical (cylindrical with curvature added along the length) arrays. The spherical arrays performed well when steered along the array axis (in depth) with G greater than 15 (for an element width of two wavelengths) when steered 15 mm toward the source, but performed poorly when steered along the length of the prostate. The cylindrical array gave the opposite result. The curved cylindrical array yielded performance between the spherical and cylindrical configurations with G greater than 3 (for an element width of two wavelengths) for steering in either direction. With the proper choice of curvature the curved cylindrical array can be designed to provide acceptable performance for steering both in depth and along the length of the prostate. [Work supported by PHS Grant Nos. 5 T32 CA09067 and 1 R43 CA81340.]

4:15

4pBB11. Hydrophone guided transcranial beam steering. Greg Clement and Kullervo Hynynen (Harvard Med. School, Brigham and Women's Hospital, 75 Francis St., Boston, MA 02115, gclement@hms.harvard.edu)

A new method for focusing ultrasound energy in brain tissue through the skull bone is investigated. The procedure is designed for use with a multielement therapeutic transducer array and a small catheter-inserted hydrophone receiver placed in the brain in order to guide the array's focus over a volume of tissue about the receiver. When performed at high intensity, cells within this volume are destroyed. The present study tests the feasibility and range of the method using *ex vivo* human skulls. The skull is placed between a focused ultrasound array and the receiver. Acoustic phase information is obtained from the receiver and used to both electrically shift the beam to new locations as well as correct aberrations due to the skull bone. The method is applied to a 104-element 1.1-MHz array and 120 elements of a 500-element 0.81-MHz array. Using these array configurations it is determined that the method can reconstruct and steer a focus over a distance of 60 mm within the brain. Application of this minimally invasive technique for ultrasound brain therapy and surgery is also demonstrated.

4:30

4pBB12. Simulation of enhanced absorption in ultrasound thermotherapy due to nonlinear effects. Marko Liebler, Siegfried Ginter, Thomas Dreyer, and Rainer E. Riedlinger (Inst. fur Hoechstfrequenztechnik und Elektronik/Akustik, Univ. of Karlsruhe, Kaiserstr. 12, D-76128 Karlsruhe, Germany, marko.liebler@etec.uni-karlsruhe.de)

Ultrasound thermotherapy (USTT) is used to induce thermal coagulation in human tissue by high intensity focused ultrasound. The absorbed ultrasound energy mainly determines the temperature achieved. It strongly depends on the counteracting processes of nonlinear ultrasound propagation and tissue absorption. To investigate the significance of these different effects for USTT, numerical simulations provide a useful tool for treatment planning. A nonlinear full-wave ultrasound propagation model is presented to simulate USTT. It is based on a high-order FDTD algorithm and combines nonlinear ultrasound propagation with a broadband frequency-power-law absorption. Calculations of temperature distributions are done using Penne's bioheat-transfer equation. The propagation model is verified by a comparison of measurements and simulations for pressure signals in water and tissue mimicking materials. Simulations with different source signals demonstrate that both higher amplitude and frequency of the source signal amplify nonlinear effects and therefore multiply the enhanced absorbed energy. It is shown that commonly used linear models considerably underestimate the heat generated and are therefore not able to simulate USTT accurately. Furthermore it is demonstrated that optimization of the source signal parameters, like the amplitude and pulse spectrum, could lead to improvement of the therapy by taking advantage of these nonlinear effects.

THURSDAY AFTERNOON, 7 JUNE 2001 PDR 18, 2:00 TO 4:15 P.M.

Session 4pEA

Engineering Acoustics and Underwater Acoustics: Magnetostrictive Transducers

Stephen C. Butler, Cochair
Naval Undersea Warfare Center, 1176 Howell Street, Newport, Rhode Island 02841

Harold C. Robinson, Cochair
Naval Undersea Warfare Center, Code 2131, 1176 Howell Street, Newport, Rhode Island 02841

Chair's Introduction—2:00

Invited Papers

2:05

4pEA1. Magnetostrictive transducer configurations and modeling. John L. Butler (Image Acoust., Inc., 97 Elm St., Cohasset, MA 02025)

The advent of the highly active magnetostrictive material, Terfenol-D ($Tb_{0.27}Dy_{0.73}Fe_{1.9}$), by Clark and co-workers in the 1970s has yielded the realization of high power magnetostrictive transducers. The national and recent international availability of this material together with price reductions should spawn more common usage of Terfenol-D in transducers. However, the incorporation of this material requires additional design considerations including a different type of equivalent circuit than used in the piezoelectric models along with associated magnetic circuitry including eddy current losses and magnetic bias considerations. We present a review of some of these considerations including properties, fundamental theory, equivalent circuits, matrix representations, and other modeling tools. Along with this background, we present examples of transducer designs, which incorporate the Terfenol-D material in some interesting unconventional designs as well as in some conventional designs.

2:30

4pEA2. A broadband Terfenol-D driven flextensional projector. Raymond Porzio, Blake A. Carnahan, Jr., Aftab Hussain, and Martin L. Marietta, Jr. (Lockheed Martin Corp., Naval Electron. and Surveillance Systems, Undersea Systems, Syracuse, NY 13084-4840)

Since 1994 Lockheed Martin has developed a number of Terfenol-D driven Class VII inverse flextensional underwater projectors, known as ''Terfenol Dogbone Flextensional Projectors.'' The underlying motivation for these projectors has been to match the high-strain, high-energy density, and relatively high-compliance characteristics of magnetostrictive Terfenol-D alloys with the high-compliance characteristics of the dogbone flextensional shell. The resulting projector is reliable, rugged, and has a smaller size, lower-frequency and a broader bandwidth relative to conventional ceramic driven flextensional projectors. The Class VII shell, by virtue of its shape, is designed to naturally accumulate axial compressive stress on the drive assembly with depth. Hence, the dogbone shaped shell is dynamically more compliant to better match the lower modulus of the Terfenol driver. Additionally, both the fundamental bending mode and the second mode of vibration provide dominantly in-phase radiation and well behaved beam patterns. Recently, a small permanent magnet biased projector demonstrated a mechanical quality factor of unity and high source level in its fundamental mode of vibration. A 3-dB band from 2.9 to 6.4 kHz (fundamental mode) and 10 to 18 kHz (second mode) where achieved. The merits, challenge, and future outlook for this technology are discussed. [Work supported by ONR Code 321.]

2:55

4pEA3. A 2.5 kHz magnetostrictive tonpilz sonar transducer design. Stephen C. Butler (Naval Undersea Warfare Ctr., Newport, RI 02841, butlersc@npt.nuwc.navy.mil)

Naval Undersea Warfare Center has fabricated and tested a 2.5 kHz magnetostrictive sonar transducer to validate various modeling techniques. The transducer selected is a longitudinal vibrator Tonpilz type consisting of Terfenol-D driver, tail mass, radiating head mass, and stress rod bolt with 21 MPa (3000 psi) prestress. The Terfenol-D drive rod is interlaced with three samarium cobalt magnets, one in the center and one on either end magnetically biasing the Terfenol to 60 kA/m (750 Oe). Both the Terfenol-D rods and magnets were laminated to reduce eddy currents. The magnetic circuit is comprised of pole piece discs on each end of the Terfenol-D magnet assembly and an external magnetic cylinder (return path) made of a high-permeability, high-resistivity, high-saturation powdered metal ''T2.'' The transducer has a 25 cm (9.8 in) diameter radiating face (piston), is 28 cm (11 in) long, and weighs 15 kg (32 lb) without the housing. It is 41 cm (16 in) long and 25 kg (56 lb) with the underwater housing. The measured results are compared to a finite element model using ''ATILA'' and distributed plan wave element model ''TRN.'' The coupling coefficient, permeability and mechanical losses effects for different prestress loads were measured on a resonant Terfenol ''dumbbell'' device. The in-water measured results indicate a mechanical Q of 2.5, an effective coupling coefficient of 0.36, an electro-acoustic efficiency of 60%, beam pattern directivity index of 6 dB, a maximum source level of 214.6 dB re 1uPa/m at 15 Amps ac drive and bandwidth of 2 kHz to 5.4 kHz $\pm$ 1.5 dB. Harmonic distortion was also measured for various drive levels. [Work supported by ONR.]

3:20–3:35 Break

3:35

4pEA4. Magnetostrictive hybrid transducers. W. Jack Hughes (Appl. Res. Lab., Box 30, State College, PA 16804, wjh2@psu.edu)

The current need in many sonar systems is for a high-power broad-bandwidth projector with a relatively flat response. One of the prime candidates to accomplish this is a hybrid transducer that uses both a piezoelectric and magnetostrictive motor section connected mechanically in series in a tonpilz (longitudinal vibrator) configuration. This transducer has a double resonance separated by about one octave. The inherent shape of the transmit frequency responses dictate using the magnetostrictive section for the lower resonance, and the ceramic section for the upper resonance, including maintaining a high response for another octave. Attention to spurious structural resonances is important. A general overview of developing the device with attention to its design, high-power application, and problem areas will be presented. [Work supported by ONR.]

Contributed Paper

4:00

4pEA5. Performance of a Terfenol-D-driven folded shell projector. Julie Slaughter (Etrema Products, Inc., 2500 N. Loop Dr., Ames, IA 50010) and Christopher Purcell (Defence Res. Establishment Atlantic, Dartmouth, NS B2Y 3Z7, Canada)

The folded shell projector (FSP), U.S. Patent No. 5,805,529, is a compact flextensional transducer originally developed for low-frequency naval sonar applications. The FSP design radiates sound from a one-piece corrugated shell. Because the shell is inherently waterproof, the projector does not need a boot. This reduces cost, improves linearity, increases thermal conductance, and, most importantly, results in stable performance with depth. The folded shell projector is being evaluated for a number of applications including towed array and expendible sonar transducers, community warning sirens, and loudspeakers. A folded shell projector driven by the giant magnetostrictive material, Terfenol-D, has been developed. This transducer was originally designed for use as a community warning siren, however, tests both in air and in water were performed to investigate its response to different acoustic loads. Experimental results of the transducer performance are presented and compared with performance estimates from an equivalent circuit model. Limitations of the current design and the modeling efforts to improve it are discussed as well as plans for future development of the FSP.

Session 4pMUa

Musical Acoustics: 3-D Spatialization for Musical Applications

Gary Kendall, Chair
School of Music, Northwestern University, Evanston, Illinois 60208

Invited Papers

2:45

4pMUa1. Compositional spatial sound manipulation: Historical overview and analyses. Durand R. Begault (NASA Ames Res. Ctr., MS 262-2, Moffett Field, CA 94035, dbegault@mail.arc.nasa.gov)

Since all listening has inherently a spatial aspect, it was inevitable that composers would manipulate spatial acoustic perceptions for their listeners just as they manipulated pitch, timbre and rhythm. Approaches to organizing the spatial component of a musical composition over the evolution of western European art music might be organized in terms of the following techniques: alternation between locations (e.g, polychoral music tradition, call and response), suggestion of environmental contexts and distant sound sources (in both 19th century orchestral music and later, with development of artificial reverberation, in electronic music), and sound source movement. Although not manipulated directly by the composer, performance space acoustics also contribute spatially; data on the temporal-spatial diffusion of reverberation in large performance spaces will be presented. With the development of loudspeaker and headphone spatial sound reproduction techniques, numerous resources and immense possibilities for spatial manipulation have occurred, including the formation of virtual performance spaces. In particular, binaural (3-D sound) signal processing techniques, room modeling auralization and environmental mapping have affected the composition of sonic landscapes in many ways, and will continue to do so.

3:10

4pMUa2. Musical considerations in the design of 3-D-sound rendering software. Gary Kendall, David Mann, Scott Robbin, and Alan Kendall (School of Music, Northwestern Univ., Evanston, IL 60203, g-kendall@northwestern.edu)

The rapid growth of 3-D-sound technology has created a need for 3-D audio tools which can be used by musicians and composers in much the same manner that visual artists use 3-D graphics tools. Experience with such audio tools reveals some issues of special concern for music. For example, the rapid movement of musical sound sources can create Doppler shifts that produce harsh detunings of pitch, and the realistic rendering of intensity loss with changing distance can cause some musical elements to be buried in the mix. Consider too that individual instruments are best spatialized in different environments—high strings in large reverberant halls, electric basses in small dry rooms—and that some musically useful spatial effects, like stereo decorrelation, are not conveniently produced with environmental models. Spatial sound rendering software needs to provide numerous exceptions to accurate physical modeling in order to adapt to the musical context and it must support a heterogeneous collection of musically useful tools. In fact, the true goal in the design of 3D software for music should be to provide a creative tool that captures the complexity and richness of spatial sound without slavishly adhering to models of acoustic reality.

Contributed Papers

3:35

4pMUa3. Does proximity govern judgments of auditory direction in the azimuthal plane? Annabel J. Cohen, M. J. Reina Lamothe, Robert MacIsaac, Richard Fleming (Dept. of Psych., Univ. of Price Edward Island, Charlottetown, PE C1A 4P3, Canada, acohen@upei.ca), and Michael Lamoureux (Univ. of Calgary, Calgary, AB T2N 1N4, Canada)

In an experiment analogous to that of Shepard which showed circularity in judgments of relative pitch of pairs of octave-complex tones [Shepard, J. Acoust. Soc. Am. **36**, 2346–2353 (1964)], 40 listeners, centered in a sound-attenuated room, judged the apparent direction (clockwise, or counterclockwise) of 396 pairs of successive complex tones emanating from all possible pairs of 12 speakers spaced at 30-degree intervals on an imaginary circumference. Judgments were influenced by speaker proximity and in general mirrored the pattern for judgments of up–down direction associated with the phenomena of circular pitch observed by Shepard. In contrast to Shepard's pitch judgment task, musical training did not increase sensitivity in the judgment of auditory spatial direction. Unique to the spatial task was the finding that approximately one-half the subjects showed systematic front–back confusion. When this confusion was taken into account, proximity effects were even more apparent. For diametrically opposed speakers (at 180-degrees absent proximity cues), direction was theoretically ambiguous, but listeners showed a small but significant clockwise bias, reminiscent of Deutsch's tritone paradox for octave-complex tones [Deutsch, Kuyper, and Fisher, Music Percept. **5**, 79–92 (1987)]. [Work supported by NSERC.]

3:50

4pMUa4. Model-based virtual room acoustics. Charles Thompson, Max Dennis, Jing Tsui, and Miroslava Raspopvic (CACT, Univ. of Massachusetts, One University Ave., Lowell, MA 01854)

A new procedure for evaluating the impulse response between source and receiver locations in small reverberant rooms is outlined. The modified image method uses the solution of the Laplace transform based formulation for the Sommerfeld integral to determine the strength of image sources and boundary reflections. The model is used to simulate sound spatialization effects in small rooms. It is shown that spatial invariance of the poles in the room transfer function allows one to use IIR models for the long-time behavior of the impulse response. The variability of the short-time response with changing observation location is to be modeled using FIR models.

4:05

4pMUa5. Directionless surround mixing. Scott Colburn (Gravelvoice, P.O. Box 45066, Seattle, WA 98145, scott@gravelvoice.com)

The first commissioned audio DVD, titled Immersions, contained a piece by Ellen Fullman and her long stringed instrument. The concept of mixing this piece for surround sound was complicated because it could not be rehearsed. It was conceptualized in advance without hearing it. It is tempting to use surround sound (like quad) to fly sounds around in the surround field but this novelty will wear thin after repeated listening. This mix was created to utilize movement within the surround field through the use of phantom images. The mix also has a defined front and back, but remains interesting wherever you sit in the sound field regardless of how you face. It also had to be good in stereo.

Invited Paper

4:20

4pMUa6. There are no rules? Perceptual challenges posed by large-scale, three-dimensional surround-sound systems. Ambrose Field (Dept. of Music, Univ. of York, York YO10 5DD, UK)

With the widespread adoption of multichannel audio gaining pace, sound designers have recognized that surround-sound systems are not simply another effects process that can be applied retroactively to musical material. This discovery has generated complex questions, e.g., where to place instruments, sounds, or dialog in a three-dimensional mix? Rather than attempting to prescribe ways in which composers can physically spatialize their material, this work proposes that by understanding the need to spatialize sound, informed methods of three-dimensional projection can be generated. Ways in which sound designers can conceive their material for surround-sound directly from the outset are examined, as is the premise that there is a need to recognize that each sound within a mix will have its own pre-existing spatial context. The relationship between the initial context in which a recording was made, and how that sound is finally spatialized plays a significant part in how ''believable'' a soundfield is. This different approach to surround mixing is detailed, demonstrating that the traditional practice of sound placement according to what a sound is can be detrimental to establishing convincing and functional surround soundfields. Finally, practical examples are given, relating to and integrating Dolby 5.1 and Ambisonics diffusion techniques.

THURSDAY AFTERNOON, 7 JUNE 2001 CRYSTAL ROOM, 5:00 TO 6:00 P.M.

Session 4pMUb

Musical Acoustics: 3-D Spatialization Concert

Gary Kendall, Chair
School of Music, Northwestern University, Evanston, Illinois 60208

Chair's Introduction—5:00

Musicians have incorporated auditory spatial features in music for many centuries, so the concept of spatial music is not historically new. The advent of multi-track tape recorders after World War II brought forth an intensified interest in music spatialization, especially in 4-channel formats, and a dramatic growth in the possibilities of spatial music. Most recently, the 8-channel ADAT format has inspired new spatial compositions which exploit the possibilities of this new medium, and now the AC-3 5.1 DVD Home Theater System promises to bring spatial music into the home. This concert includes a sampling of diverse creative approaches to musical space within these technical frameworks, featuring works by some leading composers in the musical exploration of spatial sound. From Doppler shift of rapidly moving sound sources to diffuse fields of voluminous sound, recent spatial music can provide some head-spinning effects and intense aesthetic experiences.

Session 4pNS

Noise: International Noise Initiatives

Robert D. Hellweg, Cochair
Compaq Computer Corporation, MR01-3D3, 200 Forest Street, Marlborough, Massachusetts 01752

Gilles A. Daigle, Cochair
Institute of Microstructural Science, National Research Council, Ottawa, Ontario K1A 0R6, Canada

Invited Papers

1:00

4pNS1. The International-INCE Technical Initiatives: An overview. Gilles Daigle (Natl. Res. Council, Ottawa K1A 0R6, Canada)

At Inter-Noise 91 in Sydney, Australia, the International Institute of Noise Control Engineering (I-INCE) Board took the first steps to enlarge the scope of the General Assembly's activities. In particular, the need for technical initiatives to address noise issues of international interest was recognized. Such issues usually involve important policy matters, but the focus is on technical details. The following year in Toronto, Canada, the Board authorized studies on two topics, upper noise limits in working environments and the effect of regulations on road vehicle noise, to be carried out by the General Assembly. Two further initiatives, community noise and noise barriers, were approved in subsequent years. To date, the Technical Initiatives Program has produced two final reports and one draft report. At the 1999 General Assembly in Fort Lauderdale, FL, four new technical initiatives were approved, noise policies and regulations, noise control for schoolrooms, noise from outdoor recreational activities, and noise labels for products. A fifth initiative was added by the General Assembly in Nice, France, on the topic of noise as a global policy issue. This work will give an overview of the objectives and process of the I-INCE Technical Initiatives Program.

1:20

4pNS2. Assessing the effectiveness of national and international noise policies and regulations. Lawrence S. Finegold (Finegold & So, Community & Environ. Noise Consultants, 1167 Bournemouth Court, Ctr.ville, OH 45459) and Paul D. Schomer (USA–CERL–CNN, Champaign, IL 61826)

This I-INCE technical initiative will assess the effectiveness of noise policies and regulations around the world. During the last half of the 20th century, many countries have recognized noise as an environmental and occupational problem, and various countries have been working to develop adequate noise exposure policies and noise control technologies. Considerable time and effort are devoted each year to develop noise exposure policies for places where people work, where people live, and for outdoor environments devoted to leisure activities. However, very little is known about how effective these noise policies and regulations have been in controlling the noise exposure of the populations which they are intended to protect. This policy research initiative involves a study of existing noise exposure policies and regulations in countries which have recognized noise as a problem affecting the public health and welfare. The first phase involves the collection and cataloging of noise exposure policy guidelines, regulations, and standards. The second phase involves developing a baseline of noise exposure estimates for each participating country. The third phase will entail determination of the long-term effectiveness of these policies and regulations in controlling noise exposure by examining the changes in various population noise exposure parameters over time.

1:40

4pNS3. I-INCE Technical Initiative #4: Noise and reverberation control for schoolrooms. David Lubman (David Lubman & Assoc., 14301 Middletown Ln., Westminster, CA 92683, dlubman@ix.netcom.com) and Louis C. Sutherland (Louis C. Sutherland, Consultant in Acoust., Rancho Palos Verdes, CA 90275-3908)

The International-INCE General Assembly approved Technical Initiative #4 in December 1999. The initiative creates an internationally coordinated program intended to assist participating nations with engineering issues associated with achieving satisfactory acoustics in learning spaces. Representatives of various acoustical societies were appointed to a Technical Study Group, convened by Zerhan Karabiber, professor of architecture at Yildiz University in Turkey. The initiative was well received and the group has representatives from 17 acoustical societies in 15 nations. The group is directed to submit its draft report for review by the I-INCE General Assembly by August of 2002. The report is expected to provide practical guidance to school planners, designers, code officials, and policy makers. The group met at the Inter-Noise 2000 in Nice, France. It will meet again at Inter-Noise 2001 at The Hague. An assignment made to all representatives is to assess the state of classroom acoustics and related activities in their own countries. This presentation briefly highlights energetic US activities. These include development of a draft ANSI Standard for classroom acoustics, and development of educational materials (led by the Acoustical Society of America), activities and support by other professional, acoustical, hearing, and governmental organizations, and critizen activism.

2:00

4pNS4. Status of the I-INCE technical initiative on noise from outdoor recreational activities. Nicholas P. Miller (Harris Miller Miller & Hanson, Inc., 15 New England Executive Park, Burlington, MA 01803)

Outdoor recreational activities are increasingly both creating noise and being affected by noise. A study group has been formed and met for the first time at Internoise 2000 in France. The Convener of the study group is Dr. Philip Dickinson from New Zealand, and there are approximately 12 other members who represent various countries around the world. The group is focusing both on the effects on the environment (people as well as wildlife and natural habitats) resulting from noise produced by outdoor leisure activities, as well as on the effects on leisure activities of noise from other activities. The study will consider national parks and wilderness areas, will examine the need to preserve general quiet areas, and will identify the effects on leisure activities of various sources of noise. Members have been gathering information, and the study group expects to identify specific recreational activities, with each member focusing on a particular activity, according to his or her expertise. In this presentation, the author will identify some of the specific technical and policy issues that can arise when developing a quantitative approach to monitoring and analyzing sound environments in park and wilderness areas.

2:20

4pNS5. Status of the I-INCE technical initiative on noise labels for products. Robert D. Hellweg, Jr. (Compaq Computer Corp., MRO1-3/D3, 200 Forest St., Marlborough, MA 01752, bob.hellweg@compaq.com) and Joseph Pope (Pope Eng. Co., Newton Center, MA 02459)

The I-INCE initiative on Noise Labels for Consumer and Industrial Products is a study of the labeling, and other forms of product information, that deals with noise emissions that are furnished to the purchaser of products. This I-INCE initiative will survey current methods for labeling and otherwise characterizing the noise emissions of such products. (The measurement methods used to test the products will also be included in the survey.) Such labeling methodologies are intended to provide effective means for specifying the noise properties of consumer and industrial products. It is expected that this information will benefit both the users of these products and others exposed to the noise emissions of the products, such as neighbors and bystanders. The methodologies will be compared, and an assessment will be made of their relative effectiveness. An important outcome of this study will be recommendations on how, and in what form, noise labeling can be implemented. The Study Group first met in Nice, France in August 2000. Actions to date and future plans are presented. Opportunities for ASA inputs are described including the status of draft American National Standard on Noise Labeling, currently under ballot by American Standards Committee S12.

2:40

4pNS6. US versus international environmental noise policies—a proposed agenda for the next decade. Lawrence S. Finegold (Finegold & So, Community & Environ. Noise Consultants, 1167 Bournemouth Court, Centerville, OH 45459)

During the past decade, there has been significant efforts in both Europe and Asia to develop new environmental noise control policies. During the same time in the US, there have been few major improvements in our national environmental noise policies, coupled with decreased levels of funding for noise research. The improvements in environmental noise policies being seen in Europe and Asia are due to an increasing awareness of the negative impacts of these exposures on the overall health and welfare of the exposed populations, and an awareness of the costs to society for these exposures. In particular, the rapidly evolving set of European Commission Noise Directives and the World Health Organization ''Guidelines for Community Noise'' represent a new, more aggressive philosophy concerning the steps required to minimize environmental noise exposure. In addition, the European Commission is adopting stringent noise emission criteria for a wide range of products, which could severely hinder US efforts to compete in the European marketplace. In response to this high level of international activity, it is proposed that the US embark on a 10-year effort to revitalize our national environmental noise policies. A description of a proposed philosophy for this program is presented.

3:00–4:00

Panel Discussion

Session 4pPP

Psychological and Physiological Acoustics: Masking, Modulation and Information Processing (Poster Session)

Stanley E. Sheft, Chair
Parmly Hearing Institute, Loyola University of Chicago, 6525 North Seridan Road, Chicago, Illinois 60626

Contributed Papers

All posters will be on display from 1:30 p.m. to 5:00 p.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 1:30 p.m. to 3:15 p.m. and contributors of even-numbered papers will be at their posters from 3:15 p.m. to 5:00 p.m.

4pPP1. Psychometric functions for detection of tones following on- and off-frequency forward maskers. Kim S. Schairer, Jason F. Reimer, and Walt Jesteadt (Boys Town Natl. Res. Hospital, 555 N. 30th St., Omaha, NE 68131, schairerk@boystown.org)

Recent data obtained in our laboratory suggest that the slope of the psychometric function in forward masking can vary widely across conditions, with shallower slopes occurring in conditions with higher signal levels at threshold. The goal of the present study was to separate possible effects of masker frequency, masker level, and signal delay from the effect of signal level per se. Thresholds and psychometric functions were obtained for three normal-hearing listeners using a 2IFC adaptive task. The signal was a 4-kHz, 10-ms tone. The on-frequency masker was a 200-ms tone presented at 30- to 90-dB SPL in 10-dB steps. The off-frequency masker was a 2.4-kHz, 200-ms tone presented at 60- to 90-dB SPL in 10-dB steps. Signal delays were 0, 10, or 30 ms. As expected, forward masking grew more rapidly with increases in masker level in the off-frequency than the on-frequency conditions. Signal delay had little independent effect on the slopes of the psychometric functions. The effects of masker level and signal level were similar for on- and off-frequency masker conditions in most cases. The results will be discussed in the context of a model that incorporates peripheral nonlinearity and internal noise. [Work supported by NIH.]

4pPP2. Growth of masker level with signal level as a function of preceding stimulation. Elizabeth A. Strickland (Dept. of Audiol. and Speech Sci., Purdue Univ., West Lafayette, IN 47907)

Some recent data suggest that stimulating the auditory system with sound may turn down active processing. The basic paradigm used to show this is to measure the level of a notched-noise, long-duration masker required to just mask a short-duration tone that occurs near masker onset. The preceding stimulation has been a precursor that was either a fixed-level broadband noise [Strickland, J. Acoust. Soc. Am. (in press)], or an extension of the notched-noise masker, which varied in level [Strickland, J. Acoust. Soc. Am. **107**, 2915 (2000)]. The present experiment was done to examine the effects of a notched-noise precursor that had the same frequency characteristics as the masker, but which would be fixed in level. The signal was a 10-ms, 4-kHz tone. The notched-noise masker was 200 ms, and the precursor was 205 ms. For a given precursor level, the growth of masker level with signal level was determined. This data was used to calculate the effective input–output functions of the signal and the masker. These functions will be compared to those published physiological measures from the chinchilla cochlea. [Work supported by NIH (NIDCD).]

4pPP3. Intensity discrimination and increment detection in electric hearing. Magdalena Wojtczak, Gail S. Donaldson, and Neal F. Viemeister (Univ. of Minnesota, Minneapolis, MN 55455)

Intensity DLs were measured in listeners with electric and normal hearing using increments that were gated synchronously with the pedestals and increments whose onsets were delayed relative to the onset of the pedestal (fringe condition). Two biphasic pulse durations (80 and 200 μs/phase) and three increment durations (20, 100, and 500 ms) were used. Normal-hearing listeners were tested with tonal and noise stimuli using an increment duration of 100 ms. It was found that, for short increments, the DLs of cochlear-implant users exhibit a different pattern as a function of pedestal level than the DLs observed for tonal stimuli in acoustic hearing [R. P. Carlyon and B. C. J. Moore, J. Acoust. Soc. Am. **79**, 453–460 (1986)]. For longer-duration increments, the ''gated-continuous'' difference observed in normal hearing for tones and noise stimuli was not consistently observed in electric-hearing data. For cochlear-implant users, intensity DLs measured in the gated versus fringe conditions did not differ substantially for a fringe duration that produced a difference in thresholds of normal-hearing subjects. The possible site of the mechanism mediating the gated-continuous difference in normal hearing will be discussed. [Work supported by NIH Grant P01-DCD00110 and Lions 5M Hearing Foundation.]

4pPP4. Children's detection of tonal signals in roving level noise maskers. Prudence Allen and Suzanne Turpin (Natl. Ctr. for Audiol., Univ. of Western Ontario, 1500 Elborn College, London, ON N6G 1H1, Canada, pallen@uwo.ca)

Preschool-aged children's masked detection thresholds have been measured at S/N levels for which the level of the signal plus masker exceeds that of the masker alone by approximately 6 dB and for which the temporal fine structure of the masked signal approximates that of the signal alone [Allen *et al.*, J. Acoust. Soc. Am. **104**, 2997–3005 (1998); P. Allen and L. Korpela, *ibid.* **105**, 1152(A) (1999)]. When level information is removed through randomization of masker levels, children could rely on temporal fine structure changes for detection. This study asked children and adults to detect 500-, 2000-, and 4000-Hz tonal signals presented in roving level maskers centered at 10- or 25-dB N0. Thresholds for some children remained consistent with previous data in fixed-level conditions at all frequencies suggesting that temporal fine structure information may have been used by them, even at frequencies where its coding is subject to phase locking limitations. However, there were large individual differences with many children performing very poorly. [Work supported by NSERC.]

4pPP5. A comparison of auditory suppression in older and younger normal-hearing adults. Sara Elizabeth Gehr and Mitchell Sommers (Dept. of Psych., Washington Univ., One Brookings Dr., Campus Box 1125, St. Louis, MO 63130, segehr@artsci.wustl.edu)

The present study was designed to assess the effect of age, independent of hearing loss, on auditory suppression. Two-tone suppression in a forward masking paradigm was measured for normal-hearing older (65–75 years old) and normal-hearing younger (19–24 years old) adults. Thresholds for detecting a 2000 Hz sinusoid were measured in conditions where the forward masker either contained one (nonsuppressive) or two (suppressive) tones. Young adults showed reduced thresholds in the suppressive, compared with the nonsuppressive condition. In contrast, thresholds in the suppressive and nonsuppressive conditions did not differ for older adults. This lack of a difference between the suppressive and nonsuppressive conditions for the older adults indicates that they have reduced suppression compared with young adults. Implications for the mechanisms mediating auditory suppression are discussed. [Work supported by American Psychological Association.]

4pPP6. Duration sensation when listening to complex tones. Saifuddin Kazi, Matsushima Takaji, and Ando Yoichi (Grad. School of Sci. and Technol., Kobe Univ., Rokkodai, Nada, Kobe 657-8501, Japan)

The duration sensation (DS) is defined as one of the primary sensations of the sound signal [Y. Ando, J. Sound Vib. **240**(4), to be published on 1 March 2001]. The DS of complex-tone stimuli was examined here by the paired-comparison test while changing the residue pitch or fundamental frequency, i.e., the delay time of the first peak (FP) extracted from the @ autocorrelation function. The white noise was used as a reference and complex tones (two pure-tone components) with three fundamental frequencies, 250, 500 and 1000 Hz, used as test stimuli. Ten subjects judged whether or not the white-noise duration is longer than the complex tone duration under constant 80 dB (A). The rise and fall times defined by the time at a -3 dB drop from the steady level were 1 ms constant. Previous results indicate that the DSs of the pure tone and bandpass-noise stimuli judged longer with the FP longer than that of the stimuli with shorter FP. The purpose of this investigation is to examine whether or not the DS of the fundamental frequency of complex tones is similar to the DS of the pure tone. Experimental results will be presented.

4pPP7. Do listeners include the echo portion of a damped sound when judging its duration? Robert S. Schlauch, Jeffrey J. DiGiovanni, Kristi L. Olson, and Erin E. Donlin (Dept. of Commun. Disord., Univ. of Minnesota, Minneapolis, MN 55455)

Ramped sounds (slow attack, fast decay) are perceived to be longer in duration than damped sounds (fast attack, slow decay) that are the same physical duration even when they are simply time-reversed versions of each other. Stecker and Hafter [J. Acoust. Soc. Am. **107**, 3358–3368 (2000)] suggest that this perceptual difference is a result of listeners ignoring a portion of the decay of damped sounds because they consider it an ''echo.'' To test this idea, a duration-matching experiment with two groups of 10 listeners was completed. Each group matched the duration of a broadband noise to two natural sounds, a drum strike and a word spoken in a reverberant room, presented in the normal fashion and reversed in time. The first group was told to match the duration of the sounds. The second group was told to consider the entire duration of the sound, even portions that sound like an echo. The ratio of ramped/damped at equal subjective duration was larger for the first group than for the second group. This finding supports Stecker and Hafter's idea that listeners ignore the echo portion of the sound, but it did not account for the entire effect. [Work supported by NIH.]

4pPP8. Perception of periodicity strength in chinchillas. William P. Shofner (Parmly Hearing Inst., Loyola Univ. Chicago, 6525 N. Sheridan Rd., Chicago, IL 60626, wshofne@luc.edu)

In stimulus generalization paradigms, an animal is trained to respond to a particular stimulus, and behavioral responses are then measured to probe stimuli that vary systematically along a given stimulus dimension. Unlike discrimination tasks, responses to probe stimuli are not reinforced in generalization paradigms. A generalization gradient suggests how closely the probe stimuli are perceived by the animal to resemble the training stimulus and is consistent with the hypothesis that the animal possesses a perceptual dimension related to the stimulus dimension. Chinchillas were trained to discriminate a cosine-phase harmonic tone complex from wideband noise. Probe stimuli consisted of random-phase harmonic tone complexes and infinitely iterated rippled noises. These stimuli vary in periodicity strength as measured by autocorrelation and generate a continuum in the perception of pitch strength in human listeners. Generalization gradients obtained from chinchillas are related to the height of the first peak in stimulus autocorrelation functions, suggesting that chincillas possess a perceptual dimension related to stimulus periodicity strength. Comparison of generalization data to magnitude estimation data suggests a greater influence of stimulus envelope for the perception of periodicity strength in chinchillas than for the perception of pitch strength in human listeners. [Work supported by NIDCD P01 DC00293.]

4pPP9. Spectral modulation transfer functions as a function of duration and level. David A. Eddins and Ross Harwell (Psychoacoustics Lab., Ctr. for Hearing & Deafness, Univ. at Buffalo, Buffalo, NY 14214)

Spectral modulation transfer functions (SMTF) were estimated as a function of stimulus duration and level to evaluate the influence of those parameters on spectral shape perception. Listeners discriminated between noise stimuli having either a flat or a sinusoidal spectral shape. Spectral contrast thresholds were determined by adaptively varying the peak-to-valley contrast using a 3-down, 1-up tracking algorithm in a cued, two-interval, forced-choice procedure. Carrier noises had bandwidths of either 6 octaves (200–12 800 Hz), or 1 octave (800–1600 Hz, 6400–12 800 Hz). Spectral frequencies ranged from 0.25 to 10 cyc/oct and 1.33 to 10 cyc/oct for the 6- and 1-octave conditions, respectively. In Experiment I, the stimulus durations were either 50, 100, 200, 300, or 400 ms. SMTFs were bandpass in shape with best contrast thresholds between 2 and 5 cyc/oct. Decreasing duration had a greater influence on thresholds for low- and high-spectral frequencies than mid-spectral frequencies. In Experiment II, stimulus level was varied in spectral levels from –10 to 50 dB SPL. The shape of the SMTF varied little with level down to No = 10 dB. Decreasing level had a greater influence on low- and high-spectral frequencies than mid-spectral frequencies. [Work supported by NIH NIDCD RO1DC04403.]

4pPP10. Masking patterns for simple spectral stimuli. Aniket Saoji and David A. Eddins (Psychoacoustics Lab., Ctr. for Hearing & Deafness, Univ. at Buffalo, Buffalo, NY 14214)

Auditory processing appears to consist of a series of domain-specific filtering operations that separate complex stimuli into different auditory channels. This begins with tuning in the audio-frequency domain, followed by tuning in the temporal frequency (modulation) domain, and perhaps tuning in the spectral frequency domain. To explore the hypothesis that envelope channels tuned to spectral frequency exist in the auditory system, a masking experiment was designed to measure masking patterns in the spectral domain. Noise carriers with flat spectra were either 1 (800–

1600 Hz, 6400–12 800 Hz) or 6 octaves wide (200–12 800 Hz) and were shaped with sinusoidal spectral envelopes (in cycles per octave). Masker frequencies were 1, 3, or 5 cyc/oct with a spectral contrast of 20 dB. Masked contrast thresholds were measured for signal frequencies ranging from 1.5 octaves below to 2 octaves above the masker frequency. The resulting masking patterns revealed maximum masking at the masker frequency and a gradual decrease in the masking above and below the masker frequency. This tuning is consistent with the assumption of spectral envelope channels tuned to different spectral frequencies. Secondary peaks obtained in several masking patterns may suggest inhibitory interactions among separate spectral channels. [Work supported by NIH NIDCD RO1DC04403.]

4pPP11. Detection of gaps marked by frequency-asymmetric tonal complexes. John H. Grose, Joseph W. HallIII, and Emily Buss (Dept. of Otolaryngol./HNS, 610 Burnett-Womack, CB# 7070, Univ. of North Carolina, Chapel Hill, NC 27599-7070, jhg@med.unc.edu)

When the frequency content of two acoustic markers bounding a temporal gap is asymmetric across the gap, then sensitivity to the gap depends on the order in which the markers are presented [Formby *et al.*, J. Acoust. Soc. Am. **104**, 984–998 (1998)]. Specifically, when one marker consists of a single tone while the other marker consists of that same tone plus a second synchronous tone, then performance is better when the leading marker is the two-tone complex. Analogous results have been found for cochlear implant listeners when the number of electrode channels marking the gap is asymmetric across the gap [van Wieringen and Wouters, J. Acoust. Soc. Am. **106**, 1925–1939 (1999)]. The present study extended this basic finding to conditions where the synchronous second tone within the two-tone marker was presented either diotically, or to the ear contralateral to that receiving the remaining tone (and the single-tone marker). Temporal asynchronies were also introduced into the two-tone marker in an effort to weaken the perceptual fusion of the tone pair. Results from six normal-hearing listeners indicated that the basic order effect was present for the monaural, diotic and dichotic conditions. [Supported by NIH NIDCD R01-DC01507.]

4pPP12. Effects of age and hearing loss on discimination of silent gaps in dynamic stimuli. Jennifer Lister, Kenton Tarver, and D'Arcy Cyr-Kriley (Dept. of Commun. Sci. and Disord., Univ. of South Florida, 4202 East Fowler Ave., BEH 255, Tampa, FL 33620, jlister@chuma1.cas.usf.edu)

Temporal cues in speech occur between signals that are dynamic (changing in frequency and duration). Older listeners with and without hearing loss often experience difficulty discriminating these temporal cues important for speech understanding. This study was designed to measure silent gap discrimination between dynamic stimuli for three groups of listeners: (1) young with normal hearing; (2) older with normal hearing; and (3) older with sensorineural hearing loss. Synthetic stimuli were designed to simulate four speechlike sounds (/s/, /a/, /eI/, /b/) arranged in six pairs separated by a silent gap (/s-a/, /s-eI/, /a-a/, /eI-eI/, /b-a/, /b-eI/). Two conditions of duration were imposed upon the six speech sound pairs: (1) duration of each marker fixed at values typical for the speech sounds simulated and (2) duration of each marker varied randomly within a range appropriate for each speech sound simulated. The task was a 2I/2AFC procedure targeting 70.7% correct discrimination. Results indicate that gap discrimination for frequency dynamic markers deteriorates with age and is poorer for random duration conditions than for fixed duration conditions. Significant correlations were found between age and gap discrimination and between hearing sensitivity and gap discrimination. [Work supported by USF Research and Creative Scholarship Grant No. 1219-933-RO.]

4pPP13. Modulation masking produced by masker modulation tone complexes. Jesko L. Verhey, Stephan Ewert, and Torsten Dau (AG Medizinische Physik, Universität Oldenburg, D-26111 Oldenburg, Germany)

Thresholds were measured for detecting a sinusoidal amplitude modulation in the presence of a masker-modulation tone complex. Both modulations were applied to the same sinusoidal carrier. Two different masker modulations were used: (i) a pair of components beating at the signal-modulation frequency and (ii) a three-tone complex producing a sinusoidal amplitude modulation in the modulation domain whereby the frequency of the slow amplitude variation was equal to the signal-modulation frequency. Masked thresholds were measured as function of the phase of the signal modulation relative to the slow amplitude variation in the masker modulation. In all conditions, thresholds were lower for an in-phase signal modulation compared to an antiphase condition. The maximum threshold difference was 15 dB. The results are in contrast to recent data [B. C. J. Moore *et al.*, J. Acoust. Soc. Am. **106**, 908–918 (1999)], where the lowest thresholds were obtained for the antiphase condition. Our data are in line with the idea that a nonlinearity prior to the modulation filterbank extracts the slow variation of the masker modulator. The data could be fitted well by assuming a combination of a peripheral (power-law) compression and a threshold. Peripheral compression alone did not account for the data.

4pPP14. Asymmetry of masking in the envelope-frequency domain. Stephan Ewert, Torsten Dau, Jesko L. Verhey, and Birger Kollmeier (Carl von Ossietzky Universitt Oldenburg, AG Medizinische Physik, D-26111 Oldenburg, Germany)

Masked modulation thresholds can be influenced by the detection of amplitude fluctuations of the envelope (referred to as the venelope in the following) which arise from either beating between the masker modulation and the probe modulation, or from the intrinsic envelope fluctuations of a narrowband-noise modulator itself. It was tested to see to what extent the detection of venelope fluctuations is involved in corresponding tone-in-noise, noise-in-tone, and tone-in-tone modulation masking conditions. Modulation frequencies in the range from 20–100 Hz, applied to a 2.8-kHz sinusoidal carrier, were used. Thresholds in the noise-in-tone condition are always lower than those in the tone-in-noise condition, comparable with the asymmetry of the masking effect in the audio-frequency domain. In general, it was observed that (1) subjects use venelope fluctuations as a detection cue, and that (2) venelope fluctuations interfere with an additionally applied amplitude modulation when both fall into the same frequency range. To interpret the empirical findings, a general model structure for the processing of envelope and venelope fluctuations is proposed. Similarities between the auditory system's coding strategies for audio frequencies/envelope fluctuations and envelope frequencies/venelope fluctuations are discussed.

4pPP15. Phase effects in masking in normal-hearing and hearing-impaired listeners. Torsten Dau (Med. Physik, Univ. of Oldenburg, 26111 Oldenburg, Germany) and Andrew J. Oxenham (Res. Lab. of Electron., MIT, Cambridge, MA 02139)

Changes in the phase relationships between components in a harmonic tone complex masker can lead to large changes in masked threshold for long-duration sinusoids in normal-hearing listeners. The effect is reduced or absent in hearing-impaired listeners. This reduction could be due to a loss of peripheral compression, a change in the auditory filters' phase response, or both. Signal thresholds were measured as a function of masker phase curvature, signal frequency, and masker fundamental frequency (F0). In contrast to previous studies, more than just two phase relationships were considered, allowing us to compare the overall phase response of impaired and normal ears. Peripheral compression was estimated in a separate experiment using a notched-noise masking technique. If changes in filter phase response can account for the differences between normal-hearing and hearing-impaired listeners, then certain phase relationships should produce large threshold differences even in hearing-impaired

listeners. If peripheral compression is important, threshold differences in hearing-impaired listeners should emerge at very low F0s, due to the predicted interaction between peripheral compression and temporal processing at higher stages. Preliminary results suggest that the difference between normal-hearing and hearing-impaired listeners can be explained primarily in terms of peripheral compression. [Work supported by DFG and NIDCD.]

4pPP16. Auditory abilities of experienced signal analysts. Stanley Sheft, William A. Yost (Parmly Hearing Inst., Loyola Univ., 6525 N. Sheridan Rd., Chicago, IL 60626, ssheft@luc.edu), and Michael Weir (AFRL/IFEC, Rome, NY 13440)

Signal analysts continuously monitor audio communication signals for classification and identification. Performance levels of experienced analysts generally exceed that of most automated signal-classification schemes. Auditory abilities of four experienced military signal analysts were evaluated with both basic psychoacoustic listening tasks and multidimensional scaling of utility signals. The goal of the study was to determine the aspects of analysts' auditory abilities relevant to job performance. Listening tasks were frequency discrimination, AM detection and rate discrimination, rippled-noise detection, delay discrimination, and detection of delay modulation. Across tasks, analysts' thresholds were generally within the range of normative data. Results did suggest that frequency discrimination for both simple and complex stimuli is not a crucial ability for signal analysts. Stimuli used in the scaling study were first processed through a six-channel filterbank with both envelope- and phase-modulation functions extracted from each channel. Modulation functions were then applied to narrow-band carries. For some stimuli, functions were filtered or transposed across channel center frequencies. Multidimensional scaling of processed utility signals indicated that signal classification depends on both lower-rate amplitude and phase modulation. This result may have some bearing on the design of algorithms used in automated signal classification. [Work supported by AFRL.]

4pPP17. ''Second order'' motion in the auditory system. Wendy E. Schneider and Edgar A. DeYoe (Dept. of Cell Biol., Neurobiology and Anatomy, Medical College of Wisconsin, 8701 Watertown Plank Rd., Milwaukee, WI 53226)

In both vision and audition the translation of activity across a receptotopic set of neurons (e.g., ganglion, auditory nerve fibers) produces a percept of motion in physical space (vision) or in frequency space (audition). In vision, the translating ''activity'' can consist of either a net increase or decrease in activation (1st order motion), or a local transient that, on average, results in no net change. The purpose of the present study was to determine if such ''2nd order'' motion could also be perceived in audition. Subjects were asked to listen to a stimulus consisting of 16 discrete frequencies (87–6654 Hz). Initially, each frequency was randomly assigned either high or low amplitude. Then, in either ascending or descending order, each successive frequency was ''flipped'' to the alternative amplitude (e.g., low to high or high to low) with various stimulus onset asynchronies (SOA). On each trial subjects indicated if the ''transients'' were ascending or descending the frequency spectrum. This task proved to be more difficult than a simple frequency shift paradigm (1st order motion), but subjects were able to perform successfully at sufficient SOAs, consistent with a percept of 2nd order motion.

4pPP18. Sample discrimination of target-tone duration with variable-duration context tones. Donna L. Neff and Jason F. Reimer (Boys Town Nat'l. Res. Hospital, 555 N. 30th St., Omaha, NE 68131)

This study examined the ability of three normal-hearing listeners to discriminate duration differences in target tones with and without variable-duration context tones in frequency regions remote from the target. Single target tones at 500 (''low''), 1260 (''middle''), or 3176 Hz (''high'') were combined with one or two context tones in nontarget regions (e.g., a low target paired with a high context). Across intervals in the 2AFC task, the target tone was drawn from Gaussian distributions having a mean duration of 100 or 120 ms. Listeners were to select the target drawn from the distribution with the longer mean duration. Context-tone duration was sampled from this ''correct'' distribution, so duration alone did not cue the targets. Overall performance and perceptual weights for individual tones were examined. Frequency had no effect on performance for targets alone, however, there were interactive effects of target and context frequency. The worst performance was observed for middle targets with flanking context tones. Flanking context tones were effective in combination, but relatively ineffective individually. For low and high targets, the context tone nearer the target frequency was most effective, but both influenced performance. Comparisons to similar conditions with frequency sample discrimination will be discussed. [Work supported by NIH.]

4pPP19. Informational masking as a function of masker set size. Virginia M. Richards and Zhongzhou Tang (Dept. of Psych., Univ. of Pennsylvania, Philadelphia, PA 19104)

Thresholds for the detection of a 1000-Hz tone added to informational maskers were measured using six-tone maskers. A large number of different maskers were generated by randomly choosing the frequencies of the six tones from a (log) uniform distribution ranging from 100 to 6000 Hz, exempting the 160-Hz frequency range surrounding the 1000-Hz signal. These maskers were then assigned to different masker sets, with the number of maskers in any one set being 3, 6, 12, or 24. Thresholds were measured separately for the different masker sets. Unlike the method used by Wright and Saberi [J. Acoust. Soc. Am. **105**, 1765–1775 (1999)], on each interval the masker was drawn with replacement. For small set sizes, the trials could be segregated into those in which the maskers were different across intervals and those in which the maskers were the same across intervals. The resulting thresholds did not depend on the masker composition in the two intervals. As the set size increases, the observers ability to remember/identify the maskers is diminished but the average thresholds are not obviously altered by changes in set size beyond 3. [Work supported by NIH.]

4pPP20. One factor underlies individual differences in informational masking. Robert A. Lutfi, Doris J. Kistler, Eunmi Oh, Michael Callahan, Frederic L. Wightman, and Joshua M. Alexander (Waisman Ctr., Univ. of Wisconsin, 1500 Highland Ave., Madison, WI 53706, lutfi@waisman.wisc.edu)

Masked threshold for a signal can be substantially elevated whenever the listener is uncertain about the spectral or temporal properties of the masker, an effect referred to as informational masking. Accounts often invoke multiple perceptual factors as listeners differ widely in the amount of informational masking they will exhibit under identical conditions. The present study reports results from a principal components analysis of the masking data from 76 listeners in four different studies. The studies included preschool children, adolescents, and adults, and involved only minor variations in stimuli and procedure. The premise underlying the analysis is that if different factors are involved, they should add largely independent sources of variance to the masking data. Hence, more than one principal component should be required to account for a substantial proportion of the variance in the data. The results, instead, supported the operation of a single underlying factor with all but 17% of the variance accounted for by the first principal component. The results are taken to suggest that individual differences reflect differences in the extent to which masker uncertainty adds variance to an otherwise optimal decision variable [R. A. Lutfi, J. Acoust. Soc. Am. **94**, 748–758 (1993)]. [Research supported by NIDCD and NICHD.]

4pPP21. Using reproducible noise to illuminate the nature of informational masking. Robert H. Gilkey (Dept. of Psych., Wright State Univ., Dayton, OH 45435; Air Force Res. Lab., WPAFB, OH; and HRC, Boston Univ., Boston, MA, gilk@wright.edu), Christine Mason, and Gerald Kidd (Dept. of Commun. Disord. and Hearing Res. Ctr., Boston, MA 02132)

Informational masking has been distinguished from energetic masking in that it seems to depend less on the specific characteristics of the stimulus presented on a particular trial (e.g., energy in the critical band) and more on the variability in the ensemble of stimuli presented across trials. We employ techniques that have been used in reproducible noise masking experiments [D. M. Green, Psychol. Rev. **71** (1964)] to examine the degree to which subjects responses in informational masking tasks are driven by the characteristics of the individual maskers. A single-interval task was used to collect hit and false–alarm rates to individual maskers under multiple-burst-same (MBS) and multiple-burst-different (MBD) conditions [G. Kidd *et al.*, J. Acoust. Soc. Am. **95** (1994)]. During each block the 25 MBS or 25 MBD reproducible maskers were presented in random order across trials. Both hit and false–alarm rates showed significant dependence on the masker waveform under both MBS and MBD conditions ($p<0.001$). Preliminary modeling efforts suggest that the subjects responses are at least weakly related to the energy near the signal frequency in the individual masker-alone and masker-plus-signal waveforms. [Work supported by NIH and OBR.]

4pPP22. Increasing informational masking by contralateral stimulation. Gerald Kidd, Jr., Christine R. Mason, Tanya L. Arbogast (Dept. of Commun. Disord. and Hearing Res. Ctr., Boston Univ., 635 Commonwealth Ave., Boston, MA 02215 gkidd@bu.edu), Douglas S. Brungart (Air Force Res. Lab., AFRL/HECB, WPAFB, OH 45433), and Brian D. Simpson (Veridian, Dayton, OH 45431)

Although informational masking is thought to reflect central mechanisms, substantial reductions in informational masking generally occur when the target and masker are presented to different ears. Recently, Brungart and Simpson [J. Acoust. Soc. Am.] found that the intelligibility of a target phrase in a monaural two-talker stimulus was severely degraded by the addition of an unrelated speech signal to the contralateral ear. Their results suggest that the presence of an informational masker in the unattended ear can increase informational masking in the target ear. In this study parallel experiments using nonspeech informational maskers determined whether similar results are obtained in nonspeech detection tasks. The stimuli were sequences of multitone bursts arranged in spectrotemporal patterns that produce very large or very small amounts of informational masking. The signal was a sequence of 1000-Hz tone bursts played synchronously with the maskers. Combinations of effective and ineffective informational maskers were presented in the ear ipsilateral to and contralateral to the signal. Control conditions (masker ipsilateral/contralateral only) established baseline performance. The results supported the earlier findings for multiple-talker conditions. This suggests that the phenomenon is not restricted to speech and may reflect more general perceptual processes accessible through contralateral stimulation. [Supported by NIDCD and AFOSR.]

4pPP23. An interactive similarity rating program for large timbre sets. Stephen Lakatos (Dept. of Psych., Washington State Univ., 14204 NE Salmon Creek Ave., Vancouver, WA 98686), Gary Scavone (Stanford Univ., Stanford, CA 94305), Perry Cook (Princeton Univ., Princeton, NJ 08544), and Colin Harbke (Washington State Univ., Pullman, WA 99164)

A Linux-based program is described for collecting similarity data for large numbers of sounds in order to train an automated audio classifier. Most multidimensional-scaling (MDS) programs require similarity ratings for all pairwise comparisons in order to compute an optimal low-dimensional space, thereby discouraging their use for timbre sets comprising more than 25 or 30 sounds. The current program permits the testing of much larger stimulus sets (e.g., 150–200 sounds) through two principal innovations. First, the program initially assumes that the optimal perceptual space is two dimensional, and adds additional dimensions only if they are warranted. Second, it provides flexibility to listeners by allowing them to choose stimuli for comparison and to indicate their relative similarity in a highly interactive graphical format. The program has been used in listening experiments to study the perceptual criteria listeners apply in their timbral ratings for real-world sounds, and to assess the role that mental imagery may play in such ratings. [Research supported by Air Force Grant No. F49620-99-1-0293.]

4pPP24. Consonant recognition in noise with temporal cues. III. Effects of temporal envelope enhancement on identification thresholds. Frederic Apoux (LPE, UMR CNRS 8581, Univ. Paris V, 71 Av E. Vaillant, 92774 Boulogne-Billancourt Cedex, France and INTRASON France S.A., apoux@psycho.univ-paris5.fr), Stephane Garnier (Groupement d'audioprothsistes ENTENDRE—GIPA 2, 78760 Jouars-Pontchartrain, France), and Christian Lorenzi (Inst. de Psych., Univ. Paris V, 92774 Boulogne-Billancourt Cedex, France)

In order to test the effect of temporal envelope enhancement on speech perception, masked identification thresholds of aCa stimuli (C =p,t,k,b,d,g,f,s,ch,v,z,j) were measured in a two-interval, four-alternative forced choice procedure, using two randomly interleaved two-down/one-up, adaptive staircases. aCa stimuli were either temporally expanded or left intact. The expanded stimuli were obtained by multiplying the original signal by the original envelope at each corresponding point in time. The masker was a gated, steady speech-shaped noise. Temporal expansion was not applied to noise. Thirty normal-hearing listeners (19–28 years) and 42 hearing-impaired listeners (48–79 years) participated in this experiment. Overall, the results showed that thresholds were slightly lower when the envelope was expanded rather than left intact. Moreover, a detailed analysis of the data revealed that only one consonant was adversely affected by envelope expansion (+2 and +3 dB) in both groups. In normal-hearing listeners, a 2-, 4-, and 5-dB improvement was found for consonants f, k, and j, respectively. In hearing-impaired listeners, the results showed that identification thresholds were improved by 3.5 dB for voiced consonants. In this group, the identification thresholds of five consonants (t,b,d,g,z) were substantially lowered by expansion by 2 to 9 dB.

4pPP25. Study of speech and hearing at Bell Telephone Laboratories: The Fletcher years. Christine M. Rankovic (Articulation, Inc., 36 Hampshire St., 2nd Fl., Cambridge, MA 02139)

During the first half of the twentieth century, Bell Telephone Laboratories carried out an extensive research program aimed at improving the intelligibility of speech transmitted over the telephone. Harvey Fletcher, a founding member and the first president of the Acoustical Society, led this effort. The contributions of Fletcher and his colleagues helped produce not only the world's leading telephone system, but also established the speech and hearing sciences field. A compact disk (CD/ROM) containing nearly 10 000 pages of recently discovered Bell Labs internal research documents dating from this era (1917–1933) is now available from the Acoustical Society. The majority of the documents are concise scientific reports that reveal the extraordinary thinking and insights of these world-class researchers. In addition to their historical significance, the documents also provide technical insight and clarity on topics of current interest to speech and hearing scientists. The discussion will include an overview of CD/ROM contents and a focused discussion on documents pertaining to speech intelligibility measurement and prediction.

4pPP26. Subjective listening tests by reference CD based on sound factors. Ki Duk Kim (Quality and Reliability Lab., Daewoo Electron. Co., Ltd., 412-2, Chongchun-Dong, Pupyung-Ku, Inchon, Korea, soundexpert@korea.com), Chul Whan Kim, Hoon Ki Choi, and Dong Su Ryu (Quality and Reliability Lab., Daewoo Electron. Co., Ltd., Inchon, Korea)

The audio analysis of sound reproduction equipment, such as loudspeakers, amplifiers, and so on, is composed of overall measurement (and analysis) of electric signals and acoustic signals. And the subjective listening test has been admitted as being a general measurement method also, because audio test equipment has limits of measurement and the human ear has a superior sensing ability. But the subjective listening test method varies according to personal taste and to the degree of skill for the sound. In addition, it varies according to the person's mental state. So, there has been argument about the objectivity of it. This paper proposes that sound factors (parameters) should be classified into eight categories according to the specific character of the loudspeakers and the sense of hearing. We manufactured a test reference CD for an estimate.

Session 4pSAa

Structural Acoustics and Vibration and Physical Acoustics: David G. Crighton Memorial Session I

Gideon Maidanik, Cochair
Carderock Division, Naval Undersea Warfare Center, David Taylor Model Basin, 9500 MacArthur Boulevard, West Bethesda, Maryland 20817-5000

Paul E. Barbone, Cochair
Department of Aerospace and Mechanical Engineering, Boston University, 110 Cummington Street, Boston, Massachusetts 02215

Chair's Introduction—1:00

Invited Papers

1:05

4pSAa1. The application of complex layer analysis to the vibrating fluid-loaded elastic plate. David Feit and Daniel T. DiPerna (CDNSWC, 9500 MacArthur Blvd., West Bethesda, MD 20817, feitd@nswccd.navy.mil)

Structural acoustics' canonical problem—the locally driven fluid-loaded elastic plate attracted David Crighton's attention over much of his career, and was the subject of his 1998 Rayleigh Medal Lecture to the British Institute of Acoustics. During this lecture he showed it ''to contain a great richness of effects.'' This presentation further elucidates the problem using an approach we refer to as complex layer analysis (CLA). CLA replaces the fluid half-space into which the plate radiates by a layer of acoustic fluid of finite thickness $H=H_0(1+i\eta)$ that is a complex number. By formulating the fluid-structure interaction problem in this manner the Fourier transformed surface pressure-velocity relationship which usually contains branch point singularities is approximated as a rational function of the transform variable. With the branch cut singularity removed, the resulting inverse Fourier transform is easily evaluated analytically, using contour integration, as a sum over a small number of poles. Each pole corresponds to a wave propagating in the fluid/structure system. The location of these poles is found analytically using a perturbation technique. Results for the radiation from a line-driven plate show excellent agreement with published results. This approach permits explicit evaluation and differentiation between the surface waves, i.e., the fluid-loaded structural waves and those associated with the acoustic field as affected by the plate's structural parameters. [Work supported by CDNSWC ILIR Program.]

1:35

4pSAa2. David Crighton and the unsteady Kutta condition. Allan D. Pierce (Dept. of Aerosp. and Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215)

Why do airplanes fly? Because of the Kutta condition. This is a facetious response to the popular explanation (Bernoulli effect) of the lift on an airplane wing. The Kutta condition is that the (inviscid flow) near any distinguished sharp point (such as the trailing edge) should be finite; it invariably requires that the flow further downstream should have a discontinuity in fluid velocity. *True potential flows* have no discontinuities, but they can have singularities at edges, and result in no lift (d'Alembert's paradox).

Acousticians live comfortably without invoking the Kutta condition when they use Sommerfeld's theory of the diffraction of sound by a knife edge, but it is invoked without hesitation by most aerodynamicists. Crighton made the distinction of steady versus unsteady flows, acoustic waves being an example of unsteady flows, and wrote an article in 1985 in Annual Reviews of Fluid Mechanics on the Kutta condition in unsteady flow. This summarized and appraised works by himself and others up to that time on when one should preferably use the Kutta condition; the article ends with the optimistic statement that a firm basis existed for understanding the applicability of the condition. The present paper attempts to explain Crighton's reasoning, but suggests that the subject is still murkily understood.

Contributed Papers

2:05

4pSAa3. Approximation of plate edge diffraction. Andrew Norris (Rutgers Univ., P.O. Box 909, Piscataway, NJ 08854-8058)

Edge diffraction from thin plates and structural elements remains an area closely connected with the name of David Crighton. He developed the tools for solving the Wiener–Hopf problems that were inevitably obtained, and derived useful asymptotic approximations under light and heavy fluid loading. His approach and his philosophy of obtaining physical insight are applied in this lecture to the fundamental configuration of a semi-infinite thin plate in a fluid. The acoustic diffraction coefficient is strongly dependent upon frequency, unlike the analogous coefficients from hard or soft screens. At low frequencies the diffraction is like that from a soft or pressure release screen, while at high frequencies the hard screen approximation is more appropriate. An approximate but simple diffraction coefficient is proposed which has the correct limiting behavior at low and high frequencies. The transition from the low to high frequency response occurs when the acoustic wavelength matches the ratio of the areal to volumetric densities of the plate and fluid. The approximate formula is compared with the exact diffraction coefficient, showing good agreement for all frequencies.

2:20

4pSAa4. Absolute instability or chaos?—A tribute to David Crighton. Sean Wu (Dept. of Mech. Eng., Wayne State Univ., 5050 Anthony Wayne Dr., Detroit, MI 48202)

David Crighton was the first person who revealed mechanisms leading to the onset of convective instability, and the first to derive an analytic formulation for predicting the critical mean flow speed beyond which absolute instability would occur. Crighton attributed the onset of absolute instability to a mathematical phenomenon known as a pinch of poles in a complex wave number domain. Crighton's work aroused an immense interest of the present author to search for the cause of absolute instability. The present paper is concerned with flexural vibration of a finite plate clamped to an infinite baffle in mean flow. Results show that the instabilities of a baffled plate are caused by an added stiffness due to acoustic radiation in mean flow, and controlled by structural nonlinearities. The added stiffness is shown to be negative and to increase quadratically with the mean flow speed. Without structural nonlinearities, the plate has one equilibrium, namely, its undeformed position. Under this condition, the amplitude of plate flexural vibration would grow exponentially in time everywhere, known as absolute instability, when mean flow speed exceeds a critical value. With the inclusion of structural nonlinearities, the plate may possess multiple equilibria, but cannot be stabilized on any of them. Instead, the plate may jump from one equilibrium position to another. Since this jumping is random, the plate flexural vibration may seem chaotic. This chaotic behavior disappears when viscous damping is introduced. Accordingly, the plate settles down to a buckled-down position owning to the hydraulic fluid-loading effect.

THURSDAY AFTERNOON, 7 JUNE 2001 SALON V, 3:00 TO 5:00 P.M.

Session 4pSAb

Structural Acoustics and Vibration and Physical Acoustics: Complexity and Random Matrix Theory II

Richard L. Weaver, Chair

Department of Theoretical and Applied Mechanics, University of Illinois, 216 Talbot Laboratory, 104 South Wright Street, Urbana, Illinois 61801

Invited Papers

3:00

4pSAb1. Random matrix theory for modeling random uncertainties in transient elastodynamics. Christian Soize[a] (Struct. Dyn. & Coupled Syst. Dept., ONERA, 92322 Chatillon, France)

A probability model for symmetric positive-definite real random matrices is constructed using the maximum entropy principle which allows only the available information to be used. The result obtained differs from the known results concerning Gaussian and circular ensembles for random matrices. The probability distribution of such a random matrix and the probability density function of its random eigenvalues are explicitely constructed. A fundamental mathematical result concerning convergence properties as the dimension of the random matrix approaches infinity is presented. An algebraic representation of the probability model has been obtained and is very well suited to Monte Carlo numerical simulation. This random matrix theory is used to construct a new nonparametric method for modeling random uncertainties in transient elastodynamics. The information used does not require the

description of the local parameters of the mechanical model. The available information is constituted of the algebraic properties related to the generalized mass, damping and stiffness matrices which have to be positive-definite symmetric matrices, and the knowledge of these matrices for the mean reduced matrix model. The fundamental properties related to the convergence of the stochastic transient response in elastodynamics is analyzed. Finally, an example is presented.[a]Present address: Ing. 2000, Bat. Copernic, Univ. de Marne-la-Vallée, 5 Bd. Descartes, 77454 Marne-la-Vallée, France.

3:30

4pSAb2. Irreversible energy transfer and distribution in complex systems. Adnan Akay (Mech. Eng. Dept., Carnegie Mellon Univ., Pittsburgh, PA 15143) and Antonio Carcaterra (INSEAN, Istituto Nazionale per Studi ed Esperienze di Architettura Navale, Via di Vallerano, 139, 0128, Rome, Italy)

This presentation examines the conditions under which irreversible energy transfer or irreversible energy redistribution can take place within dynamical systems that do not possess classical damping properties. Under usual circumstances, energy input to a system returns to its original form after a delay during which the system components undergo oscillatory motions. However, depending on the complexity of the system, the period during which energy travels within the system may become very large, reaching infinity. While conditions that lead to irreversible energy transfer in linear systems are rare, nonlinearities in the system can more readily provide continuous absorption of energy as in the case of thermalized vibrations of atoms. Numerical simulations suggest that the irreversibility attribute for a given system is deeply related to the observation time scale. Examples show that a process may appear reversible when its instantaneous energy time history is observed while its time-average trend exhibits irreversibility. By introducing entropy as a measure of the irreversibility associated with the energy transfer, this presentation suggests three controlling parameters of irreversibility: system complexity as measured by the number of its degrees of freedom, the effect of nonlinearities, and the time-average observation scale. An analysis of their interrelationships discloses new insights into the thermalization phenomenon. [Research sponsored by NSF and INSEAN.]

Contributed Papers

4:00

4pSAb3. Long and finite time approximations of high modal density systems. Paul E. Barbone (Dept. of Aerosp. and Mech. Eng., Boston Univ., Boston, MA 02215 and Inst. of Cancer Res., Sutton, Surrey, UK)

Our recent results concerning modeling of undamped structures with high modal density show that such structures can be replaced by very simple structures with damping. In the extreme case, a structure with infinitely many degrees of freedom can be replaced by a single mass-dashpot-spring system. When applied to a particular structure, however, such a model has a finite time of validity, predicted to be $T=O(\Delta\omega)^{-1}$. Here, $\Delta\omega$ represents the spacing of natural frequencies in the original structure. In this contribution, we present models valid for simulation times $T\gg(\Delta\omega)^{-1}$. The new models are based on long time rather than high modal density approximations. We show how the two models can be combined to form a single model which gives accurate results for all times.

4:15

4pSAb4. With a pinch of dissipation all operators can be inverted. John J. McCoy (The Catholic Univ. of America, Washington, DC 20064)

With domination of computation over analysis, issues of convergence and uniqueness have largely given way in the engineering/physics literature to issues of numerical stability. A common cause is sometimes the genesis of a lack of convergence or of uniqueness when seeking an analytical prediction, and of a numerical instability when seeking a numerical prediction. If the underlying cause is the same, the actions prompted are sometimes quite different. The introduction of dissipation is widely accepted as a ''tool'' for ''regularizing'' numerical codes, by virtue of incorporating a truer physics into the model. Three experiment scenarios are described for which the introduction of dissipation also serves to ''mask'' errors in the development behind the code, thereby precluding the prediction of significant physical effects. Two of the scenarios are drawn from the published literature—the errors in modeling arise in including finite frequency effects in the scattering by a rough surface, and in including inclusion/inclusion interaction in estimates of the settling rate for nondilute fluid suspensions. The third scenario has not been discussed in the published literature—the modeling error applies to the generalization of a nonreflecting boundary condition at an interior plane of an unbounded range-independent full space, to now model the reflection that obtains at a plane interface connecting two, dissimilar range-independent half-spaces.

4:30

4pSAb5. Insuring causality of frequency-response functions with hysteretic damping. M. Strasberg (Code 702, David Taylor Model Basin, NSWCCD, 9500 MacArthur Blvd., West Bethesda, MD 20817)

When calculating the temporal response $h(t)$ of a system responding linearly to an external impulsive force applied at $t=0$, the availability of personal computers that calculate Fourier transforms may make it convenient to first calculate the frequency-response function $H(\omega)$ of the system and then obtain $h(t)$ as the inverse Fourier transform of $H(\omega)$. Dissipation can be accounted for by multiplying the stiffness and elastic quantities in H by the complex factor $(1+i\eta)$, assuming ''hysteretic'' damping with loss factor η. However, this makes H a complex function of frequency whose inverse transform may be neither real nor causal. As is well known, the inverse transform can be made real by setting $H(-\omega)=H^*(+\omega)$, where * indicates complex congugate, or replacing η with $\eta\,\mathrm{sgn}(\omega)$, where $\mathrm{sgn}(\pm\omega)=\pm 1$. But this may not lead to causality. It is suggested that causality be insured by expressing the temporal function simply as $h(t)=[1+\mathrm{sgn}(t)]\hat{h}(t)$, where $\hat{h}(t)$ is the inverse Fourier transform of either the real *or* the imaginary part of $H(\omega)$, not both. It will be shown that this is just what the Kramers–Kronig causality relations require.

4:45

4pSAb6. On the spatial coupling between deterministic and fuzzy subsystems in the hybrid approach to the midfrequency problem. Philip J. Shorter and Bryce K. Gardner (Vibro-Acoust. Sci., 12555 High Bluff Dr., Ste. 310, San Diego, CA 92130)

The analysis of the dynamic behavior of a structural–acoustic system across a broad frequency range presents a number of challenges to an analyst. For a typical structural–acoustic system it is quite common to find that the modal density varies significantly between the various subsystems, across the frequency range of interest. The vast number of modes in the system as a whole can render a detailed deterministic analysis impractical, while the low modal density of certain subsystems is problematic for statistical energy analysis. One approach to the problem is to construct a hybrid model that combines deterministic and statistical descriptions of the system dynamics. One of the key requirements of the hybrid approach is an accurate estimate of the spatial coupling that occurs between the deterministic and statistical (or fuzzy) parts of the model. This paper discusses recent work which has investigated both a spatial correlation approach to calculating this coupling and also an asymptotic modal approach. The relative merits of the different approaches are discussed and a number of numerical examples are presented.

Session 4pSC

Speech Communication: Cross-Linguistic Studies: Acoustics and Perception (Poster Session)

Ann Bradlow, Chair
Department of Linguistics, Northwestern University, 2016 Sheridan Road, Evanston, Illinois 60208

Contributed Papers

All posters will be on display from 1:30 p.m. to 5:00 p.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 1:30 p.m. to 3:15 p.m. and contributors of even-numbered papers will be at their posters from 3:15 p.m. to 5:00 p.m.

4pSC1. Interlanguage benefit for non-native speaker intelligibility. Tessa Bent (Dept. of Linguist., Northwestern Univ., 2016 Sheridan Rd., Evanston, IL 60208, t-bent@northwestern.edu)

Non-native speakers (NNSs) of English must often communicate in English even if they share an L1. This study investigated how the talker–listener match with respect to language background affects L2 intelligibility. Two non-native speakers of English with an L1 of Chinese and two monolingual English speakers were recorded in English reading sentences and producing spontaneous speech. Listeners were 32 monolingual native speakers of English, 18 NNS of English with an L1 of Chinese and 14 NNS of English with an L1 other than Chinese. Results showed patterns of objective intelligibility and subjective comprehensibility ratings across listener groups that were strongly influenced by the relationship between the language backgrounds of talker and listener. As expected, native English listeners showed a preference for the native English talkers over either of the Chinese talkers. In contrast, the Chinese listeners found one of the Chinese talkers most intelligible, and the non-Chinese non-native listeners found one of the Chinese talkers equally intelligible to one of the native talkers. From these patterns of intelligibility, we conclude that, provided the talker is above a certain threshold of intelligibility, if talker and listener both share an interlanguage or are both native speakers, intelligibility is maximized.

4pSC2. Investigating the nature of cross-language perceptual comparisons: Evidence from production. Wendy Baker and Pavel Trofimovich (Univ. of Illinois at Urbana–Champaign, Urbana, IL 61801)

In cross-language speech perception research, a metric of perceived phonetic similarity between native and non-native sounds both predicts and describes non-native listeners' perceptual accuracy [Guion *et al.*, J. Acoust. Soc. Am. **107**, 2711–2724 (2000)]. By hypothesis, judgments of perceived phonetic similarity reflect the match between a non-native sound and the sound category used to process it. Whereas measures of perceptual accuracy often mask the nature of the perceptual categories used to process non-native sounds, the acoustics of non-native segment articulations may unambiguously indicate whether non-native sounds are processed in terms of perceptually similar native or newly established non-native sound categories. This study examines the relationship between judgments of phonetic similarity between Korean sounds and six English vowels in bVd and bVt words and productions of the same vowels by 40 Korean listeners who differed in age and amount of exposure to English. A metric of perceived phonetic similarity between English and Korean vowels was correlated with results of acoustic comparisons between English stimulus tokens and the listeners' productions, revealing that a metric of perceived phonetic similarity also predicts production accuracy and that the nature of the sound categories used to process non-native sounds determines cross-language perceptual relationships.

4pSC3. Training /v/–/w/ perception with native speakers of Indian languages. Radhika Aravamudhan and Anna Marie Schmidt (School of Speech Pathol. & Audiol. P.O. Box 5190, Kent State Univ., Kent, OH 44242, raravam1@kent.edu)

Experiments in training listeners to perceive non-native categories are an important aspect of cross-language perception research. In this study, ten native speakers of various Indian languages participated in training designed to improve perception of the English /v/–/w/ contrast using a fading technique (Jamieson and Morosan, 1986). First the subjects labeled English /p, b, f, v, w/ using native language (L1) orthography to ensure that other L1 consonants were not evoked by acoustic cues to English /v/, /w/. They also rated the closeness of the English sound they heard to the sound in their L1. In the five-day training, subjects heard natural stimuli produced by four (two males and two females) English speakers. The steady-state portion of the initial consonant was reiterated by varying amounts with subjects hearing progressively shorter consonants as the training progressed. Both discrimination and identification training with feedback were used. Pre- and post-training testing with stimuli from new speakers as well as recording of production of Indian subjects was completed. Results and implications will be discussed.

4pSC4. Perception and production development in citation versus sentence contexts by learners of Japanese. Yukari Hirata (Dept. of EALL, Colgate Univ., Hamilton, NY 13346, yhirata@mail.colgate.edu)

This study compares the degrees of L2 learners perception versus production development in isolated word contexts with those in sentence contexts. Native speakers of English learning Japanese participated in the perceptual training program used in Hirata [Proc. Int. Congr. Phon. Sci. **2**, 1413–1416 (1999)] in which subjects identified, and received immediate feedback on, the numbers of morae in target words in either citation or sentence contexts. Before and after training, subjects were tested on their ability to accurately perceive and produce words that included difficult durational contrasts (1) in citation contexts (word tests), and (2) in frame sentences (sentence tests). The results of the word tests showed no positive or negative correlation between the degrees of perception and production improvement. However, in the sentence tests, a negative correlation was found ($r = -0.651$): the greater the improvement in one domain, the less the improvement in the other. These contrasting results suggest that development captured in isolated word contexts does not readily predict that in fluent speech contexts, expanding Greenspan *et al.* [J. Exp. Psychol. **14**(3), 421–433 (1988)]. The factors responsible for the negative correlation in the sentence contexts are discussed in light of the mechanisim of fluent L2 acquisition. [Work supported by Consortium for Language Teaching and Learning.]

4pSC5. Foreign accented speech: Encoding and generalization. Shawn Weil (Dept. of Psych., Ohio State Univ., Lazenby Hall, 1827 Neil Ave., Columbus, OH 43210)

Speech produced by a non-native talker of a language is influenced by the talkers native language, and the result is Foreign accented speech (FAS). Previous research regarding talker variability and normalization has found that indexical talker characteristics are encoded into memory along with phonetic information, and implicitly helps subsequent speech perception. Are the overarching properties of FAS encoded in the same manner? The current study explores what talker information is encoded into memory, and how learned talker qualities improve subsequent speech recognition. Listeners were exposed to one speaker for four experimental sessions via a battery of tests measuring speech intelligibility. On the fifth day, listeners were given a post-test in signal correlated noise which measured speech intelligibility for either the talker they had been trained on, a similarly accented talker, or a talker with a different, unrelated accent. Three control groups received only the post-tests. In sentential, exposure to a talker with an accent seemed to improve performance for talkers who have a similar accent. This was not the case for single word contexts. Implications onto models of speech perception are discussed.

4pSC6. The effect of a non-native accent in Dutch on speech intelligibility. Sander J. van Wijngaarden, Herman J. M. Steeneken, and Tammo Houtgast (TNO Human Factors, P.O. Box 23, 3769ZG Soesterberg, The Netherlands, vanwijngaarden@tm.tno.nl)

It is well known that second-language talkers are usually less intelligible than native talkers, especially under adverse listening conditions. Background noise tends to influence the intelligibility of foreign accented speech more profoundly than unaccented speech. In order to quantify this effect, speech reception threshold (SRT) experiments were carried out using Dutch speech by American, German, Polish, Chinese and (native) Dutch talkers (15 speakers, 3 talkers per language background). The experience with the Dutch language varied between talkers from 3 months to 28 years. Ten Dutch university students participated as listeners. In order to obtain 50% sentence intelligibility, non-native talkers were found to require a 0–3-dB better speech-to-noise ratio than native talkers do. A significant correlation between self-reported proficiency and SRT-score was observed ($R=0.78$). Also, an objective estimate of the degree of foreign accent of the individual speakers was obtained by means of a pairwise comparison experiment, in which 19 native Dutch listeners participated. The accent ratings which were obtained (after Thurstone normalization) were also significantly correlated with the SRT-scores ($R=0.85$). The conclusion may be drawn that the effect of noise on the intelligibility of foreign accented speech can be predicted quantitatively from easily obtained measures of perceived foreign accent.

4pSC7. Speaking clearly for non-native listeners. Ann R. Bradlow, Tessa Bent, and Rebecca Schwarzlose (Dept. of Linguist., Northwestern Univ., 2016 Sheridan Rd., Evanston, IL 60208, abradlow@northwestern.edu)

Non-native listeners often have particular difficulty perceiving novel phoneme contrasts, and are disproportionately affected by background noise relative to native listeners [e.g., Mayo *et al.*, J. Speech Hear. Res. **40**, 686–693 (1997); Meador *et al.*, Bilingualism **3**, 55–67 (2000)]. This study investigated whether these difficulties could be overcome by naturally produced clear speech. Based on previous findings that hearing-impaired adults [Picheney *et al.*, J. Speech Hear. Res. **28**, 96–103 (1985)] and learning-impaired children [Bradlow *et al.*, J. Acoust. Soc. Am. **108**, 2603 (2000)] derive great benefit from clear speech (17%–20% and ~9%, respectively), we predicted that non-native listeners would also exhibit a substantial clear speech effect. The 32 non-native and 32 native listeners were presented with English sentences. Factors that varied were speaking mode (conversational versus clear), signal-to-noise ratio (-4 vs -8 dB) and talker (male versus female). Contrary to our prediction, non-native listeners exhibited a relatively small clear speech effect (~5%). This result contrasts sharply with the large clear speech effect for hearing-impaired adults and learning-impaired children. We interpret these cross-population differences as reflecting differential effects of speech perception deficits that arise from problems of access to the underlying linguistic code versus those that arise from problems of access to the signal.

4pSC8. Perceptual and acoustical similarity of six European languages. Verna B. Stockmal and Zinny S. Bond (Linguistic Dept., Gordy Hall, Ohio Univ., Athens, OH 45701)

Listeners have shown considerable sensitivity to the ''acoustic signature'' or ''sound'' of languages when discriminating or identifying unknown foreign languages. Which acoustic/phonetic properties of languages listeners employ when making these judgments are only imperfectly understood. Bond and Stockmal [Acoust. Soc. of Am. Berlin (1999)] found that initially, listeners made judgments about Asian languages based on regional characteristics, what might be termed voice setting [J. Esling and R. Wong, TESOL Q. **17**, 89–95 (1983)], rather than by rhythmic properties. Perceptual focus shifted and rhythm became more salient when the set of languages shared regional characteristics. At least some of the rhythmic properties of languages can be extracted by phonetic measurements of the acoustic signal [F. Ramus, M. Nespor, and J. Mehler, Cognition **73**, 265–292 (1999)]. This study presents instrumental measurements based on consonant/vowel segmentation of six European languages, Bulgarian, Czech, Greek, Hungarian, Italian, and Romanian. The acoustical measurements were used to plot the relationship of the languages to one another. Listeners were asked to judge similarity of spoken samples of the languages and to indicate which salient feature of the language was the basis for each judgment. Comparison of the perceptual judgments and instrumental measurements was made.

4pSC9. Effects of noise and proficiency level on intelligibility of Chinese-accented English. Catherine L. Rogers (Dept. of Commun. Sci. and Disord., Univ. of South Florida, 4202 E. Fowler Ave., BEH 255, Tampa, FL 33620-8150), Jonathan M. Dalby (Commun. Disord. Technol., Inc., Bloomington, IN 47404), and Kanae Nishi (Univ. of South Florida, Tampa, FL 33620-8150)

It is known that native speech intelligibility is degraded in background noise. This study compares the effect of noise on the intelligibility of English sentences produced by native English speakers and two groups of native Mandarin speakers with different English proficiency levels. High-proficiency Mandarin speakers spoke with detectible accents, but their speech was transcribed at about 95% of words correct in a previous study, in which no noise was added [C. Rogers and J. Dalby, J. Acoust. Soc. Am. **100**, 2725 (1996)]. Low-proficiency Mandarin speakers were transcribed at about 80% correct in the same study. Forty-eight sentences spoken by six speakers (two native, two high proficiency, and two low proficiency) were transcribed by listeners under four conditions: with no added noise and mixed with multi-talker babble at three signal-to-noise ratios ($+10$, 0, and -5 dB). Transcription accuracy was poor for all speakers in the noisiest condition, although substantially greater for native than for Mandarin speakers. For the other three conditions, the rate of degradation of intelligibility with increasing noise was much greater for accented than for native speech. These results confirm and extend those reported in Munro [Stud. Sec. Lang. Acq., 139–153 (1988)]. [Work supported by University of South Florida.]

4pSC10. Perceptual weighting in fricative/vowel syllables for speakers of English and Mandarin. Sreedivya Radhakrishnan and Anna Marie Schmidt (School of Speech Pathol. & Audiol., Kent State Univ., P.O. Box 5190, Kent, OH 44242, sradhakr@kent.edu)

Adult English speakers have been shown to pay more attention to noise spectra in labeling syllable–initial fricatives than to formant transitions. However, cross-language research indicates that languages differ in weights assigned to aspects of the acoustic signal. In this study, native speakers of English and Mandarin Chinese labeled and rated for goodness (in the L1) both English and Mandarin alveolar and post-alveolar fricative noises from four vowel environments. They also labeled and rated natural FV syllables in each language. Finally, subjects heard F + V combinations (from their L1 only) in which fricative noise from the anterior fricative was combined with vowel transitions from more posterior fricative and vice versa. Vowel length was manipulated in one set of stimuli. Results indicate differences in weighting by the two groups of subjects.

4pSC11. Acoustic study of vowel nasalization in Thai. Chutamanee Onsuwan and Patrice Speeter Beddor (Dept. of Linguist., Univ. of Michigan, Ann Arbor, MI 48109)

Patterns of coarticulatory nasalization and segmental duration were investigated in Thai, a language with contrastive vowel length. Four Thai speakers produced multiple repetitions of real and nonsense CV(:)N items, where C=[b], V=/i e ae/, V:=/i: e: ae:/, and N=bilabial, alveolar, or velar nasal. Three temporal measures were taken: duration of V(:), N, and vowel nasalization. Acoustic onset of vowel nasalization was primarily identified from FFT spectra, sampled across the vowel in 10-ms increments, as onset of the nasal formant and/or broadening of formant bandwidth in the low-frequency region. Preliminary results from two speakers showed that long V: was nearly twice as long as short V, and final N was nearly two times longer following short V than long V: [A. S. Abramson, Int. J. Am. Linguist. **28** (1962)]. Vowel nasalization varied as a function of V(:)N duration: relative to N onset, nasalization started earlier in long V: than short V. However, long V: was proportionately less nasalized (44% for long V: vs 49% for short V). These results are compared to findings for English, and interpreted relative to look-ahead and coproduction models of coarticulation.

4pSC12. Cross-language perception of rate induced resyllabification. Kyoko Nagao, Byung-jin Lim, and Kenneth de Jong (Dept. of Linguist., Indiana Univ., Bloomington, IN 47405)

This paper investigates the degree to which speakers of languages with different syllabic inventories are similar in their perceptions of syllabic affiliation. Stetson (1951) noted that repeated coda (VC) structures become perceived as onset (CV) structures as repetition rates increase. Stimuli were extracted from a repetitive production experiment in which English talkers produced voiced and voiceless labial stop onsets and codas at tempi controlled by a rate varying metronome. Native English, Japanese, and Korean listeners were asked to label the repeated syllable with one of four choices: ''bee,'' ''pea,'' ''eeb,'' ''eep.'' All three language groups showed perceptual shifts from VC to CV as the speech rate increased. However, Japanese listeners are biased toward identifying VC productions as CVs, and very rarely identified CV productions as VCs. These results indicate a very consistent cross-language perception of syllabic affiliation, regardless of a language's syllabic inventory, although the language's inventory does have a small effect on such perceptions. These results differ from those for the voicing contrast. Voicing contrasts in coda position which are non-native for Koreans were more poorly categorized by the Korean listeners. [Work supported by NIDCD and NSF.]

4pSC13. Voice onset time of Jordanian Arabic stops. Fares Mitleb (Dept. of English, Yarmouk Univ., Irbid, Jordan)

This paper reports on findings of an acoustic measure of voice onset time (VOT) in Jordanian Arabic stops in syllable initial position. Results point out that VOT values corrolate to voice contrast. Also, a difference between VOT values is found due to the length contrast of the following vowel. In addition, our results seem to depart from the literature with regard to the effect of place of articulation on VOT values. The findings of this piece of research are discussed with respect to the importance of acoustic codes to the proper treatment of some low-level phonetic parameters. [This research was supported by Yarmouk University.]

4pSC14. Acquisition of dialectal differences in English by native Japanese speakers. Julie McGory, Elaina Frieda, Shawn Nissen, Robert A. Fox (Dept. of Speech and Hearing Sci., The Ohio State Univ., Columbus, OH 43210-1002), Kelly Friedman, and Kim Rosenbauer (The Ohio State Univ., Columbus, OH 43210-1002)

Many previous investigations of second language (L2) learning have focused on the acquisition of phonetic segments by groups of non-native speakers living in diverse regions in the US. These studies have not, however, assessed possible dialectal variation in L2 productions. The current study investigates the extent to which dialectal differences found in the vowel systems of two different American English (AE) dialects are acquired by adult Japanese speakers. Four subject groups were identified: native English speakers of a midwestern dialect of AE, native English speakers of a southern dialect of AE, Japanese speakers whose target language is a midwestern AE dialect, and Japanese speakers whose target language is a southern AE dialect. Subjects were recorded saying word and nonword utterances in isolation. Acoustic measures of vowels that can potentially capture differences between the dialects were made. These include vowel duration, and $F0$, $F1$, $F2$, and $F3$ frequencies measured at five equal distances throughout the vowel. $F0$–$F3$ measurements were used to calculate bark differences. Preliminary findings suggest that adult Japanese learners acquire some of the acoustic properties that are consistent with the dialect of the target language. [Work supported by an INRS Award from research funding from ATR (Fox, P.I.).]

4pSC15. Emphatic Koreans and neutral Americans? Soohee Kim, Emily Curtis, and Lesley Carmichael (Univ. of Washington, Box 353520, Seattle, WA 98195)

Intonation is one of the subtle grammatical aspects of a second language that often escapes formal instruction and thus acquisition. If intonation is acquired early on in first language, and no formal instruction is received, speakers are likely to carry over native language intonation patterns in a second language. The neutral intonation pattern for indicative sentences may differ across languages, for example. It is hypothesized that the neutral intonation pattern of the Seoul dialect (mid-high–low–high-low) is carried over to English by native speakers of Korean. It so happens that mid-high–low–high is an intonation contour that expresses emphasis or dissatisfaction in English, often resulting in perception of unintended emotion or inappropriate conversational implicature on Korean speakers' part. To test our hypotheses, five native Seoul dialect speakers' intonation contours in five statements are compared in both English and Korean. Koreans' English productions are also rated for emotionality by 10 native speakers of English. Rules of intonation are an integral part of the gram-

mar of a language and should be explicitly taught to learners of a second language, especially where their grammatical (or pragmatic) import differs.

4pSC16. Production and perception of authentic and feigned Spanish accent. Rebecca Hill (Prog. in Linguist., Univ. of Florida, P.O. Box 115454, Gainesville, FL 32611-5454)

This experiment will be undertaken in order to examine the ability of native English speakers to distinguish between authentic and feigned Spanish accent. To that end, 50 English CVC words produced by 6 native Spanish speakers and 6 native English speakers will be recorded. The same six native English speakers will also be asked to produce the same words using self-determined Spanish-accented English. The English consonants chosen for analysis are [p/b, t/d, k/g] and the vowels chosen are [i,e,æ,u] and their lax counterparts. These consonants and vowels were chosen because they have been shown to be problematic for native speakers of Spanish learning English. The production experiment included the acoustic measurement of VOT and closure duration for the stop consonants and $F1$, $F2$, and duration for the vowels. The perception experiment will be conducted using ten native English speakers, five of whom are familiar with Spanish and five who are naïve. Listeners will hear the three types of speech (authentic Spanish accent, authentic American English accent, and feigned Spanish accent) and rate the degree of accentedness using a five-point scale. Results of the perception experiment will be examined and discussed in relation to the production data.

4pSC17. Lexical tone contrast effects related to linguistic experience. Alexander L. Francis and Valter Ciocca (Dept. of Speech and Hearing Sci., Univ. of Hong Kong, Hong Kong)

Listeners' auditory discrimination of vowel sounds depends in part on the order in which stimuli are presented (Cowan and Morse, 1986). Such contrast effects have been argued to be language-independent (Polka and Bohn, 1996), and to result from psychophysical (not speech- or language-specific) order-of-presentation effects such as decay of memory traces over time or increased weighting of later-occurring stimuli (Macmillan, Braida, and Goldberg, 1992; Repp and Crowder, 1990). In the present study, native Cantonese speakers' discrimination of a linguistic tone continuum is shown to exhibit contrast effects similar to those shown for vowels in previous studies. When presented with two syllables differing in $F0$ by approximately 4 Hz, listeners were significantly more sensitive to this difference when the first syllable was higher in frequency than the second. However, American English-speaking listeners with no experience listening to Cantonese showed no such contrast effect when tested in the same manner using the same stimuli. Neither English nor Cantonese listeners showed any contrast effects in the discrimination of a nonspeech continuum with the same frequency properties. These results suggest that tone contrast effects, unlike vowel effects, may be language-specific and not merely the consequence of general properties of auditory processing.

4pSC18. A cross-linguistic study on discourse and syntactic boundary cues in speech. Janice Fon and Keith Johnson (Dept. of Linguist., The Ohio State Univ., 1712 Neil Ave., Rm. 222, Columbus, OH 43210, jfon@ling.ohio-state.edu)

Online segmentation is a formidable task for any speech parser. Yet, human beings seem to handle this task with ease. This study focuses on the relationship between discourse/syntactic boundaries and prosodic cues in divergent languages—English, Japanese, and two varieties of Mandarin (Guoyu and Putonghua). Speech was elicited by having talkers describe the events in a film [W. L. Chafe (unpublished)]. Recorded data were transcribed, digitized, and will be segmented into discourse and syntactic units while measurements of F0, syllable duration, and rms amplitude will be taken on the digitized data. Prosody will be partially labeled following ToBI (Tones and Break Indices) conventions of each language. A comparison of different dimensions of data-discourse/syntax, acoustics, and intonation, will be made in order to examine boundary cues in speech. Our previous work using a smaller set of Mandarin (Guoyu) data has shown that syllable onset interval isochrony is generally preserved until the pre-boundary syllable. Pre-boundary syllables are characterized by a significant lengthening effect, which differs by boundary types. Lengthening before discourse units is longer than that before syntactic units such as clauses or phrases. [Work supported by NIDCD.]

4pSC19. Language-specific knowledge and the perception of tonal contrasts in Italian and English. Mariapaola D'Imperio (LORIA, Univ. of Nancy, France, dimperio@loria.fr)

Intonation, including details of tonal alignment [i.e., the synchronization of tones and segments (D'Imperio, 2000)], has been claimed to be part of the phonological knowledge of native speakers. This linguistic knowledge can be assumed to influence the perception of tonal contrasts. Both American English and Neapolitan Italian have two rising pitch accents (L+H* and L*+H) whose alignment is contrastive, although details of their implementation differ. The accents also cue different pragmatic functions (e.g., cuing the question/statement contrast only in Neapolitan) and are subject to different syntagmatic constraints in the two languages. This study tested the hypotheses that the American listeners would be able to perceive the contrast between the two Neapolitan accents, despite the aforementioned differences, and that they would respond similarly to a linguistic (question/statement) and a psychoacoustic (early/late peak) identification task. The stimuli were constructed by manipulating the alignment of a Neapolitan utterance through PSOLA. These stimuli were employed in a linguistic task for both language groups and a psychoacoustic task for the Americans only. Although the Americans could perform the linguistic task, scoring similarly to the Neapolitans, they could not perform the categorization in the psychoacoustic task. The results bear upon the ''universality'' of tonal perception.

4pSC20. Modality salience in multimodal speech perception under degraded presentation conditions. Janet M. Weisenberger, Sandra M. Kreidler, Katherine L. York, and Michelle A. Yeary (Speech and Hearing Sci., Ohio State Univ., Columbus, OH 43210)

Studies of the McGurk effect in auditory-visual speech perception have demonstrated that both modalities contribute to the overall percept, even when both modalities are perfectly rendered. Initial studies in our laboratory found that a McGurk-type effect can also be observed with tactile-visual presentation, employing a tactile speech perception device. The present study investigated the question of whether subjects' relative reliance on a particular sensory modality could be shifted when presentation via one modality was degraded. Subjects were tested with a McGurk paradigm in unimodal (visual, auditory, or tactile) and multimodal presentations, in both nondegraded and degraded presentation conditions. Auditory degradation was achieved by low-pass filtering with cutoffs at 400 or 1000 Hz, and visual degradation was achieved by use of a diffusing screen placed in front of a video monitor. Tactile stimuli were not degraded. Results showed that subjects quickly shifted response patterns when one modality was degraded to show increased reliance on the nondegraded modality. However, increased reliance on the nondegraded tactile input was not found, possibly because the initial training with the tactile device was limited. Nonetheless, tactile-visual McGurk effects were replicated. [Portions of this work were supported by NIH.]

Session 4pUW

Underwater Acoustics and Signal Processing in Acoustics: Underwater Acoustic Communications II: Algorithm Design and Analysis for Realistic Propagation Conditions

James C. Preisig, Chair

Department of Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution, M/S 11, Bigelow 210, Woods Hole, Massachusetts 02543-1053

Contributed Papers

2:00

4pUW1. High-speed acoustic communication in shallow water using multiple coherent path beamformer technique. Pierre-Philippe J. Beaujean and Lester R. LeBlanc (Florida Atlantic Univ., 101 N. Beach Rd., Dania Beach, FL 33004)

A multiple coherent space-time processing method is used for acoustic communication at high data rate in very shallow water, between 16 and 32 kHz. The multiple coherent path beamformer (MCPB) method for processing the signals uses an adaptive processor that forms beams in the direction of a collection of coherent signals representing the strongest path. The source is the FAU Utility Acoustic Modem used for acoustic networking during AUV operations. The system encodes data using coherent modulation schemes with coded rates varying between 4000 and 48 000 bps. The receiver system is composed of a 64-channel Mills–Cross array attached to an FAU embedded 64-channel acquisition system. Results already obtained show that the MCPB can decode data reliably at 3200 m for rates up to 32 000 bps, with a platform speed of 2–3 knots and a sea-state of 1–2. For a source level of 186 dB, BPSK encoding and 62.5s symbols (16 000 bps), the MCPB achieves a bit error rate of 0.2% without any Forward Error Coding. These results demonstrate the long-range and high-data rate transmission ability of the system, and make possible real-time transmission of side-scan data and other low-resolution video images.

2:15

4pUW2. Performance analysis of an underwater acoustic communications time diversity receiver. Michael S. Richman, Scott L. Whitney, and Geoffrey S. Edelson (Advanced Systems, BAE Systems, P.O. Box 868, Nashua, NH 03061-0868)

One method for introducing time diversity into a decision feedback equalizer (DFE) spatial diversity communications receiver operating in an underwater acoustic environment is presented and analyzed. Time diversity refers to the repetition of transmitted packets and the subsequent batch processing of these packets by an appropriately modified multichannel receiver. To overcome the effects of the time-varying nature of the channel, the modifications made to the spatial diversity-based multichannel DFE (SD-DFE) include appropriate Doppler shift compensation for each received packet and determination of the proper alignment among the time-displaced packets to allow for successful equalization. The proposed time diversity DFE (TD-DFE) receiver is shown to achieve a high rate of packet success under a variety of channel conditions. More significantly, the TD-DFE algorithm is able to equalize packets when a comparable SD-DFE might fail. The packet-to-packet coherence of a series of transmissions is analyzed to provide insight into how the characteristics of the channel affect the performance of the TD-DFE. This analysis illustrates why a time diversity-based equalization approach can be more robust for the underwater acoustic channel as compared to a standard TD-DFE algorithm at the possible expense of reducing throughput.

2:30

4pUW3. A channel subspace filtering approach to adaptive equalization of realistic acoustic channels. Raj R. Nadakuditi (MIT/WHOI Joint Program, Cambridge, MA 02139) and James C. Preisig (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

One of the major problems in underwater acoustic communications is compensating for the large amount of time varying intersymbol interference (ISI) due to multipath. Associated with each of the deterministic propagation paths are macro-multipath fluctuations which depend on large scale environmental features and geometry and micro-multipath fluctuations which are dependent on small scale environmental inhomogeneities. For arrivals which are unsaturated or partially saturated, the fluctuations in the ISI are dominated by the macro-multipath fluctuations resulting in correlated fluctuations between different taps of the sampled channel impulse response. Traditional recursive least squares (RLS) algorithms used for adapting channel equalizers do not exploit this structure. A post-filtering algorithm is presented to exploit this channel correlation structure in least squares based channel estimators and equalizers. The improvement in the performance of the algorithm with respect to the traditional least squares algorithm is predicted theoretically and demonstrated using both simulation and experimental data.

2:45

4pUW4. Quantitative bit error analysis of time-reversal communication sequences. G. F. Edelmann (Scripps Inst. of Oceanogr., Univ. of California, San Diego, La Jolla, CA 92093-0238), T. Akal (SACLANT Undersea Res. Ctr., 19138 La Spezia, Italy), W. S. Hodgkiss, S. Kim, W. A. Kuperman, and H. C. Song (Scripps Inst. of Oceanogr., Univ. of California, San Diego, La Jolla, CA 92093-0238)

Underwater acoustic communications must mitigate the intersymbol interference caused by the time-varying multipath dispersion. An experiment was conducted in June 2000 demonstrating that the time-reversal process recombined the temporal multipath resulting in reduced bit errors for communication. Quantitative bit error results will be shown for BPSK (binary phase shift keying) and QPSK (quadrature phase shift keying). Communication sequences were transmitted over a distance of 10 km both in range-independent and range-dependent environments north of Elba Island, Italy. The range-independent transmissions were made in 100-m-deep water and the range-dependent transmissions were made upslope from 100-m-deep water into 40-deep-water. [Work supported by ONR.]

3:00

4pUW5. Passive phase-conjugate signaling using pulse-position modulation. Paul Hursky, Michael B. Porter (Sci. Applications Intl. Corp., La Jolla, CA 92037), Vincent K. McDonald, and Joseph A. Rice (Space and Naval Warfare Systems Ctr., San Diego, CA)

Acoustic communications is an attractive option for networking distributed assets, such as UUVs, autonomous sensors, and other systems that must talk to one another in the ocean. Although the ocean is often a complex multipath channel, impressive progress has been made in developing equalization algorithms to overcome this. Unfortunately, many of these algorithms are computationally demanding and not as power efficient as we would like, operating at the bottom of the ocean. At the same time, the networking applications being envisioned often also require low rate, but highly reliable signaling, for purposes such as wakeup. The passive phase-conjugate concept [J. Acoust. Soc. Am. **95**, 1450–1458 (1994)] is a very simple demodulation technique which nevertheless is coherent and can overcome signal distortion due to multipath. To explore its use in low power, lightweight applications, we have extended this approach to use pulse-position modulation (PPM) and have tested it during several SignalEx experiments using an 8–16 kHz band. We present results of these tests, discuss synchronization and Doppler compensation, and compare the PPM signaling scheme to frequency-shift keying, which is often also considered for low power, lightweight applications.

3:15–3:30 Break

3:30

4pUW6. Multipath of sound propagation in shallow water and underwater spread spectrum communication. Yan Chen and Geng Chen (Inst. of Acoust., Chinese Acad., Beijing 100080, PROC, chy@oceana.ioa.ac.cn)

Due to the influence of the sea bottom and the sound speed profile, the propagation of sound in shallow water appears to be a complicated multipath. The time spread and the amplitude fluctuation of different arriving paths cause strong fading and intersymbol interference (ISI). At the receiving point the multipath signal can be expressed as $r(t)=A_0 s(t-\tau_0)+\Sigma_{i=1}^{N-1} A_i s(t-\tau_i)+n(t)$, where the first term is arriving signal directly, the second term is arriving signals by different paths, the third term is noise. When the difference of time delays is over the duration of code chip, a strong ISI occurs. As we know, the pseudo-random signal has high resolution in both time and frequency domains. It is often used in spread spectrum communication. In this paper a diversity technique with spread spectrum is used for realizing coherent accumulation of the arriving paths. The results of the lake experiment and the sea experiment are shown in this paper. The efficiency of coherent accumulation is obvious.

3:45

4pUW7. Using continuous additive training sequences for high rate, Doppler tolerant communications. Tim Collins, Philip R. Atkins, Claire Bongiovanni (School of Electron., Univ. of Birmingham, Edgbaston, Birmingham, UK, t.collins@bham.ac.uk), Jon J. Davies, Shaun M. Dunn, and Steve A. Pointer (DERA Winfrith, Winfrith Tech. Ctr., Dorchester, Dorset, UK)

An underwater acoustic communications system designed to transmit 20 kbit/s over ranges from 100 m up to 5 km between a semi-autonomous underwater vehicle and a surface vessel moving with up to 20 knots relative velocity is presented. Under these conditions, attention must be given not only to the steady-state equalization capability of the receiver, but also to its ability to respond to sudden changes in channel profile due to sudden movements of the remote vehicle. In addition to conventional packet-based transmissions, a novel continuous transmission system was also implemented during the experimental stages. Instead of using distinct channel probe, training and data blocks, a continuous training sequence was added on top of a continuous data transmission. Advantages of continuous additive training sequences (CATS) such as an improved tolerance to burst errors, greater data throughput and flexible re-synchronization ability will be highlighted. These are balanced against the disadvantages of increased receiver complexity and compromised ambient noise performance. Results of several sea-trials will be presented to verify the performance of the system at short and long ranges using both static and dynamic platforms. [Work sponsored by DERA Winfrith, UK.]

4:00

4pUW8. Sea trial results of a robust and spectral-efficient OFDM underwater communication system. Byung-Chul Kim and I.-Tai Lu (Dept. of Elec. and Computer Eng., Polytechnic Univ., 901 Rte. 110, Farmingdale, NY 11735, itailu@poly.edu)

In single-carrier, high data-rate systems symbol durations are much smaller than channel delay spreads. Reliable equalization in such systems is extremely difficult to attain because the channel is time varying and highly frequency selective. One way to circumvent the need of using costly equalization schemes, while maintaining high data rates, is to use multicarrier modulation in which broadband data is transmitted in parallel on multiple narrowband subcarriers. Symbol duration T in such systems is usually larger than delay spreads such that each narrowband subcarrier experiences only nonfrequency-selective flat fading. To improve spectral efficiency and eliminate the use of guard bands between subcarriers, each subcarrier can be chosen to be orthogonal to the others. Such a multicarrier system is referred to as orthogonal frequency division multiplexing (OFDM). Unfortunately, the orthogonality is destroyed in typical underwater channels due to large Doppler shifts and spreads (compared to the inverse of symbol durations). Therefore, previous work in UWA required the spacing of adjacent subcarriers to be $2/T$, which is twice the minimum spacing $1/T$. In this paper, time-rescaling and pilot-symbol aided channel estimation techniques are employed to combat Doppler shift and spreads. Reliable results are obtained with $1/T$ subcarrier spacing in sea trials.

4:15

4pUW9. Blind multichannel decision-feedback equalization of precoded OQPSK signals. Andreas Waldhorst, Rolf Weber, and Johann F. Böhme (Lehrstuhl für Signaltheorie, Ruhr-Universität Bochum, 44780 Bochum, Germany, andreas.waldhorst@ieee.org)

A new special modulation format is derived from Offset Quadri-Phase Shift-Keying (OQPSK) and is applied to high data-rate acoustic digital transmission in very shallow water. A novel multichannel receiver structure for this signal type is introduced. It is based on an adaptive decision feedback equalizer (DFE), which is extended so that both its feedforward and feedback sections are allowed to operate fractionally spaced. This is accomplished by exploitation of the special properties of the introduced modulation format at the receiver's decision device. To compensate for Doppler effects, explicit feedforward timing and decision-directed carrier recovery are performed. This together enables the receiver to operate in a completely self-recovering or blind manner, which is verified by analyzing experimental data gathered in the North Sea during the ROBLINKS 1999 experiments. To demonstrate the good performance of the proposed blind receiver, results of successful data recovery after transmission over a few kilometers through shallow water in cases with both a fixed and a moving source are given. [Work supported by the European Commission.]

Session 5aAA

Architectural Acoustics and Noise: Privacy and Articulation in Classrooms and Other Occupied Spaces

David Lubman, Chair
D. Lubman & Associates, 14301 Middletown Lane, Westminster, California 92683

Contributed Papers

8:30

5aAA1. Ease of hearing in various classroom geometries. Abigail Stefaniw (Architecture Dept., Rensselaer Polytechnic Inst., 110 8th St., Troy, NY 12180, stefaa@rpi.edu)

There are many proposed standards by various organizations for classroom acoustics. However, none of these acoustical professionals include architectural designs that would make their criteria appear pragmatic to architects. This paper will summarize current research on classroom acoustics, and further it by linking it to geometrical acoustics. Current classroom standards can be condensed into three settings for three criteria: reverberation time, signal-to-noise ratio, and background-noise level. Each of these settings will be auralized in three different geometries for elementary school classrooms. The auralization will be used to administer intelligibility tests and ease of hearing tests to trained listeners. The ease of hearing questions will measure the concentration effort required to achieve intelligibility. This information will allow an architecturally realistic evaluation of the best acoustic geometry needed for high-intelligibility speech communication in learning spaces.

8:45

5aAA2. Listening and learning in classroom acoustical design. Elizabeth Petry (Univ. of Hartford, 200 Bloomfield Ave., West Hartford, CT 06117), James McClellan, and Patricia Myler (Fletcher Thompson, Inc., Hartford, CT 06103-2703)

Listening and learning are proven ingredients to the successful education of our nation's children. Research studies confirm that the acoustical environment in a classroom is an important variable in the academic achievement of children. For effective communication in a classroom, unoccupied ambient noise levels should not exceed 30 to 35 dBA, while reverberation times (RT) should not surpass 0.4 s. Unfortunately, these acoustic conditions are seldom found. Even children with ''normal'' hearing sensitivity often encounter significant difficulty understanding noisy or reverberated speech. Excessive classroom reverberation and/or noise can compromise speech perception, but also reading/spelling ability, behavior, attention, concentration, and academic achievement in children with normal hearing and in children with hearing loss. The concept of an ideal acoustical classroom environment provides design, engineering, and budgetary challenges for architects, engineers, school administrators, and community leaders. To address these somewhat conflicting challenges, a team of architects and engineers has collaborated to develop successful auditory and architectural solutions. These real life solutions in Connecticut schools, John Trumball Primary School, Watertown; Rochambeau Middle School, Middlebury; and East Lyme High School, East Lyme will serve as models to other communities, educators, planners, and architects as we enter the new millennium.

9:00

5aAA3. Acoustic design of the Charles R. Drew Charter School media center—a case study. Jesse J. Ehnert (Newcomb & Boyd, One Northside 75, Atlanta, GA 30318-7761)

The new Charles R. Drew Charter School in the East Lake community of Atlanta, GA, slated to open in August 2001, is a privately managed public school focusing on mathematics and reading for grades K–8. In addition to several classrooms, music rooms, laboratories, offices, and a cafetorium, the project includes a media center and reading area which share a common volume with the school's main entrance lobby. Primary concerns in the design included providing adequate sound isolation for the various spaces within this common volume, ensuring suitably low background noise levels from the HVAC system, and providing high levels of speech intelligibility in the reading area. Designing to these goals while keeping the various areas all within the same space posed unique challenges to the design team. An iterative and collaborative design effort, which included computer modeling, was used to provide a final design which achieved a successful compromise among the various requirements.

9:15

5aAA4. Optimum reverberation time for speech intelligibility in classrooms. Eva-Marie Nosal (Dept. of Mathematics, Univ. of British Columbia, 1984 Mathematics Rd., Vancouver, BC V6T 1Z2, Canada, eva-marie@math.ubc.ca) and Murray Hodgson (Univ. of British Columbia, Vancouver, BC V6T 1Z3, Canada)

This work addresses the topical question of the optimum reverberation time (RT) for speech intelligibility in classrooms. Existing literature is reviewed and is found to be inconsistent. Experimental methods generally predict zero optimum RT, while theoretical methods predict nonzero values. These differences are discussed; they appear to stem from different treatments of the background noise. Assuming diffuse-field theory, a theoretical model to predict optimum RT was developed from a previously proposed model [S. R. Bistafa and J. S. Bradley, J. Acoust. Soc. Am. **107**, 861–875 (2000)]. It incorporates noise in a physically realistic way and predicts nonzero optimum RTs. Subsequently, to reduce reliance on diffuse-field theory, the model was modified using empirical methods to predict speech level and reverberation times in classrooms [M. R. Hodgson, J. Build. Acoust. (in press, 2001)]. Again, nonzero optimum RTs were predicted. The prediction models and results are presented, practical conclusions are drawn, and suggestions for further work are made.

9:30

5aAA5. Subjective evaluation of speech transmission quality in rooms with sentence materials. Hiroshi Sato, Jun Kanda, Shuichi Nishitani, Hiroshi Yoshino (Dept. of Architecture and Build. Sci., Grad. School of Eng., Tohoku Univ., Japan), and Masayuki Morimoto (Environ. Acoust. Lab., Kobe Univ., Japan)

The situation is often encountered that the score of syllable, word or sentence intelligibility test or measured indexes of speech intelligibility (i.e., STI, RASTI) is different from the listeners impression. This is caused by the difference between subjective impression and speech information recognition. Four kinds of listening tests were performed to obtain a subjective evaluation for speech transmission in auditoria; (a) easiness of speech perception by paired comparison method, (b) easiness of speech perception by rating scale method, (c) acceptability of speech perception by yes–no-form method, (d) difficulty of sentence listening. The results from method (a) treated as a standard data because of its accuracy. The results demonstrate that (1) easiness of speech perception presented only

relative value, (2) acceptability is not suitable for rating the speech transmission quality of sound fields because of the difficulty of judgment, (3) difficulty is easy to understand from its meanings and gives absolute evaluation.

9:45–10:00 Break

10:00

5aAA6. Speech privacy in open-plan offices. I. Systematic measurements. John Bradley and Chong Wang (Natl. Res. Council, Montreal Rd., Ottawa K1A 0R6, Canada)

This work reports the results of a systematic series of measurements of speech propagation between workstations in a conventional open-plan office. The propagation measurements were repeated as the parameters describing the open-office configuration were systematically changed. The values of the following were varied: workstation panel height, panel absorption, workstation plan size, ceiling height, and ceiling absorption. In addition the effects of the location and type of ceiling light fixtures were investigated as well as details of the workstation furnishings and configuration. The results are evaluated in terms of Speech Intelligibility Index values, SII (ANSI S3.5 1997). The SII is a new version of the articulation index and is a weighted speech-to-noise ratio that is a good measure of the expected speech privacy. The height of the partial height panels separating workstations and the absorption of the ceiling have the most important effects on speech privacy. Workstation size, panel absorption, and lighting configuration have intermediate effects on speech privacy. Ceiling height and workstation configuration had smaller effects on expected speech privacy.

10:15

5aAA7. Speech privacy in open-plan offices. II. Developing a model. Chong Wang and John Bradley (Natl. Res. Council, Montreal Rd., Ottawa K1A 0R6, Canada)

This work gives an overview of the development of a mathematical model of speech propagation between workstations in a conventional open-plan office. The work was carried out in two stages. A model was first developed for sound propagation over a single screen and including realistic floor and ceiling reflections. Maekawa's screen diffraction result was used and reflections were modeled using an image sources technique. Interference effects are seen to be important at lower frequencies but a simple energy addition of diffracted and reflected sound gives adequate results at speech frequencies. The second phase of the model also included reflections due to the panels of a complete pair of rectangular workstations and again using an image sources technique. Correctly modeling the ceiling absorption and the influence of ceiling-mounted light fixtures are seen to be critical to achieving accurate predictions. The current model was found to predict Speech Intelligibility Index values with an rms error of only 0.02.

10:30

5aAA8. An architectural design tool for the prediction of the acoustic comfort conditions in architectural spaces. Fausto Rodriguez-Manzo (Lab. de Diseno Acustico, Div. de Ciencias y Artes para el Diseno, Universidad Autonoma Metropolitana-Azcapotzalco, Mexico, rfme@correo.azc.uam.mx)

It is usual in many countries that the study programs of architectural studies lack acoustical education; as a consequence, most of their projects and buildings suffer from acoustical problems and noise intrusions, bad acoustical insulation, and spaces reverberant enough to affect speech inteligibility conditions to a large extent. Indeed, this is a problem of acoustical criterion. Due to budgetary reasons and fairly complex calculations for relatively small projects, the architect may not want to be involved in the above, or look for an acoustical consultant. This work presents a simplified model of the acoustical analysis of acoustic comfort conditions in architectural spaces. The model is useful preferably for architectural design purposes, allowing an advanced knowledge of the acoustic comfort conditions of an architectural space. In this model, most of the cases are included, although it is not designed for those spaces in which critical acoustic requirements are called for, such as theaters, and concert halls. The present model is based on the classification of comfort as applied for architecural spaces and human activities. A sample case will be presented to enhance the capabilities of the model.

10:45

5aAA9. A study on speech privacy in closed offices. Daryl J. Caswell, Yong Ma (Univ. of Calgary, 2500 University Dr. NW, Calgary, AB T2N 1N4, Canada, djcaswel@ucalgary.ca), Jim T. Goodchild (SMED Intl., Inc., Calgary, AB T2C 4T5, Canada), and Liming Dai (Univ. of Regina)

Quantification of speech privacy in closed spaces is dependent on several factors: attenuation and absorption characteristics of the construction, background sound levels, speaking levels of the talker, the language and dialect, as well as accent used in the speech. Speech privacy is the opposite concept of speech intelligibility and can be assessed by the predictors of speech intelligibility. In this paper, several assessment methods of speech intelligibility are introduced. They are adapted to measure the speech privacy in closed offices, including standard ASTM E1130, speech transmission index (STI) and early-to-late ratio (clarity). The experimental results are used to evaluate the speech privacy index (PI), which is the rating number proposed for assessing the speech privacy. The subjective measurements offered by human subjects are also conducted using different languages: English and Mandarin Chinese. The test materials used in the speech intelligibility testing are single words (MRT words), sentences and conversations. Both measurements are described under the same acoustical environment. The results of the subjective and objective measurements indicate that the current standard needs some modifications when it is used in closed offices.

Session 5aBB

Biomedical Ultrasound/Bioresponse to Vibration and Physical Acoustics: Brad Sturtevant Memorial Session in Lithotripsy I

Pei Zhong, Chair

Department of Mechanical Engineering and Materials Science, Duke University, 1 Science Drive, Box 90300, Durham, North Carolina 27708-0300

Chair's Introduction—8:00

Invited Papers

8:05

5aBB1. Brad Sturtevant's contributions to research on shock waves and lithotripsy. David T. Blackstock (Dept. of Mech. Eng. and Appl. Res. Labs., Univ. of Texas, Austin, TX 78712-1063)

Bradford Sturtevant (1933–2000) spent his entire academic career in the Graduate Aeronautical Labs at Caltech (GALCIT). He was a distinguished and internationally respected member of the fluid mechanics community, known particularly for his research on shock waves. Although a member of the Acoustical Society since 1974, he rarely if ever published in JASA. Nevertheless, his contributions to acoustics, mainly in what we think of as nonlinear acoustics although he probably saw them simply as part of fluid mechanics, are significant. Examples are: shock focusing, nonlinear resonance in closed- and open-end tubes, sonic booms, and propagation of weak shocks through a random medium. In the late 1980s he became interested in lithotripsy. When Andrew Evan (Indiana University School of Medicine) organized a multidisciplinary consortium to solve many of the mysteries of lithotripsy, it is not surprising that Brad was picked for his expertise on shock waves and his interest in shock waves in the body. The research effort that ensued, now in its eighth year, owes much to the work of Brad and his students and to his counsel. Brad was especially concerned about the mechanism(s) by which stone breakup occurs.

8:25

5aBB2. Pretreatment with low-energy (12 kV) shockwave lithotripsy (SWL) protects kidney from subsequent high-energy application. Andrew Evan, Lynn Willis, Bret Connors, Philip Blomgren (Dept. of Anatomy and Cell Biol., Indiana Univ. School of Medicine, 635 Barnhill Dr., Indianapolis, IN 46202), Youzhi Shao (Indiana Univ. School of Medicine, Indianapolis, IN 46202), and James Lingeman (Methodist Res. Inst., Indianapolis, IN 46220)

A clinical dose of 2000 shockwaves applied at 12 kV induces renal vasoconstriction but causes no hemorrhagic lesion. Two thousand shockwaves applied at 24 kV causes the same vasoconstriction as 12 kV but lesion size is increased. These findings led us to ask whether lesion size would be smaller after 2000 shockwaves at 24 kV to one pole if a vasoconstrictive state was induced in that kidney by first applying 2000 shockwaves at 12 kV to the lower pole of that kidney. Anesthetized pigs received 2000 shockwaves to the lower pole followed by 2000 shockwaves to the upper pole of the same kidney. Structural and functional analyses were performed 4 hours post-SWL. Results from these experiments show renal vasoconstriction induced from the first application of SWL at 12 kV to one pole appears to limit bleeding/hemorrhage caused by 24 kV to the other pole.

8:45

5aBB3. *In vivo* quantitation of regional versus systemic SWL-induced free-radical activity. Brian Auge, Ravi Munver, John Kourambas, Glenn Preminger (Duke Univ. Medical Ctr., Div. of Urologic Surgery, 305 Baker House, DUMC Box 3167, Durham, NC 27710), and Pei Zhong (Duke Univ., Durham, NC 27708)

Renal tissue injury in shock wave lithotripsy (SWL) may be attributed directly to vascular trauma and indirectly to ischemia-reperfusion with resultant free-radical formation. We have employed a microdialysis system to assess areas of renal injury in relation to the site of SWL treatment. Swine were assigned to a control group ($N=3$) and a SWL-treated group ($N=3$). Microdialysis probes were inserted into the parenchyma of the lower pole, upper pole of right kidney, and the lower pole of left kidney. Seventy-two hours following probe insertion, baseline dialysate samples were collected, after which SWL was administered exclusively to the right kidney lower pole in the treated group, using a Dornier Compact S lithotripter at its highest intensity. Dialysate samples were collected every 1000 shocks. It was found that SWL caused a significant increase in free-radical activity, as measured by lipid peroxidation, in the right lower pole, followed by a moderate increase in the right upper pole. Yet, no free-radical activity was noted in the left kidney. These findings suggest that SWL-induced free-radical activity is localized and confined to the ipsilateral kidney. There does not appear to be systemic circulating nephrotoxic factors following SWL treatment. [Work supported by NIH.]

9:05

5aBB4. Cell lysis due to shear in shock wave lithotripsy. Murtuza Lokhandwalla, Bradford Sturtevant (Grad. Aeronautical Labs., California Inst. of Technol., Pasadena, CA 91125, murtuzalok@hotmail.com), James McAteer, and James Williams, Jr. (Indiana Univ. School of Medicine, Indianapolis, IN 46202-5120)

Injury to isolated red blood cells (RBCs) and deformation of aluminum foils due to focused shock waves in a cavitation-free environment were investigated. The lithotripter-generated shock wave was refocused by a parabolic reflector. This refocused wave field had a tighter focus (smaller beamwidth and a higher amplitude) than the lithotripter wave field, as characterized by a membrane hydrophone. Cavitation was eliminated by applying overpressure to the fluid. Aluminum foils were used to study shock wave damage and had distinct deformation features corresponding to exposure conditions, i.e., pitting and denting accompanied with wrinkling. Pitting was eliminated by high overpressure ($\sim$12 MPa) and so was due to cavitation bubble collapse, whereas denting and wrinkling were caused by the reflected shock wave refocused by the parabolic reflector. RBCs suspended in phosphate-buffered saline (PBS) were exposed to the reflected wave field from a parabolic reflector and also from a flat reflector. Exposure to the wave field from the parabolic reflector increased hemolysis fourfold compared to untreated controls and was twice that of cell lysis with the flat reflector. Thus, hemolysis was shown to be directly related to the shock-strength gradient and validates shearing as a cell lysis mechanism in SWL. [Work supported by NIH Grant No. P01 DK43881.]

9:25

5aBB5. Localized detection of cavitation *in vivo*. Michael R. Bailey, Lawrence A. Crum, Nathan Miller, Lisa N. Couret (Ctr. for Industrial and Med. Ultrasound, Appl. Phys. Lab., Univ. of Washington, Seattle, WA 98105), Oleg A. Sapozhnikov, Yuri A. Pishchalnikov (M. V. Lomonosov Moscow State Univ., Moscow, Russia), George Keilman (Sonic Concepts, Woodinville, WA 98072), James A. McAteer, Bret Connors, and Andy P. Evan (Indiana Univ. School of Medicine, Indianapolis, IN 46202)

A step in determining cavitation's deleterious role in shock wave lithotripsy (SWL) is spatially correlating cavitation and tissue damage: Can cavitation be detected in kidney tissue? A system for localized cavitation detection and tests is described. Our lithotripter source and two single-element ring transducers (1.1 MHz, focal length 6.3 cm, radius 1.1–3.5 cm) were aligned confocal and orthogonal. An ultrasound scanhead fit in the central opening. *B*-mode imaging and fluoroscopy showed the position of the focus within the kidney. Hyperechoic clouds in the image indicated pockets of bubbles during SWL. A coincidence-detection algorithm and the confocal transducers made it possible to localize cavitation to within an 8-mm^3 region. Cavitation was detected in a pig kidney during SWL. Following SWL, the tissue region interrogated can be marked with a lesion produced by using the detection transducers as high-intensity focused ultrasound (HIFU) sources. This technique allows one to determine if the site of cavitation detection was located within the kidney tissue. Lesions were produced in pig kidney. Ultimately a ring of lesion might be drawn around the interrogation region so that SWL damage to the region might be assessed and coorelated with cavitation. [Work supported by NIH DK43881 and DK55674.]

9:45

5aBB6. Suppression of large intraluminal bubble expansion in SWL. Pei Zhong and Yufeng Zhou (Dept. of Mech. Eng. and Mater. Sci., Duke Univ., P.O. Box 90300, Durham, NC 27708)

Large, rapid expansion of intraluminal bubbles has recently been demonstrated as a potential mechanism for the rupture of capillaries and small blood vessels in shock wave lithotripsy (SWL). To reduce vascular injury without compromising the fragmentation capability of a lithotripter, we developed a strategy to modify the ellipsoidal reflector of an HM-3 lithotripter so that following each spark discharge a weak compressive pulse is generated to superimpose onto the tensile component of a lithotripter shock wave. This modified waveform was found to significantly suppress bubble expansion; and the propensity of vascular injury, as assessed by the rupture of a cellulose vessel phantom, is reduced by an order of magnitude both at 20 and 24 kV. At 24 kV, stone fragmentation is reduced by about 30% using the modified reflector, which, however, is comparable to that produced by the standard reflector at 20 kV. These results suggest that the modified reflector may be used to produce sufficient stone comminution with much reduced potential for vascular injury in SWL. [Work supported by NIH.]

10:05

5aBB7. Generation of very high-pressure pulse using time reversal in a solid waveguide: Application to lithotripsy. Gabriel Montaldo, Philippe Roux, Arnaud Derode, Mathias Fink (Laboratoire Ondes et Acoustique, ESPCI, 10 rue Vauquelin, 75005 Paris, France), and Carlos Neigreira (Inst. De Fisica, Montevideo, Uruguay)

The use of a piezoelectric transducers array has opened the possibility of electronic steering and focusing the beam to track the stone. However, due to the limited pressure delivered by each transducer (typically 10 bars), the number of transducers needed to reach the focus an amplitude of the order of 1000 bars is typically of some hundred of elements. Here, we present a new solution that combines the use of the time reversal method with a small number of transducers that generate low-amplitude waveforms in a solid waveguide to obtain the shock wave of very high amplitude in tissues located at the front of the waveguide. We use the fact that, due to time reversal invariance, for every burst of ultrasound diverging from a source and reflected by the boundaries of the waveguide,

there exists in theory a set of waves that retraces all of these complex paths and converges in synchrony, at the original source. As the temporal dispersion of the waveguide increases, the time reversed wave is temporally recompressed with a stronger amplitude amplification. Such an effect allows a very high spatio-temporal recompression to be reached and to obtain a small number of transducers the amplitude needed to break stones.

10:25–10:45 Break

Contributed Papers

10:45

5aBB8. Measurements of the pressure field and *in vitro* stone fragmentation of a Storz Modulith SLX lithotripter. Robin O. Cleveland, Parag V. Chitnis (Dept. of Aerosp. and Mech. Eng., Boston Univ., Boston, MA 02215), Ronald Anglade, and Richard K. Babayan (Boston Univ. School of Medicine, Boston, MA 02118)

The Storz Modulith SLX electromagnetic lithotripter can produce peak pressures higher than most other clinical lithotripters. The pressure field of a clinical SLX was measured with a PVDF membrane hydrophone: the peak-positive pressure varied from 10–115 MPa and the peak-negative pressure from -5 to -15 MPa. The focal spot, based on peak positive pressure, was a cigar shaped volume 25 mm long by 5 mm in diameter. The fragmentation performance was determined by treating artificial kidney stones (cylinders, 9.5 mm long and 6.5 mm diameter, made from Ultracal-30 gypsum) with 400 SWs. After treatment, stone remnants greater than 2 mm in diameter were dried and weighed. We found that fragmentation was very sensitive to correct positioning of the stone. Displacement from the focus by 4 mm or more in the horizontal plane led to a statistically significant decrease in fragmentation ($p<0.03$). In the vertical direction a displacement of more than 10 mm was necessary to produce a statistically significant reduction in fragmentation. We also found that a slower firing rate of 1 Hz was more effective than 2 Hz ($p<0.033$) indicating that shock waves can be shielded by cavitation bubbles. [Work partially supported by NIH through P01-DK 43881.]

11:00

5aBB9. Time-lapse assessment of SWL shock wave damage to kidney stones using micro-computed tomography (μCT). Robin O. Cleveland (Dept. of Aerosp. and Mech. Eng., Boston Univ., Boston, MA 02215), Ralph Müller (Inst. for Biomed. Eng., Swiss Federal Inst. of Tech. (ETH), Zürich, Switzerland), James C. Williams, and James A. McAteer (Indiana Univ. School of Med., Indianapolis, IN 46202)

SWL could be improved with a better understanding of the mechanisms of stone failure. We used μCT to visualize the initiation and progression of fracture in COM stones in vitro and to determine if the stone's environment influences its mechanism of failure. A desktop μCT (17 μm resolution) was used to collect time-lapse images of the internal and surface structure of kidney stones during administration of SWs using an electrohydraulic lithotripter (20 kV, 1 Hz). Stones were either constrained by water-saturated gauze or were surrounded by water. The constrained stone showed first evidence of fracture on the distal surface at 200 SWs. Subsequent SWs dislodged a slab from the distal side (spall) and separated the stone's core and shell. In contrast, a stone surrounded by water showed damage at the proximal surface at only 10 SWs. The data imply that the failure mechanism can vary depending on the conditions surrounding a stone. When cavitation was restricted, damage by spall predominated. With sufficient free fluid surrounding the stone, cavitation caused damage at the proximal surface. [Partially supported by NIH through P01-DK 43881 and the M. E. Müller Professorship in Bioengineering at Harvard Medical School.]

11:15

5aBB10. Effect of SW rate on stone fragmentation *in vivo*. James E. Lingeman, Ryan F. Paterson, David A. Lifshitz (Methodist Hospital Inst. for Kidney Stone Disease, 1701 N. Senate Blvd., Indianapolis, IN 46202, jlingeman@clarian.com), James A. McAteer, James C. Williams, Jr., Drew L. Rietjens, Bret A. Connors, and Andrew P. Evan (Indiana Univ. School of Medicine, Indianapolis, IN 46202)

The trend toward ungated SWL means that SW's are often delivered at a fast rate ($\geqslant 2$ Hz). This is a concern, since rapid SW delivery causes increased renal injury in experimental animals. Also, *in vitro* studies show that slowing the SW rate (<2 Hz) improves stone fragmentation. In the current study, the effect of SW rate on stone comminution was tested in an *in vivo* model. Gypsum stones were inserted into both kidneys of 100 lb pigs via percutaneous access. SWL was performed (Dornier HM3, 400 SW's uninterrupted, 20 kV, 2 Hz or 0.5 Hz), and stone fragments were collected and sieved through 2 mm mesh. The percent of stone weight lost (particles <2 mm), was greater in stones shocked at the slower rate, $55.1\pm4.9\,\%$ ($n=6$), compared to the faster rate, $19.4\pm5.6\,\%$ ($n=6$) [$p<0.0008$]. These data indicate that slowing the SW rate improves the efficiency of stone fragmentation *in vivo*. It is observed *in vitro* that bubbles along the SW axis are more numerous with faster SW rate; this could impose a barrier to transmission of SW energy into the stone. Localization of the site of rate-dependent bubble activity could provide clues to help improve SWL. [Work supported by NIH Grant P01 DK43881.]

11:30

5aBB11. Compact piezoelectric transducers for lithotripsy. Thomas Dreyer, Rainer E. Riedlinger (Inst. fur Hoechstfrequenztechnik und Elektronik/Akustik, Univ. of Karlsruhe, Kaiserstr. 12, D-76128 Karlsruhe, Germany, thomas.dreyer@etec.uni-karlsruhe.de), Edgar Bauer, and Werner Krauss (Richard Wolf GmbH, Pforzheimer Str. 32, D-75438 Knittlingen, Germany)

Self-focusing piezoelectric transducers are widely used in extracorporeal shockwave lithotripsy. This application demands high focal pressure pulse amplitudes, small focal beamwidths and large penetration depths, which results in large transducers with typical dimensions of 350-mm focal distance and 500-mm diameter. Against the background of the development of smaller and therefore more economical devices it is still desirable to reduce the transducer size but keep the acoustic properties. This can be achieved only by an increased initial pressure amplitude at the transducer surface. Since the single piezoelectric elements are driven close to their electrical limit, a different mechanical design is necessary. Piezoelectric elements placed in a layered manner on both sides of a spherical cap could provide a significant increase of the pressure pulse amplitude generated at the transducer surface. Three-dimensional FEM simulations of the transducer structure including all important design parameters, like the electrical drive for both sides, show the ability of generating higher amplitudes by keeping similar pulse shapes. Measurements made on a prototype transducer with a diameter of 270 mm and a focal distance of 200 mm show comparable results to the standard single-layered design regarding peak focal pressure amplitude and spatial amplitude distributions.

Session 5aMU

Musical Acoustics: General Topics

Uwe J. Hansen, Chair
Physics Department, Indiana State University, Terre Haute, Indiana 47809

Contributed Papers

8:30

5aMU1. Perception of single octave-related complexes in relation to the tritone paradox. Frank Ragozzine (Dept. of Psych., Southwest Missouri State Univ., 901 S. National Ave., Springfield, MO 65804, frr807f@smsu.edu)

In the tritone paradox, listeners are presented with a sequential pair of octave-related complexes separated by an interval of half an octave. Listeners tend to perceive these stimuli such that half of the pitch class circle is heard as higher, and the opposite half is heard as lower. It is theorized that listeners compare these tones to a circular pitch class template, having an orientation with respect to height that differs among listeners, in making these relative height judgments. In the present experiment, listeners were presented with two sets of stimuli. In the first, listeners were presented with the tritone paradox stimuli. In the second, listeners were presented with a single octave-related complex (rather than a pair of such tones) on each trial, and were asked to determine whether each tone was high or low without comparison to any other tone. An influence of pitch class on perceived height was shown for both tritone and single tone stimuli, and a significant relationship between perception of the tritone paradox and perception of the single octave-related complexes was obtained. These results are in accord with the pitch class template theory.

8:45

5aMU2. Regulating glottal airflow in singing: Application of the maximum power theorem. Ingo R. Titze (Dept. of Speech Pathol. & Audiol. and Natl. Ctr. for Voice and Speech, The Univ. of Iowa, Iowa City, IA 52245)

Two competing views of regulating glottal airflow for maximum vocal output are investigated. The maximum power transfer theorem is used as a guide. A wide epilarynx tube (laryngeal vestibule) matches well with low glottal resistance (the yawn-sigh approach), whereas a narrow epilarynx tube matches well with a higher glottal resistance (the twang-belt approach). A simulation model is used to calculate mean flows, peak flows, and oral radiated pressure ratio between the vocal tract (the load) and the glottis (the source). Results show that when the ratio approaches 1.0, maximum power is transferred and radiated from the mouth. [Work supported by NIH.]

9:00

5aMU3. Vibrational modes and sound spectra of bass steelpans. Thomas D. Rossing (Phys. Dept., Northern Illinois Univ., DeKalb, IL 60115) and Uwe J. Hansen (Indiana State Univ., Terre Haute, IN 47809)

The vibrational modes and sound spectra from bass steelpans by two well-known makers are reported. The large note areas generally radiate a strong fundamental along with an octave and a twelfth. The (1,0) mode is most often tuned to the octave, and the (0,1) to the twelfth, but in a few notes these roles are reversed. In some notes, the partial tuned to the twelfth dies out rather slowly, and gives an audible ''aftersound.'' Modes of vibration in the skirt are made up of bending waves that travel around the circumference. The two circular stiffening ribs appear to have a rather noticeable effect on the vibrational modes of the skirt.

9:15

5aMU4. Measuring residual stresses in a steelpan. Andres Peekna (Innovative Mechanics, Inc., 265 Coe Rd., Clarendon Hills, IL 60514-1029) and Thomas D. Rossing (Northern Illinois Univ., DeKalb, IL 60115)

By bonding strain gage rosettes to a steelpan and drilling a small hole, it is possible to measure the residual stresses at selected locations on a steelpan. The accuracy of the method depends upon the extent to which the stresses and strains change with depth below the surface on which the strains are measured. The residual stresses in a tenor steelpan, ''sunk'' by hammering in the usual way, appear to be mostly compressive in the middle ring of notes, changing to tensile in the outer ring. The effect of heating on the residual stresses is discussed.

9:30

5aMU5. Normal modes of a finite element lip reed model. Daniel O. Ludwigsen and William J. Strong (Dept. of Phys. and Astron., N283 ESC, Brigham Young Univ., Provo, UT 84602, dan_ludwigsen@byu.edu)

Models of the lip reed for brass instruments show utility in time-domain simulation with one or two degrees of freedom. Higher-dimensional models may hold further insight into the actual behavior of lip structures in playing conditions. The present model represents a mid-sagittal slice of lip tissue using ANSYS, a commercial software package for solid modeling and finite element analysis. Modes of the lip model structure provide a first step in understanding its response to pressure loads on the surface in a subsequent time-domain simulation of the player/trombone system. The dependence of the mode frequencies and mode shapes on muscle tissue stiffness can help to inform stiffness parameter selection for simulation, and may offer clues to understanding how the player controls the embouchure.

9:45

5aMU6. Symmetries and dynamics of musical scales. Alpar Sevgen (Dept. of Phys., Boğaziçi Univ., Bebek 80815, Istanbul, Turkey)

Configurational potential energy, as well as complexity, may be used as a measure for musical scales. To show this, complexities $C(\mathbf{n})$ and energies $E(\mathbf{n})$ of equally tempered scales with the interval structure $\mathbf{n}=\{n_1,n_2,\ldots,n_M\}$ for $N=12$ semitones and $M=7$ notes are compared. Complexity is the sum of sharps and flats in a scale. Energy $E(\mathbf{n})$ is due to an assumed interaction energy n^{α} between the notes. It is found that the symmetries of $C(\mathbf{n})$ and $E(\mathbf{n})$ are the same: rotations and reflection (i.e., cyclical permutations and reverse ordering, respectively, of the components of $\mathbf{n}$), and note↔no-note transformation (e.g., major↔pentatonic). But beyond that, the complexity and energy models are apparently unrelated. Symmetries explain the groupings of scales but not the differences in complexity or energy of the groups. There is, however, a *remarkable agreement* between the $C(\mathbf{n})$ and $E(\mathbf{n})$ tables for a range of values of α. (As an example, the harmonic oscillator value $\alpha=2$ is employed.) That these tables display an almost identical pattern in the ordering of 38 groups comprising 462 scales of 12-tuples implies that energies $E(\mathbf{n})$ can also be used as a measure for musical scales.

5a FRI. AM

10:00–10:15 Break

10:15

5aMU7. Active control and adaptive structures for a smart violin. Sathya Hanagud and Xia Lu (School of Aerosp. Eng., Georgia Inst. of Technol., Atlanta, GA 30332-0150)

String instruments, like the violin, have been studied for the past 350 years. These studies have focused on the quality of the wood, dynamic characteristics of these instruments and structural acoustics and psychoacoustics. The emphasis of the previous research was to understand the secrets of great designers of these instruments. For example, one would like to explain reasons for the legendary quality of the sound radiated by a Stradivarius and the price that it commands today. Many researchers have also suggested passive techniques to improve the quality of these musical instruments. However, no attempts have been made to use active structural control and smart structures to modify the quality of the music radiated by these instruments like the violin. In this paper, we start with the assumption that the legendary quality of the sound radiated by a Stradivarius is due to its special structural dynamic or structural-acoustics characteristics. Our approach is to use the concepts of adaptive or smart structures and active structural control to improve the quality of the sound radiated by these musical instruments. Procedures are developed to change the structural dynamic characteristics, actively, to replicate the structural dynamic characteristics of musical instruments developed by great masters.

10:30

5aMU8. Improvement of the concert harp through the application of composites. Curt Preissner and Thomas J. Royston (Univ. of Illinois at Chicago, Chicago, IL 60607, troyston@uic.edu)

The replacement of a Sitka spruce grand concert harp soundboard with a fiber-reinforced plastic soundboard was considered for improved durability and long-term stability. First, a means of nondestructively identifying the structural acoustic properties of a conventional wooden soundboard was developed based on a multi-mode dynamic identification technique available in the literature. This was followed by the development of a design process to reproduce the same structural acoustic properties in the composite replacement. It was found that specific matching criteria must be met to achieve a dynamic response that closely approximates the wooden soundboard. Lay-up of the composite soundboard was guided by this matching criteria and Tsais orthotropic plate design procedure. Experimental verification was then performed utilizing the same system used to identify the wooden plate properties. The identification technique and composite replacement design process may be applicable to other musical instruments, as well as other nonmusical, orthotropic plate structures.

10:45

5aMU9. An acoustical study on double bass bridge height adjusters. Andrew W. Brown (Institut füer Wiener Klangstil, Univ. of Music and Performing Arts, Singerstr. 26a, A-1010 Vienna, Austria, brown@mdw.ac.at)

A study on double bass bridge height adjusters was made using local and international surveys, a listening test survey, and sound spectrum analysis to learn more about the acoustical properties of various bridge adjusters and their current use. A questionnaire sent via e-mail to members of the International Society of Bassists confirmed that preferences and opinions about adjusters vary widely, though some general tendencies are recognizable. Responses also gave some indications why approximately 60 to 80 bassists use bridge adjusters, while adjusters in central Europe are rarely seen. Computerized listening tests showed that all types of bridge height adjusters cause an audible difference in sound compared to a bridge with no adjusters, and that individual models and materials have unique tonal characteristics. Finally, a digital sound analysis was made comparing a bridge with no adjusters to the same bridge fitted successively with six types of bridge height adjusters. The analysis showed that bridge height adjusters make a significant difference in pizzicato decay time depending on the model and material in question, and FFT spectrum analysis of bowed tones showed the sound characteristics of bridge height adjusters throughout the range of the double bass.

11:00

5aMU10. Normal modes of vibration in balalaikas. Andrew Morrison and Thomas D. Rossing (Phys. Dept., Northern Illinois Univ., DeKalb, IL 60115)

Using holographic interferometry, we have studied the vibrational behavior of a prima balalaika, a secondo balalaika, and an alto balalaika, all constructed at the Folk Musical Instrument Manufacturing Company in St. Petersburg, Russia. The modes of lowest frequency are characterized by the in-phase motion of the soundboard, while at higher frequencies we observe modes with longitudinal and transverse nodal lines where the soundboard pivots as it vibrates, much like the guitar. The modal frequencies are compared with the sound spectra recorded when the strings are plucked.

11:15

5aMU11. Traditions versus acoustics in the evolution of the Baltic psaltery. Ain Haas (Dept. of Sociology, Indiana Univ.–Indianapolis (IUPUI), Indianapolis, IN 46202)

The Baltic psaltery figures prominently in the musical heritage, folklore, and national symbolism of Finns, Estonians, Latvians, and Lithuanians. Despite its sacred aura, instrument makers have had considerable latitude in design. Positing the ancient Northern European lyre as the likely source of inspiration for the Baltic psaltery, the paper analyzes how acoustical considerations seem to have affected the transformation of the instrument. Archeological finds, manuscript illustrations, ethnographic museum specimens, and instrument reconstructions are examined for clues to the likely sequence of forms. Cultural interaction between Finnic and Indo-European peoples is presented as a major factor in opening the way for acoustically driven experimentation of the design of a traditional instrument.

11:30

5aMU12. Acoustical measurements on carved Baltic psalteries. Andres Peekna (Innovative Mechanics, Inc., 265 Coe Rd., Clarendon Hills, IL 60514-1029) and Thomas D. Rossing (Northern Illinois Univ., DeKalb, IL 60115)

Modes of vibration in several successful replicas of carved Estonian kannels, an outstanding Latvian–American kokle, and a reconstruction of a Novgorad gusli were determined by electronic TV holography. From the characteristics of the various resonances, as well as the sound of the instruments, we conclude that a successful Baltic psaltery should have: (1) a maximum number of resonances within the tuning range; (2) a strong resonance between the keynote and the note above; and (3) a strong resonance also at the low dominant. Not only the overall soundhole area but their locations can have significant effect on sound quality and volume.

11:45

5aMU13. Tuning the lower resonances of carved Baltic psaltery by adjusting area of the soundholes. Andres Peekna (Innovative Mechanics, Inc., 262 Coe Rd., Clarendon Hills, IL 60514-1029) and Thomas D. Rossing (Northern Illinois Univ., DeKalb, IL 60115)

A 12-string kannel was constructed by Ain Haas, initially without any soundholes. Subsequently, an initial soundhole pattern was drilled and then enlarged incrementally. The frequencies of the two lowest dipole modes changed significantly, from 382 and 492 Hz to 411 and 454 Hz, respectively. Introducing distributed soundholes appears to affect the internal air motion. The mode damping was also observed to increase as soundholes were added, although damping of the lowest symmetric mode decreased upon adding holes of larger diameter. The sound quality was judged to improve as more soundholes were added.

Session 5aPP

Psychological and Physiological Acoustics: Lateralization Localization, and Spatial Perception

Raymond H. Dye , Jr., Chair
Parmly Hearing Institute, Loyola University of Chicago, 6525 North Sheridan Road, Chicago, Illinois 60626

Contributed Papers

9:00

5aPP1. Transposed stimuli reveal similar underlying sensitivity to interaural timing information at high and low frequencies. Leslie R. Bernstein and Constantine Trahiotis (Dept. of Neurosci. and Dept. of Surgery (Otolaryngol.), Univ. of Connecticut Health Ctr., Farmington, CT 06030, Les@neuron.uchc.edu)

Threshold interaural temporal disparities (ITDs) obtained at high frequencies are typically much larger than those obtained at low frequencies. Colburn and Esquissaud [J. Acoust. Soc. Am. Suppl. 1 **59**, S23 (1976)] attributed this to differences in the peripheral, monaural processing of the stimuli. They hypothesized that neural impulses synchronized to the waveform at low frequencies and to only the envelope at high frequencies serve as inputs to a binaural processor that functions similarly at low- and high-signal frequencies. This hypothesis was supported in ITD-discrimination experiments employing novel, high-frequency ''transposed stimuli'' designed to provide high-frequency channels with envelope-based information that mimics information that is normally available only in low-frequency channels. That is, when the inputs to the high-frequency binaural channels were provided with information like that normally available at low frequencies, sensitivity to ITD in the two spectral regions was found to be similar. The data at both low and high frequencies were well accounted for, quantitatively, in terms of a constant-criterion change in normalized interaural correlation computed subsequent to bandpass filtering. compression, rectification, and low-pass filtering. [Work supported by NIH.]

9:15

5aPP2. Sound localization by interaural time differences at high frequencies. Zachary A. Constan (Phys. and Astron., Michigan State Univ., East Lansing, MI 48824) and William M. Hartmann (Phys. and Astron., Michigan State Univ., East Lansing, MI 48824)

It is possible for human listeners to lateralize high-frequency noise on the basis of interaural time differences in the fine structure of the envelope. However, the ability to do so makes great demands on the interaural coherence of the noise. This paper explores listeners' ability to lateralize a broadband, high-passed, coherent-noise signal in the presence of a broadband incoherent masker. Results showed that as the high-pass cutoff frequency increased through the critical region from 1 to 4 kHz, the required interaural coherence increased rapidly, especially for the smaller changes in lateral position. The observed functional behavior can be successfully predicted by a neural model of lateralization based on the centroids of bandwise cross-correlation functions of model peripheral inputs that have been rectified and low-pass filtered [e.g., Bernstein and Trahiotis, J. Acoust. Soc. Am. **100**, 1754–1763 (1996)]. The results have implications for the localization of broadband sounds in rooms, where the interaural coherence tends to increase with increasing frequency, but often not rapidly enough to allow the envelope timing information to contribute usefully. [Work supported by NIDCD.]

9:30

5aPP3. Temporal resolution in an interaural temporal difference discrimination task with distracter stimuli. Eric A. Erpenbeck and D. Wesley Grantham (Vanderbilt Bill Wilkerson Ctr., Dept. of Hearing and Speech Sci., VUMC, 1114 19th Ave. S., Nashville, TN 37212, d.wesley.grantham@vanderbilt.edu)

Interaural temporal difference (ITD) thresholds were determined for brief (14-ms) 500-Hz tone bursts in a two-interval forced-choice paradigm in which the interstimulus interval was 36 ms. Preliminary data indicate that the ITD threshold can be affected by the presence of a third (distracter) burst temporally centered between the two primary bursts on each trial: When the distracter has the same ITD as the final burst of the pair, the ITD threshold is unchanged or decreases relative to in the no-distracter condition; when it has the same ITD as the first burst, the threshold increases. The results are consistent with the hypothesis that, when presented with a rapid three-burst triad, the binaural system processes the ITD of the first burst normally (i.e., as if presented in isolation), but confuses or averages the ITDs of the latter two bursts. Follow-up experiments will measure the effects of the discrete nature of the distracter (i.e., by removing the silent gap between the distracter and the burst that shares its ITD), and the frequency of the distracter on the magnitude of its effect. Results will be discussed in terms of binaural temporal resolution of onset versus post-onset information. [Supported by NIDCD.]

9:45

5aPP4. Detection of amplitude modulation and discrimination of interaural differences in modulation phase for high-frequency carriers. Mark A. Stellmack and Neal F. Viemeister (Dept. of Psych., Univ. of Minnesota, Minneapolis, MN 55455, stell006@umn.edu)

Using a 4-kHz pure-tone carrier and modulation frequencies (f_m) of 2–512 Hz, thresholds for monaural detection of sinusoidal amplitude modulation (SAM) were measured, as were threshold modulation depths for the discrimination of interaurally in-phase and out-of-phase SAM envelopes. For one listener, interaural differences in envelope phase for f_m up to 256 Hz were discriminable at modulation depths (m) for which the monaural modulation itself was just detectable, and all monaural and binaural thresholds were within a range of 5 dB (20 log m) across f_m. For a second listener, binaural thresholds increased more rapidly with increasing f_m than monaural thresholds. Another experiment examined the notion that interaurally delayed SAM envelopes are processed as time-varying interaural level differences. In this task, listeners discriminated interaurally in-phase SAM envelopes from SAM envelopes carrying a fixed interaural phase difference (IPD). Threshold m was measured for several f_m and IPD combinations. For a given IPD, thresholds were either constant or decreased with increasing f_m. The data suggest that dynamically varying interaural level differences are the cue to discrimination at low f_m, while the fixed interaural time difference is the likely cue at high f_m. [Work supported by NIDCD DC00683.]

10:00

5aPP5. Peripheral auditory processing and the precedence effect. Klaus Hartung and Constantine Trahiotis (Dept. of Neurosci. and Dept. of Surgery (Otolaryngol.), Univ. of Connecticut Health Ctr., Farmington, CT 06030, tino@neuron.uchc.edu)

This work addresses how a consideration of peripheral auditory processing can help one to understand the relative salience of binaural information conveyed by successive binaural transients in precedence experiments. It appears that much of the variability in the data is amenable to an explanation based on peripheral interactions that result from auditory filtering and the functioning of hair cells in combination with a binaural model based on cross correlation. This approach does not include inhibitory mechanisms that are commonly considered as being necessary in order to account for the precedence effect. [Work supported by NIH.]

10:15

5aPP6. Observer weighting of interaural delays in echo clicks preceded by source clicks that have been attenuated. Raymond Dye, Jr., Jose Gallegos, and Christopher Brown (Parmly Hearing Inst., Loyola Univ. Chicago, 6525 N. Sheridan Rd., Chicago, IL 60626, rdye@luc.edu)

This investigation examined the effect that the relative levels of leading and lagging clicks have on binaural precedence. A diotic pulse was presented during the first interval to mark the intracranial midline. In the second interval, two dichotic pulses were presented, separated by an ''echo delay'' ranging from 1 to 32 ms. The interaural delays of the two pulses were independently selected from a Gaussian distribution. Listeners were instructed to respond according to the laterality of the lagging click. The leading click was attenuated by from 0 to 30 dB. Performance was measured by proportion correct, relative echo weight, and the proportion of responses predicted by the derived weights. At the shortest echo delays, the effect of attenuation on echo weight and proportion correct occurred after merely 6 dB of attenuation. At echo delays longer than 8 ms, the echo weights started out higher but were not as dependent upon source attenuation. The same effect was found for proportion correct. At 4 ms, there were steady increases in both echo weight and proportion correct as the source was attenuated. The proportion of responses accounted for by the weights was dependent upon neither source attenuation nor echo delay. [Work supported by NIDCD.]

10:30–10:45 Break

10:45

5aPP7. Effect of reverberation on spatial unmasking for nearby speech sources. Barbara G. Shinn-Cunningham (Hearing Res. Ctr., Depts. of Cognit. and Neural Systems and Biomed. Eng., Boston Univ., 677 Beacon St., Boston, MA 02215, shinn@cns.bu.edu), Lisa Mraz, and Norbert Kopčo (Hearing Res. Ctr., Boston Univ., Boston, MA 02215)

Individualized HRTFs were measured in a moderately reverberant room ($T60 = 450$ ms) for sources directly in front of and to the right of a listener for both near (15 cm) and far (1 m) distances. The full HRTFs (including reverberation) and pseudo-anechoic HRTFs (time windowing out the reverberation) were used to simulate a speech target and a speech-shaped noise masker over headphones. Speech reception thresholds were measured adaptively, varying the target level while keeping the masker level constant at the better ear. Thresholds were measured for both left and right monaural signals as well as for binaural signals. Results show the magnitude of spatial unmasking that can arise for sources very close to the head, where large interaural level differences (ILDs) arise, and determine the degree to which spatial unmasking is due to better ear and binaural effects. These results are compared to previous anechoic results simulating sources near a listener in which large ILDs appear to degrade binaural performance below predicted better-ear performance. Comparisons between pseudo-anechoic and realistic reverberation conditions address the degree to which reverberation interferes with spatial unmasking. [Work supported by a grant from the Air Force Office of Scientific Research.]

11:00

5aPP8. The role of masking in the Franssen effect. William M. Whitmer, William A. Yost, and Stanley Sheft (Parmly Hearing Inst., Loyola Univ. of Chicago, 6525 N. Sheridan Rd., Chicago, IL 60626, wwhitme@luc.edu)

The sudden onset of a sound which slowly decays at one loudspeaker can occlude the presence and location of the same sound presented simultaneously with a slow rise time at another loudspeaker. This phenomenon, known as the Franssen effect, has been shown in previous research to be specific to low-frequency pure tones in reverberant rooms. To examine the possible mechanisms involved, listeners heard pure-tone transient/steady-state signal pairs at frequencies ranging from 250–4000 Hz from either one loudspeaker or two contralateral loudspeakers in an eight-speaker array. Signals were masked with Gaussian noise. Using a two-interval, forced-choice tracking procedure, thresholds for detecting the steady-state signals were measured. The transient tone was present in both intervals. The results showed increases across frequencies in threshold of approximately 6 dB for two-source (Franssen) conditions compared to single-source conditions. In an auxiliary experiment, signals were masked with tones of differing frequencies. For Franssen conditions, masking was dependent on the interaction of signal and masker frequencies. Both results are discussed in terms of onset dominance, nonecho suppression and free-field masking. [Work supported by NIDCD.]

11:15

5aPP9. Contralateral masking effects in dichotic listening with two competing talkers in the target ear. Douglas S. Brungart (AFRL/HECB, WPAFB, OH 45433, douglas.brungart@wpafb.af.mil) and Brian D. Simpson (Veridian, Dayton, OH 45431)

Some of the most influential experiments in selective auditory attention have been based on a dichotic cocktail party task where listeners are asked to respond to a speech signal presented to one ear while ignoring a simultaneous competing speech signal presented to the other ear. These experiments have generally shown that the intelligibility of a monaural speech signal is unimpeded by the presence of an interfering speech signal at the opposite ear. However, recent results in our laboratory indicate that listeners cannot ignore a speech signal at the unattended ear when two simultaneous speech signals are presented to the target ear. In this study, the intelligibility of a target phrase in a two-talker stimulus presented to one ear was measured monaurally and with a speech or noise signal in the opposite ear. Performance in this task was unaffected when noise was added to the unattended ear, but degraded substantially when speech was added to the unattended ear. These results suggest that there are strong interactions between the monaural processes that listeners use to segregate two spatially colocated voices and the binaural processes they use to segregate voices originating from different apparent locations in space. [Work supported by AFOSR.]

11:30

5aPP10. Measurements of the directional performance of commercial hearing aids. Robert B. Schulein, Laurel A. Christensen, and Andrew J. Haapapuro (Etymotic Research, 61 Martin Ln., Elk Grove Village, IL 60007, r_schulein@etymotic.com)

Various estimates of directional microphone performance are possible, yielding a wide variety of published and advertised signal-to-noise improvements. There appears to be a growing concensus that the ratio of on-axis sound to diffuse sound (the ''directivity index'') provides the most realistic measure for real-world use of hearing aids. We undertook to measure several commercial digital and analog hearing aids, mostly of recent design, using the methods under consideration for ANSI standard adoption: anechoic polar measurements with data numerically integrated to obtain a diffuse-field directivity index estimate and direct anechoic and reverberation-room measurements. These measurements generally agreed, and showed a wide range of performance across designs.

11:45

5aPP11. Modeling the effect of torso on the head-related transfer function via the boundary-element method. Nail A. Gumerov, Zhihui Tang, Ramani Duraiswami (Inst. for Adv. Computer Studies, Univ. of Maryland, College Park, MD 20742), Richard O. Duda, V. Ralph Algazi (Univ. of California, Davis, CA 95616), and S. T. Raveendra (Collins and Aikman Automotive Interior Systems, Plymouth, MI 48170)

The head-related transfer function (HRTF) depends on many parameters including the parameters of torso. The goal of this study is to understand and evaluate the influence of torso on the HRTF. For this purpose a simplified model consisting of a spherical head and an ellipsoidal torso was investigated using a boundary-element method. The use of this simplified model for modeling of the human body is justified for relatively low frequencies (frequencies below 1500 Hz). Solutions of a 3-D Helmholtz equation for monopole source and various positions and orientations of the body relative to the source of sound are obtained using the boundary-element method. The results are compared with analytical results of Duda and Martens [J. Acoust. Soc. Am. **104**, 3048–3058 (1998)] and computations in the absence of torso and results of Avendano *et al.* [IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, 1999] in the presence of torso. The influence of the torso on the angular and frequency dependencies of the HRTF is then identified, evaluated, and discussed. Also discussed is the influence of the torso in modifying the HRTF with the pose of the head. [Work supported by NSF.]

12:00

5aPP12. Effects of visual-feedback training in 3-D sound displays. Pavel Zahorik (Dept. of Psychol., Univ. of California, Santa Barbara, CA 93106, zahorik@psych.ucsb.edu), Clement Tam, Kenneth Wang, Philbert Bangayan, and V. Sundareswaran (Rockwell Sci. Ctr., Thousand Oaks, CA 91360)

Current low-cost 3-D sound displays do not use individualized head-related transfer functions (HRTFs) to render acoustic space. As a result, sound source localization accuracy is often degraded when compared to the accuracy using real sources, or to higher quality displays using individualized HRTFs. Here, a way to improve accuracy was examined in which listeners were provided with paired auditory and visual feedback as to the correct sound source location. Sound localization accuracy was assessed for six listeners, using a large number of virtual sound sources sampled from a spherical grid surrounding the listener, before, during, and after feedback training. Feedback training markedly improved localization accuracy compared to a control group of five listeners that did not receive training. The largest improvements in accuracy resulted from listeners' enhanced abilities to distinguish sources in front from sources behind. Further, these improvements were not transient short-term effects, but lasted at least 4 days between training and testing sessions. These results suggest that simple and relatively short periods of feedback training (two 45-min sessions) can effectively facilitate perceptual re-mapping to modified spatial cues, and therefore mitigate technical deficiencies in 3-D sound systems due to nonindividualized HRTFs. [Work supported by ARL-FedLab.]

FRIDAY MORNING, 8 JUNE 2001 CRYSTAL ROOM, 7:55 A.M. TO 12:00 NOON

Session 5aSA

Structural Acoustics and Vibration and Physical Acoustics: David G. Crighton Memorial Session II

Gideon Maidanik, Cochair

Carderock Division, Naval Undersea Warfare Center, David Taylor Model Basin, 9500 MacArthur Boulevard, West Bethesda, Maryland 20817-5000

Paul E. Barbone, Cochair

Department of Aerospace and Mechanical Engineering, Boston University, 110 Cummington Street, Boston, Massachusetts 02215

Chair's Introduction—7:55

Invited Papers

8:00

5aSA1. David Crighton and structural acoustics with mean flow. Nigel Peake (Dept. of Appl. Math. and Theoretical Phys., Cambridge Univ., Silver St., Cambridge CB3 9EW, UK, n.peake@damtp.cam.ac.uk)

David Crighton wrote a number of significant papers on structural acoustics, but the one which has provoked the most discussion is concerned with the causal response of an elastic plate to unsteady forcing in the presence of mean flow [Crighton and Oswell, Philos. Trans. R. Soc. London, Ser. A **335**, 557–592 (1991)]. A number of exceedingly unusual features arise in this problem, including absolute instability at high flow speeds, negative energy waves and the existence of neutral modes which violate the usual Rayleigh–Lighthill condition of out-going group velocity. A number of developments have been made since 1991, by Crighton himself and by Peake, and some of these will be described in this presentation. These include the effects of plate curvature and the presence of a boundary layer, both of which act to push the absolute instability to unrealistically high flow speeds; the dynamics of composite plates which allow shear deformation; and the effects of nonlinearity. In the latter regard it will be shown how negative energy waves are saturated at finite amplitude, and how solitary waves can develop over a wide parameter range.

8:30

5aSA2. Approximate factorization of Wiener–Hopf kernels. I. David Abrahams (Dept. of Math., Univ. of Manchester, Oxford Rd., Manchester M13 9PL, UK)

The key step in the solution of a Wiener–Hopf equation is the factorization of the Fourier transform of the kernel, $K(\alpha)$ (here, α is the Fourier variable). Exact factorizations are seldom useful, either because they do not exist or because they are too cumbersome to evaluate accurately. It is desirable, therefore, to construct approximations of $K(\alpha)$ which are both easy to factorize and simple to evaluate. Recently, Crighton examined the case that K depends on a small parameter, ϵ; i.e., $K=K(\alpha;\epsilon)$. Crighton showed that matched asymptotic expansions (MAE) can be used to construct a composite asymptotic approximation to K for $\epsilon\rightarrow 0$, valid uniformly over the full required range of α. The composite approximation comprises several relatively simple factors in typical applications, and these factors are then decomposed into their Wiener–Hopf factors by standard techniques. Abrahams has introduced an alternative approximation method for Wiener–Hopf kernels based on Padé approximants. Padé approximants give explicit uniform approximations to K valid in the full relevant range of α, do not require a small expansion parameter, and can be computed with relatively little effort. Although Crighton's and Abrahams's approaches seem at first glance unrelated, this talk will show that under certain circumstances, they are closely related.

9:00

5aSA3. Nonlinear acoustics: Discussions with David Crighton. Lev A. Ostrovsky (NOAA Environ. Technol. Lab., Boulder, CO 80305 and Inst. of Appl. Phys., Nizhny Novgorod 603006, Russia)

This lecture is a brief overview of Professor David Crighton's important role in nonlinear acoustical research which, as in many other areas, went far beyond his own valuable publications. He was an active organizer of symposia on Nonlinear Acoustics and relevant Euromech meetings. Also personal meetings and discussions with David G. Crighton regarding the topic are memorable for the author.

9:30

5aSA4. Computation of the sources of sound in turbulent flow. Tim Colonius (California Inst. of Technol., Pasadena, CA 91125, colonius@caltech.edu)

Direct numerical simulations of turbulent flows and their radiated acoustic fields offer a detailed description of the acoustic sources at low Reynolds number. High Reynolds number jets will remain inaccessible to direct computation for many years to come, and there is thus a great need for good models of the acoustic sources. The low Reynolds number simulations can provide insight and data for such source modeling efforts. In this brief survey of research in this area, issues that were anticipated in the research of Professor David Crighton, for whom this session has been dedicated, are discussed. These include the technical challenges that must be overcome in order to accurately compute aeroacoustic flows, the modeling of acoustic sources in jets and mixing layers as wave packets, and the scattering and refraction of sound by turbulence.

10:00–10:15 Break

10:15

5aSA5. Asymptotic theory of propeller noise. Anthony B. Parry (Rolls-Royce plc, P.O. Box 31, Moor Ln., Derby DE24 8BJ, UK)

This paper describes noise radiation from open rotors. The starting point is the expression for noise radiation due to either the blade loading distribution or the blade thickness distribution. These expressions, involving integrals over the blade surface, are themselves obtained from the Ffowcs WIlliams-Hawkings equation. Using the assumption that the number of blades is large, it is possible to evaluate the integrals asymptotically and to show precisely how the farfield radiation is dominated by contributions from a few regions of the blade. At subsonic conditions, the noise is, of course, completely tip generated. At supersonic conditions the radiation is, to leading order, dominated by the Mach radius: the radius that approaches the observer at precisely sonic speed. At supersonic conditions there is also a significant lower order term, originating at the blade tip. The results for all cases show the explicit dependence on tip Mach number, harmonic number and radiation angle. Excellent agreement is obtained between the asymptotic results and full numerical evaluation of the acoustic field. The formulas are also extended to predict the acoustic benefit of sweep.

10:45

5aSA6. Structural acoustics with mean flow: Instability waves on force excited submerged elastic structures. Sevag H. Arzoumanian (Dept. of Appl. Mathematics and Theoretical Phys., Univ. of Cambridge, Cambridge CB3 9EW, UK, sevag@damtp.cam.ac.uk)

Computational verification of the Crighton and Oswell predictions for the response of a line excited, doubly infinite, submerged plate in mean flow are presented. The phenomena highlighted for validation include: the occurrence of absolute instability above a critical flow speed, the presence of convectively unstable waves downstream from the driver for all flow velocities below this critical speed, and the existence of anomalous propagating waves with group velocities pointed towards the driver over a narrow range of excitation frequencies. Extensions of the Crighton and Oswell model to include the effects of compressibility in the fluid, finite thickness in the structure and the presence of compliant coatings are discussed. Extensions of this model to three dimensions are also presented through analytical studies and numerical simulations of instability waves on a point excited 2-D plate submerged in a 3-D fluid in mean flow. [Work supported by ONR and St. John College, Cambridge.]
906105pas

11:00

5aSA7. Hydrofoil sound generation at high Reynolds number. Carolyn Judge, Dwayne Bourgoyne, Joshua M. Hamel, Steven L. Ceccio, and David R. Dowling (Dept. of Mech. Eng., Univ. of Michigan, Ann Arbor, MI 48109-2121)

The unsteady separated turbulent flow near the trailing edge of a loaded hydrofoil is often a source of hydroacoustic noise. Intense turbulence in this region may produce noise directly, but, at low Mach number, hydrodynamically forced structural motions may also radiate sound. In fact, interactions between the foil's unsteady vortical wake flow and the hydrofoil structure may produce undesired self-excited hydrodynamic-vibratory resonances sometimes called ''singing.'' This presentation reports experimental results from a series of recent experiments focused on understanding and documenting these phenomena at chord-based Reynolds numbers up to 60 million. The measurements include foil surface static and dynamic pressures, foil vibration, LDV-determined average flow velocities and turbulence quantities, and PIV flow fields in the immediate vicinity of the foil's trailing edge. The experiments are conducted at the US Navy's Large Cavitation Channel with a two-dimensional test-section-spanning hydrofoil (2.1 m chord, 3.0 m span) at flow speeds from 0.5–18.3 m/s. Special interest is focused on the trailing edge of the foil where the unsteady Kutta condition is investigated. [Significant assistance provided by personnel from NWSC-CD. Sponsored by Code 333 of ONR.]

11:15

5aSA8. Acoustic backscattering and coupling processes for elastic and plastic circular disks in water: Direct and holographic observations. Philip L. Marston (Dept. of Phys., Washington State Univ., Pullman, WA 99164-2814) and Brian T. Hefner (Univ. of California at Davis, Davis, CA 95616)

Coupling and mode conversion processes can cause significant backscattering enhancements for scattering by tilted circular disks in water. These were investigated by direct observation of the backscattering as a function of the tilt of the disk and, in the case of elastic disks, using acoustic holography. The coupling of symmetric Lamb waves onto the elastic disk and an associated made conversion to (and from) in-plane shear waves was important for the tilted elastic disk [B. T. Hefner and P. L. Marston, Acoust. Res. Lett. Online **2**, 55–60 (2001)]. In a different range of tilts and at a sufficiently high frequency, a (higher-order) antisymmetric edge wave that propagated around the perimeter of the elastic plate was important. Tilted plastic disks displayed anomalous enhancements for tilts near (but beyond) a critical angle associated with the internal total reflection of a shear wave from the edge of the plate. Other enhancements are also present for tilted plastic plates. [Work supported by ONR.]

11:30

5aSA9. Visibility of David Crighton and ultrasonic waves. M. A. Breazeale (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, University, MS 38677)

Visibility of David Crighton is assured through his many useful publications. He was able to make his theoretical results seem simple and understandable. Visibility of ultrasonic waves is another matter. In our understanding we are aided by another Fellow of the Royal Society. Lord Rayleigh used collimated light incident on a grating. He showed that behind the grating there is a series of real images with an optical periodicity of magnitude $(2b^2)/l$ where b is the grating spacing and l is the optical wavelength. For the ultrasonic analog we substitute a standing ultrasonic wave for the grating, then we superimpose a scale on the images of the ultrasonic wave-fronts. This becomes a simple but sensitive means of measuring ultrasonic wave velocity in transparent media.

11:45

5aSA10. Scattering of sound by laminar wake of symmetric moving body. Andrew G. Semenov, Samuel A. Rybak, and Vadim N. Alekseev (N. N. Andreev Acoust. Inst., Russian Acad. of Sci., 4 Shvernik St., Moscow, 117036 Russia)

The report is based on analysis of sound propagation in the vicinity of a rather general class of localized flows induced by the motion of symmetric structures of the type of solid spheres and droplets or vortices in the liquid. Acoustic scattering on such inhomogeneities, situated inside various flows, was performed with the aid of boundary value problem solution for the M. Lighthill equation. The role of a moving body which generates an inhomogeneity as well as a role of surface at which the tangential component of flow velocity could suffer a discontinuity are demonstrated in a rather general form. Comparison of body and flow contribution to observed scattered field is presented. Flow presumable yield in the forward scattering direction situation of the type encountered in conventional underwater or biomedical tomography system, comprising the source–receiver observation line, is underlined. Details and specific features are demonstrated for well-known (classical) hydrodynamic flow types, such as potential and viscose flows around moving sphere, and especially for symmetric laminar wake. Scattered sound angular structure and frequency dependencies are derived. The influence of flow Mach and Reynolds number value ranges on sound scattering intensity are discussed.

Session 5aSCa

Speech Communication and Signal Processing in Acoustics: Speech Analysis and Synthesis Techniques and Their Applications to Acoustics

Jose A. Diaz, Cochair
VLN 258, P.O. Box 025685, Miami, Florida 33102-5685

Emily A. Tobey, Cochair
Callier Center, University of Texas at Dallas, 1966 Inwood Road, Dallas, Texas 75235

Chair's Introduction—8:00

Invited Papers

8:05

5aSCa1. Speech analysis in automatic speaker recognition schemes in both forensic and commercial applications. Javier Ortega-Garcia and Joaquin Gonzalez-Rodriguez (Universidad Politecnica de Madrid, Spain and EUIT Telecomunicacion-UPM, Ctra Valencia km 7, E-28031 Madrid, Spain, jortega@diac.upm.es)

Speaker identity is a complex information included in the speech signal and codified in several levels of knowledge. Regarding the phonetic segmental level, this information resides in the specific acoustic resonances of the vocal tract that can be observed in the spectral envelope. Considering the nonstationarity of speech signals, short-time speech analysis techniques are employed through time-windowing procedures. Homomorfic cepstral analysis is then usually performed, as deconvolution between glottal information and vocal tract information is efficiently accomplished in this domain, guiding to a parametrization process of each frame through different techniques, as Mel-frequency-derived (MFCC) or linear prediction-based (LPCC) cepstral analysis. After the parametrization stage, probabilistic characterization of speakers identity is attained, usually through the use of left-to-right or ergodic hidden Markov models (HMM). Two different operating modes are clearly found in speaker recognition, namely (i) speaker identification, task in which a speaker has to be selected from a given set of speakers models, and (ii) speaker verification, in which a binary decision (accepted/rejected) has to be taken regarding the input unknown speech. Automatic speaker recognition leads to development of both commercial and forensic applications, but the cost of the decision of the system clearly separates the technical specificity of them.

8:35

5aSCa2. A robust fundamental frequency detection algorithm based on the spectrogram. Jose Diaz (Universidad de Carabobo, Valencia, Venezuela), Christine Sapienza, Howard Rothman, and Yaser Natour (Univ. of Florida, Gainesville, FL 32601)

The detection of fundamental frequency (Fo) in speech has often been shown to be a particularly difficult signal processing problem. This parameter is a necessary one for documenting vocal fold vibration and alterations to these vibratory patterns in the presence of pathology. There exist a variety of algorithms for extracting and analyzing Fo. It has been reported that these techniques do not work well for different types of talkers and decrease in performance as the noise level increases. The main objective of this research was to develop a robust algorithm for the extraction and analysis of Fo from normal and pathological voices. This algorithm is based on the spectrogram and makes use of artificial intelligence techniques to extract Fo. An algorithm was developed and tested with six normal and six abnormal samples, which contain a sustained vowel. These 12 samples were also analyzed by two commercial software packages, which make use of other techniques, and the results were compared to the results provided by the proposed algorithm. The results provided by the proposed algorithm agreed with those of the other two software packages, and the algorithm was able to detect Fo for a longer period of time.

Session 5aSCb

Speech Communication: Speech Recognition, Analysis, and Signal Processing (Poster Session)

Renetta G. Tull, Chair
Waisman Center, University of Wisconsin, 1500 Highland Avenue, Room 461, Madison, Wisconsin 53705

Contributed Papers

All posters will be on display from 9:15 a.m. to 10:45 a.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 9:15 a.m. to 10:00 a.m. and contributors of even-numbered papers will be at their posters from 10:00 a.m. to 10:45 a.m.

5aSCb1. Dynamic synapse neural network (DSNN): A new configuration based on wavelet filter bank and genetic algorithm. Hassan Heidari Namarvar, Jim-Shih Liaw, and Theodore W. Berger (Dept. of Biomed. Eng., Univ. of Southern California, Los Angeles, CA 90089-1451, heidarin@usc.edu)

A new DSNN architecture has been developed by the wavelet filter bank and the genetic algorithm (GA) training algorithm. The original DSNN [J.-S. Liaw and T. W. Berger, Hippocampus **6**, 591–600 (1996)] was based on the integrate-and-fire-based neurons using Hebbian learning. Implementing input neurons of DSNN as wavelet filters has been shown to increase the retention of input information and the system performance in comparison to the integrate-and-fire neural network with a Hebbian learning algorithm. This study has shown improvements in achieving convergence of the neural network for difficult discrimination conditions. The Hebbian and anti-Hebbian learning rules do not take into account all of the system parameters which may have a significant impact on system performance. Hence we used the GA as the training algorithm to provide an efficient way for searching parameter spaces. This novel network has been tested by raw speech waveforms for a speech recognition task. The system performance during training phase was highly improved in comparison with the prior version of DSNN. [Work supported by DARPA.]

5aSCb2. Automatic classification of the telephone-listening environment in a hearing aid. Peter Nordqvist (Dept. of Speech Music and Hearing, Royal Univ. of Sweden, Drottning Kristinas vg 31, Stockholm, Sweden) and Huanping Dai (GN ReSound Corp., Redwood City, CA 94063)

An algorithm is developed for automatic classification of the telephone-listening environment in a hearing instrument. The system would enable the hearing aid to automatically change its behavior when it is used for a telephone conversation (e.g., decrease the amplification in the hearing aid, or adapt the feedback suppression algorithm for reflections from the telephone handset). The telephone-listening environment and a reference listening environment are modeled with discrete hidden Markov models. The reference listening environment contains information about clean speech and a few different types of background noises. The probabilities for the different listening environments are calculated with forward algorithm for each frame of the input sound, and are compared with each other in order to detect the telephone-listening environment. Preliminary results demonstrate accurate classification of telephone speech by the hidden Markov system. [Work supported by GN ReSound.]

5aSCb3. A new technique for joint optimization of excitation and model parameters in parametric speech coders. Khosrow Lashkari (DoCoMo USA Labs., 181 Metro Dr., Ste. 300, San Jose, CA 95110, lashkari@dcl.docomo-usa.com) and Toshio Miki (DoCoMo USA Labs., 181 Metro Dr., Ste. 300, San Jose, CA 95110, toshio@dcl.docomo-usa.com)

In a speech coding system, synthesis error (the difference between the original speech at the encoder input and the reproduced speech at the decoder output) is a more relevant measure of signal distortion than linear prediction (LP) error. By minimizing the synthesis error instead of the linear prediction error, the analysis and synthesis stages become more compatible. While LP error is linear in filter parameters, synthesis error is a highly nonlinear function of these parameters, making it computationally intractable for real-time applications. This paper presents a computationally feasible solution for minimizing the synthesis error applicable for joint optimization of the excitation and model parameters in real time. Using a gradient search in the root domain, synthesis error is minimized by re-optimizing the filter parameters for a given excitation. Starting the gradient search from the LPC solution, the resultant synthesis error produced by the proposed technique is guaranteed to be lower than the synthesis error using the LPC filter. By adding an extra minimization step, this technique can be incorporated into any parametric speech coder including LPC, multipulse LPC and CELP speech coders. The new technique and results of application to real speech will be presented.

5aSCb4. Speech enhancement for an in-ear communication system using bone-conducted and acoustic signals. Kyriakos Papanagiotou, Boaz Rafaely (Inst. of Sound and Vib. Res., Univ. of Southampton SO17 1BJ, UK), and Christine H. Shadle (Univ. of Southampton SO17 1BJ, UK)

The main requirement for mobile radiotelephony nowadays is an ergonomic, robust, and functional system. One weak point of existing devices, such as hands-free kits for mobile telephones, is poor speech intelligibility due to factors such as environmental noise and poor signal reception. This work investigates the potential of speech enhancement using post-processing of signals from two types of transducers: a miniature microphone placed on the opening of the ear canal, and a miniature accelerometer mounted on four locations around the head. The microphone provides good speech quality with the drawback of being sensitive to external noise, while the accelerometer senses vibrations that are immune to ambient noise, but are a distorted, low-pass-filtered version of speech sounds. The use of optimal filtering, which enhances the signals acquired from the miniature transducers by matching them to a reference recording of the same material, was investigated. The methods investigated here could then be applied to improve speech quality and reduce noise in communication earsets. Initial results with speech, based on sus-

tained phonemes and phonetically balanced sentences, showed potential for improvement in the quality of the speech signals and also suggested that different filters might be used for different phonemic contexts. [Research funded by EPSRC Grant No. GR/M86026.]

5aSCb5. Improvements in methodology for the pitch-shifting technique. Jay J. Bauer and Charles R. Larson (Dept. of Commun. Sci. and Disord., Northwestern Univ., 2299 N. Campus Dr., Evanston, IL 60208, j-bauer2@northwestern.edu)

In recent years several studies have been published on voice control that utilized the pitch-shifting technique. The technique requires averaging of multiple vocalizations to repeated presentations of pitch-shifted feedback in order to eliminate random fluctuations in voice fundamental frequency. Time-locked reflexive responses to a stimulus parameter are observed quite clearly after averaging. The present study was undertaken to improve the pitch-shift methodology for data acquisition and analysis. Auditory pitch stimuli were presented at rates of 10, 5, and 2 per 5-sec vocalization, as opposed to previous studies that used one stimulus per vocalization. Results of testing these rates of stimulus presentation on a group of 22 subjects indicated there was no difference in reflex magnitude or latency between the present methods and rates up to 5/vocalization. Rates of 10/vocalization led to a reduction in reflex magnitude. Using the 5/vocalization rate, it was possible to reduce the number of vocalizations required of each subject per experimental condition with no loss of data accuracy. These improvements allowed for a greater amount of data collection in a shorter period of time, while possibly reducing the effects of subject fatigue and vocal instability. [Research supported by NIH Grant No. DC02764.]

5aSCb6. Evaluating consonant errors in automatic speech recognition programs. Renetta G. Tull, Erin E. Molin, Meaghan M. Lindstedt, and Jessica R. Dykstra (Dept. of Communicative Disord., Univ. of Wisconsin–Madison, 1975 Willow Dr., Madison, WI 53705, tull@waisman.wisc.edu)

Many voice-enabled computer applications (e.g., voice-controlled web commands, automatic call centers) use automatic speech recognition (ASR) software that allow users to command the computer by voice. Problems occur because ASR does not exhibit 100% accuracy for all speech patterns. In preparation for future evaluation of nonprototypical voices and speech technology, this project used three subjects with prototypical (no known speech pathology) to test three commercially available ASR software programs: Voice Xpress(R), Naturally Speaking(R), and Via Voice(R). Phonetic analyses revealed consonant errors in each of the three programs. Vowel errors are not addressed in this study. Naturally Speaking(R) 4 was the most accurate of the ASR programs tested. Errors were identified at the word level and analyzed by phoneme and manner of articulation. Similarities among the consonant errors were found in each program. Tests showed stop-to-stop (e.g., /t, d, b, p, g, k/), fricative-to-fricative (e.g., /s, z, f, v, sh, zh/), and fricative-to-stop errors. Nasal to nasal exchanges (e.g., /m/ /n/) were noted for each software package in each of the initial tests. Constant errors are expected to decrease as errors are identified and disseminated among software developers and User Centered Design is employed. [Work supported by NIDRR.]

5aSCb7. Laser Doppler vibrometer and subminiature electret microphone: New methods for objective evaluation of nasality. Masanobu Kumada (Dept. of Speech Physiol., Univ. of Tokyo, Tokyo, Japan, kumada@sol.dti.ne.jp), Toshiaki Kaneko (Sci. Univ. of Tokyo, Tokyo, Japan), Takahiko Ono, Masanao Ohashi, Koji Ishida (Ono Sokki, Yokohama, Japan), and Seiji Niimi (Univ. of Tokyo, Tokyo, Japan)

It is very important to evaluate nasality objectively in terms of both intersubject (severity of nasality) and intrasubject (validity of treatments, i.e., operation, speech therapy etc.). But there is no definite methodology for an objective evaluation of nasality. Our objectives were to evaluate the degree of nasality with a laser Doppler vibrometer (LDV) and a subminiature electret microphone (SEM) (using an accelerometer as a known standard), and to discuss the possibility of clinical use of these methodologies. LDV, used recently in the industrial field as well as in the medical/biological field, is a generator and detector of lasers to detect the velocity of the object's vibration by the Doppler effect. The subjects were five healthy Japanese males (Tokyo dialect speakers). Used were sustained phonations (about 2 s) of /m/ and five Japanese vowels (/i/, /e/, /a/, /o/, /u/), and the Japanese words, shinkansen and shimbunshi (six repetitions for each phoneme). By both LDV and SEM systems, a significant difference was found beween /m/ and each vowel, which means that both systems can detect nasal/non-nasal differences and seem valid to evaluate the severity of pathological nasality (e.g., a cleft palate). Further study is planned for not only sustained phonation and word but also for sentence level, using patients (pathology) as well as healthy subjects (physiology). Furthermore, LDV seems to be a powerful tool for speech physiology where vibration of the body wall can be measured during speech and singing, without any contact receiver.

5aSCb8. Noise-robust spectral estimation using nonlinear transform filters. Kiyoaki Aikawa and Kentaro Ishizuka (NTT Commun. Sci. Labs., 3-1 Morinosato-Wakamiya, Atsugi-shi, Kanagawa 243-0198, Japan and Dept. of Psych., Boston Univ., Boston, MA 02115)

This paper proposes a new spectral estimation method for robust speech recognition. The new method suppresses the noise superposed on a speech spectrum using a novel nonlinear-transform-filtering algorithm. Vocal energy is concentrated at the harmonic components in a voiced portion of speech. Therefore, the signal-to-noise ratio (SNR) is high at the harmonic frequencies. If these high-SNR components selectively contribute to obtaining a spectral envelope, then the overall SNR will be improved compared to conventional spectral estimation methods. The new algorithm is composed of two stages of nonlinear-transform-filtering. In the first stage, the spectral gain is nonlinearly converted using the power of a real number. This nonlinear conversion emphasizes the high-energy components of the spectrum. The converted spectrum is smoothed in the frequency domain by a linear filter. The impulse responses of the high-energy components spread over noisy portions with this filtering. The second stage uses the nonlinear conversion performed using the power of the inverse of the exponent used in the first stage. Linear filtering follows after the conversion to compensate the smoothing filter used in the first stage. Simulation experiments demonstrated that the proposed algorithm successfully suppressed the noises superposed at the valleys of harmonic-structured speech spectra.

5aSCb9. Chemical impairment measure from speech. Harb S. Hayre (Chemical Fitness Screening, P.O. Box 19756, Houston, TX 77224-9756, sant2@aol.com)

Speech involves over 67.5% of cerebral functions, and chemical intake is sensed by nerve cells connected to various cerebral subsystems, therefore speech is an optimum representation of cerebral signal with chemical impairment correlates of cognition, reaction time, coordination, and motor function degradation. Five clean words were selected to avoid nasal, throat or th or sh sounds. Impairment measure was then subject normalized. This measure does not identify chemicals ingested. First field test involved 20 DWI arrestees by a local police unit, their BAC levels were compared with impairment measure, and everyone with BAC over 1.0 were found impaired. Furthermore individuals with the same BAC levels had different impairment levels. It can be explained on the basis of varying tolerance levels, food intake, weight, and medicine use. Similarly in a study of 40 university student binge drinkers were found to be not impaired at 1.0 BAC, and many others were impaired at much lower BAC levels, thus demystifying the alleged safety of 0.6 BAC for drivers! Over 6000 subjects have been tested in north America and Europe, and nearly 30% of polydrug/chemical users passing DOT drug tests were found impaired or unfit to function safely.

5aSCb10. Practical considerations in evaluating directional hearing devices. Huanping Dai, Srdjan Petrovic, Nick Michael, Laurel Olson, and Chaslav Pavlovic (GN ReSound Corp., 220 Saginaw Dr., Seaport Ctr., Redwood City, CA)

In theory, a directional hearing device could apply large attenuation to a single noise source and achieve great enhancement of signal-to-noise ratio relative to an omni-directional device. However, as observed in many clinic studies, the actual benefit of a directional device measured in typical rooms rarely exceeds 6 dB, even in conditions with a single noise source located in the direction of maximum attenuation of the device. This dramatic reduction from the theoretical benefit is a manifestation of many factors, including mismatch in the microphone responses, misalignment of the device relative to the target source, head movement, head shadow, room reverberation, and so on. In the present study, we examine the relative contributions of these factors using computer simulations and electroacoustical measurements. The results provide a basis for setting realistic expectations of the benefit of directional devices for specific listening environments.

5aSCb11. Acoustic-phonetic approach to speech recognition based on event detection and linear discriminant analysis. Amit Juneja and Carol Espy-Wilson (Dept. of Elec. and Computer Eng., Boston Univ., 8 St. Mary St., 3rd Fl., Boston, MA 02215, ajuneja@bu.edu)

An approach to speech recognition based on phonetic-feature theory was developed to recognize the E-set. The speech signal is first segmented into broad classes based on parameters for the manner phonetic features *sonorant*, *syllabic*, *continuant*, and *strident*, in addition to silence. The speech signal is divided into the regions: vowel, sonorant consonant, strong fricative, weak fricative, and stop. The events specified by the manner features are then used to extract parameters relevant for the place phonetic features. In particular, we extract parameters for the place phonetic features *labial* and *alveolar* for stops; and *alveolar* and *palatal* for strong fricatives. Decisions at this stage of recognition are based on an optimal linear combination of parameters. The linear weights are obtained from the training data by applying linear discriminant analysis. The TI46 database is used for training and testing. The results of our event-based system are compared to a standard MFCC and HMM based recognizer. [Work supported by NSF Grant SBR-9729688.]

5aSCb12. Robust speech event detection using strictly temporal information. Om D. Deshmukh, Carol E. Wilson, and Ariel Salomon (Speech Commun. Lab, Dept. of Elec. & Computer Eng., 8 St. Mary St., Boston, MA 02215)

A major problem in the development of robust speech recognition systems is our understanding of how to deal with noise and/or reduced spectral information. Previous studies have shown that the temporal structure of speech is robust and a good source of information for recognizing manner. This work discusses the development of algorithms that specifically target temporal information in the speech signal that can be used to identify different manner classes. In particular, we extract events associated with sonorant regions and obstruent regions. We compared the performance of our algorithms on clean speech (sentences taken from the TIMIT database) and on 1 channel spectrally impoverished speech [as described by Q.-J. Fu, House Ear Institute, Dept. of Auditory Implants and Perception, Tigersoft]. The results show no degradation in performance for the detection of landmarks associated with stops, affricates, strident fricatives, and vowels. There is a small degradation of six landmarks associated with the ''weaker'' consonants (nonstrident fricatives and sonorant consonants).

FRIDAY MORNING, 8 JUNE 2001 PDR 17, 10:25 A.M. TO 12:00 NOON

Session 5aSP

Signal Processing in Acoustics: Signal Processing for Auditory and Speech Systems

Robert C. Bilger, Chair
Speech and Hearing Science, University of Illinois, 901 South 6th Street, Champaign, Illinois 61820

Chair's Introduction—10:25

Contributed Papers

10:30

5aSP1. Robustness of beamforming algorithms with head-related transfer functions. Michael E. Lockwood, Douglas L. Jones, Robert C. Bilger, Charissa R. Lansing (Beckman Inst., Univ. of Illinois at Urbana–Champaign, 405 N. Mathews Ave., Urbana, IL 61801), William D. O'Brien, Jr., Bruce C. Wheeler, and Albert S. Feng (Beckman Inst., Univ. of Illinois at Urbana–Champaign, Urbana, IL 61801)

This study examines the robustness of the frequency-banded minimum-variance beamformer (FBMVB) when processing signals from in-the-ear microphones or signals that include head-related transfer function (HRTF) effects. In previous work using simulated signals from a free-field array [Lockwood *et al.*, J. Acoust. Soc. Am. **106**, 2278 (1999)], the two-channel FBMVB proved to be highly effective in extracting the signal of a target sound from a known direction amidst multiple interfering sound sources from other unknown directions. When the signals were modified to include the effects of a measured HRTF [G. Gardner and K. D. Martin, J. Acoust. Soc. Am. **97**, 3907–3908 (1995)], the performance of the FBMVB improved as measured by the SNR gain and intelligibility-weighted SNR gain. The FBMVB also outperformed conventional algorithms in both the free-field and HRTF tests. Additionally, analytical results show that under certain conditions the FBMVB may extract a target source and perfectly cancel multiple interfering sound sources despite differences in the frequency-dependent intersensor gains for each source. Consequently, the FBMVB may extract signals from the on-axis direction without knowledge of the HRTF, as long as the HRTF in the on-axis direction is identical for both ears. [Work supported by NIH-NIDCD Grant R21DC04840.]

10:45

5aSP2. Human performance in a multisource environment with a frequency-banded minimum-variance beamforming algorithm. Jeffery B. Larsen, Michael E. Lockwood, Charissa R. Lansing, Robert C. Bilger (Beckman Inst., Univ. of Illinois at Urbana–Champaign, 405 N. Mathews Ave., Urbana, IL 61801), Bruce C. Wheeler, William D. O'Brien, Jr., Douglas L. Jones, and Albert S. Feng (Beckman Inst., Univ. of Illinois at Urbana–Champaign, Urbana, IL 61801)

The effectiveness of a newly developed signal-processing algorithm to extract speech in the presence of multiple interferers was evaluated through intelligibility ratings from normal- and hearing-impaired listeners. Sentences were presented with competing speech and multitalker babble (MTB) at four different SNRs to groups of individuals with sensorineural hearing loss and normal hearing. The sentences and competition had been processed through a real-time, frequency-banded minimum-variance beamforming (FBMVB) algorithm in multitalker environments (Lockwood *et al.*, 1999). Competing messages and MTB were simulated at $+22$ and $+45$ azimuths relative to the target source. Each target sentence with competing messages was submitted to the FBMVB algorithm to yield a set of 24 ''processed'' and 24 ''unprocessed'' sentences per SNR condition. Listeners rated the intelligibility (0%–100%) of the target sentences and repeated the last word of each sentence. Results from the ratings and the word recognition scores show improvement for every listener when the ensemble speech and competition were processed through the FBMVB algorithm. Results from these experiments indicate the advantage of the algorithm in selective extraction of speech in complex environments. Continued evaluation of the algorithm under real-room conditions is in progress. [Work supported by R21DC04840 ONR N00014-99-1-0612.]

11:00

5aSP3. A mathematical theory of comodulation. Barry D. Jacobson (Harvard–MIT Div. of Health, Sci. and Technol., Cambridge, MA and Johns Hopkins Univ., Baltimore, MD 21218), Gert Cauwenberghs (Johns Hopkins Univ., Baltimore, MD 21218), and Leonid M. Litvak (Massachusetts Eye and Ear Infirmary, Boston, MA)

Amplitude comodulation is the phenomenon that for many types of sound sources, all of the individual frequency components emitted will be amplitude modulated in a related manner, i.e., they will all start/stop and rise/fall together. It is known from various psychophysical studies that comodulation plays an important role in perceptual grouping of sound components. In this work, amplitude comodulation is examined for possible use as a basis for auditory source separation. A derivation is presented of the equations governing a comodulated system for the case of constant frequency sources. A theorem is proved giving necessary and sufficient conditions for uniquely decomposing a sound mixture into its constituent sources. A by-product of these conditions is an understanding of the importance of common onsets and offsets in source separation. An algorithm is presented for computing the unique solution in those cases where one exists. Finally, possible future extensions to the case of frequency varying sources are discussed. [Work supported by NIH Training Grant No. 5T32 DC00038.]

11:15

5aSP4. Comparison of binaural beamformers for speech extraction in complex auditory scenes. Yang Zheng, Michael E. Lockwood, Bruce C. Wheeler, Douglas L. Jones (Beckman Inst., Univ. of Illinois at Urbana–Champaign, 405 N. Mathews Ave., Urbana, IL 61801), Albert S. Feng, William D. O'Brien, Jr., Robert C. Bilger, and Charissa R. Lansing (Univ. of Illinois at Urbana-Champaign, Urbana, IL 61801)

The goal of this study was to assess and compare the capacities of several array-processing algorithms to extract on-axis signals in the presence of one or more off-axis interferers. Performance comparisons are made between the localization/extraction (L/E) algorithm [Liu *et al.*, J. Acoust. Soc. Am. **108**, 1888–1905 (2000)], the frequency-banded minimum-variance beamformer (FBMVB) [Lockwood *et al.*, J. Acoust. Soc. Am. **106**, 2278 (1999)], the Frost beamformer, the Griffiths–Jim generalized sidelobe canceller (GSC), and the Peissig–Kollmeier (P-K) algorithm [Kollmeier *et al.*, J. Rehab. Res. Dev. **30**, 82–94 (1993)]. The two metrics used to evaluate algorithm performance were the signal-to-noise ratio gain (SNRG) and the intelligibility-weighted signal-to-noise ratio gain (IWSNRG). The signals used simulated an omnidirectional microphone array in free field and were composed of sentence-length speech by male and female speakers against a background of one to four interfering sound sources. Overall, the IWSNRGs for the L/E and FBMVB algorithms were similar to those for the Frost, GSC, and P-K algorithms when there was a single interferer, but 3–4 dB higher when there were two or more interferers. This study shows that FBMVB and L/E algorithms hold promise for hearing aids to be used in noisy environments. [Work supported by NIH-NIDCD Grant R21DC04840.]

11:30

5aSP5. Beamformer improvement with increasing number of sensors. Matthew J. Berry, Michael E. Lockwood, Douglas L. Jones, Robert C. Bilger, Charissa R. Lansing, William D. O'Brien, Jr., Bruce C. Wheeler, and Albert S. Feng (Beckman Inst., Univ. of Illinois–Urbana-Champaign, 405 N. Mathews Ave., Urbana, IL 61801)

The two-sensor frequency-banded minimum-variance beamformer (FBMVB) [Lockwood *et al.*, J. Acoust. Soc. Am. **106**, 2278 (1999)], shown to extract a target speech signal from multiple interfering speech sources, is extended to three and four sensors. The extended FBMVB retains the real-time computational advantages of the two-sensor FBMVB. Results from a battery of simulated and real acoustic environment tests suggest that the extended FBMVB offers better extraction of the target speech than the two-sensor FBMVB. Measuring the quality of extraction as signal-to-noise ratio gain and intelligibility-weighted signal-to-noise ratio gain, the extended FBMVB demonstrates performance improvements of 2–3 dB over the already high performance of the two-sensor FBMVB for most tests. In general, results suggest that the extended FBMVB is more robust with respect to crowded interferer settings and room reverberation.

11:45

5aSP6. Eyes and ears: Attentive teleconferencing utilizing audio and video cues. Bill Kapralos, Michael Jenkin, John Tsotsos (Dept. of Computer Sci. Ctr. for Vision Res., York Univ., North York, ON M3j 1P3, Canada), and Evangelos Milios (Dalhousie Univ., Halifax, NS B3H 1W5, Canada)

The multiple speaker teleconferencing systems currently available typically focus on a single speaker and provide limited, if any, automatic speaker tracking technologies. However, in a multiple-speaker setting, speakers must be localized and tracked in both the video and audio domains. Although many fast and portable video trackers capable of locating and tracking humans exist, they employ conventional cameras thereby providing a narrow field of view. In addition, audio localization systems are expensive, nonportable, and computationally intensive. Furthermore, there have been very few attempts to combine both audio and visual systems. This work investigates the development of a simple, economical, and compact teleconferencing system utilizing both audio and video cues. An omni-directional video sensor is used to provide a view of the entire visual hemisphere thereby providing multiple dynamic views of all participants. Using a statistical color model and simple geometrical properties, the location of each participant's face is determined and provided to the audio system as a possible direction to a sound source. Beam forming with a small, compact microphone array allows the audio system to detect and focus on the speech of each participant. The results of experiments conducted in normal, reverberant environments indicate the effectiveness of the system.

Session 5aUW

Underwater Acoustics: Time Reversal and Propagation

David R. Dowling, Chair
Department of Mechanical Engineering and Applied Mechanics, University of Michigan, 2019 Autolab, 2121, Ann Arbor, Michigan 48109-2121

Contributed Papers

8:30

5aUW1. Shallow-water reverberation from a time reversed mirror: Data-model comparison. B. Edward McDonald (US Naval Res. Lab., Washington, DC 20375) and Charles Holland (Saclant Undersea Res. Ctr., 19138 La Spezia, Italy)

A July 1999 bottom reverberation experiment and accompanying simulation results will be discussed. The experiment was carried out as part of the Focused Acoustic Fields joint program between Saclantcen and Scripps Institution of Oceanography. A 3.5-kHz probe source 1 m off the bottom just north of Elba produced a signal which was recorded on a vertical source/ receive array (SRA) at ranges between 3 and 7.4 km. The SRA then time reversed and retransmitted the signal, resulting in a focal annulus near the bottom at the probe source range. Reverberation from the SRA signal was recorded at the SRA. A distinct maximum in reverberation at the time corresponding to the probe source range indicates selective bottom ensonification near the focus. Simulation studies carried out using RAM and the Saclantcen PROSIM reverberation models reproduced salient features of the experiment and led to suggestions for future experiments. Model sensitivity to array aperture and focal range will be discussed. Results imply that reverberation from a time reversed mirror may be useful in rapidly probing mildly varying bathymetry. [Work supported by Saclantcen.]

8:45

5aUW2. Time-reversed reverberation focusing in a waveguide. J. F. Lingevitch (Naval Res. Lab., Washington, DC 20375), H. C. Song, and W. A. Kuperman (Scripps Inst. of Oceanogr., Univ. of California–San Diego, La Jolla, CA 92093)

Time reversal mirrors have been demonstrated to focus acoustic energy in the ocean at ranges of tens of kilometers from a probe source [W. A. Kuperman *et al.*, J. Acoust. Soc. Am. **103**, 25–40 (1998)]. Time reversal has also been exploited to focus acoustic energy at discrete scatters of unknown location using the eigenvectors of the time reversal operator which can be formed by measuring the interelement transfer function of a source/receiver array [C. Prada *et al.*, J. Acoust. Soc. Am. **99**, 2067–2076 (1996)]. This paper extends these principles to the case of reverberation from a rough interface in a waveguide. The reverberation data collected on a source/receiver array are used to form a time reversal operator. The most energetic eigenvector of this operator focuses to a specific range along the rough interface. Improved signal-to-noise measurements of the time reversal operator can be obtained by probing the water column with orthogonal array beams. Since this method does not depend upon *a priori* environmental information, it is applicable to complex shallow water environments. Numerical simulations with a Pekeris waveguide are used to demonstrate this method. [Work supported by ONR.]

9:00

5aUW3. Echo to reverberation enhancement with time-reversal methods in the ocean. Seongil Kim, William A. Kuperman, William S. Hodgkiss (Scripps Inst. of Oceanogr., La Jolla, CA 92093-0238), Tuncay Akal (SACLANT Undersea Res. Ctr., 19138 La Spezia, Italy), Hee Chun Song, and Geoffrey F. Edelmann (Scripps Inst. of Oceanogr., La Jolla, CA 92093-0238)

Reverberation from a rough ocean boundary often degrades the performance of active systems in shallow water. In this study, two time-reversal methods are investigated to enhance the echo-to-reverberation ratio by focusing more acoustic energy in the water column rather than on the ocean boundary. The first method is with a probe source in the middle of the water column. A time-reversal process with the probe signal realizes a focus at the probe signal position and shadows the boundaries below and above the focus resulting in reduced reverberation and increased target echo. The second method is without a probe source. An eigendecomposition of the reverberation signal provides the transfer function information between the source and the scatterers where the reverberation is generated. The backprogation from a source array weighted by nondominant eigenvectors minimize the acoustic field incident on the corresponding boundary. Both of these methods place a notch in the returning reverberation signal. The results from numerical simulations as well as an ocean experiment will be discussed. [Work supported by ONR.]

9:15

5aUW4. Shallow water performance of a time-reversing array with a moving source. Karim G. Sabra and David R. Dowling (Dept. of Mech. Eng., Univ. of Michigan, Ann Arbor, MI 48109-2121, drd@engin.umich.edu)

Future active sonar systems may be developed from the automatic spatial and temporal focusing properties of time-reversing arrays (TRAs) in unknown environments. In the case of source and/or array motion, both the forward and backward propagation necessary for TRA operations may be influenced by the Doppler effect. Predictions for TRA focusing performance for a single source emitting broadband or narrowband Doppler-shifted signals in a shallow-water waveguide can be obtained from a Fourier superposition of stationary but spatially distributed time-harmonic sources. This approach is illustrated for a Pekeris waveguide and extended to range-dependent environments with the wide-angle parabolic equation code RAM (by Collins). Results for the characteristics of the retrofocused field will be presented for different combinations of source-array motion and geometry for acoustic center frequencies of several hundred Hertz and nominal ranges of up to 20 km in a range-independent sound channel. Results from fully range-dependent sound channels may also be presented. [Work sponsored by ONR.]

9:30

5aUW5. The meaning of negative group velocity and amplitude modulated signals. M. F. Werby (NRL Code 7180, Stennis Space Center, MS 39529 and Dept. of Phys., The Catholic Univ. of America, Washington, DC 20064) and H. Uberall (The Catholic Univ. of America, Washington, DC 20064)

Negative group velocities have been reported in calculations for waves generated on elastic objects. It is usual to interpret a group velocity with the travel time of a wave front or the time that the energy of a wave generated travels along some trajectory. One may look at the group velocity as the rate at which an envelope of a wave train travels. More generally, one may exploit the principle of stationary phase along with the Fourier transform in frequency and derive the concept of group velocity from the first order asymptotic term. This derivation requires that the modulus of the integral is changing slowly relative to the phase term. In addition it is required that the group velocity is not near a stationary point in frequency. In that case first order asymptotics is no longer valid and one must employ higher asymptotics. Here, the standard concept of group velocity is no longer meaningful. Further, the signal becomes amplitude modulated rather than frequency modulated. We examine this for some examples, assuming losslessness. In the event of loss the concept of group velocity has already been reported to be of dubious value.

9:45

5aUW6. Optimal array design and sensitivity for mode filtering. Tsung-Jieh Shiao (Dept. of ECE & SMAST, Univ. of Massachusetts–Dartmouth, 285 Old Westport Rd., North Dartmouth, MA 02747, g_tshiao@umassd.edu) and John R. Buck (Univ. of Massachusetts–Dartmouth, North Dartmouth, MA 02747)

The Cramer–Rao lower bound (CRLB) gives the smallest possible variance of any unbiased estimator. Buck, Preisig, and Wage [J. Acoust. Soc. Am. **103** (1998)] derived this bound for the variance of unbiased mode filters and observed that the array geometry is the only factor controlling performance under scientists' control. The difficulty lies in designing an appropriate array for variable ocean environments. This research includes two parts. First, the minimax optimization criteria and the gradient descent method are applied to find optimal arrays over a range of possible deployment environments. In several simple prototype problems, it is possible to design nonuniform arrays whose worst case CRLB is significantly lower than a uniformly spaced array. This contradicts conventional wisdom that uniformly spaced arrays provide the most flexibility for multiple or extended deployments. Moreover, the gradient descent method is more computationally efficient than exhaustive search methods of optimizing arrays. Second, the sensitivity of the optimal array's performance to environmental mismatch is studied and the perturbation of this performance is analytically bounded. This bound works well in the preliminary results and predicts the deterioration of mode filter performance due to mismatch. [Work supported by ONR Young Investigator Program.]

10:00–10:15 Break

10:15

5aUW7. Generalized optical theorem for two-dimensional scattering: Applications to backscattering and arbitrary scattering angles. Philip L. Marston (Dept. of Phys., Washington State Univ., Pullman, WA 99164-2814)

A generalized optical theorem has been demonstrated for scattering by three-dimensional targets having no energy dissipation as well as inversion symmetry [P. L. Marston, J. Acoust. Soc. Am. (accepted for publication)]. A component of the complex scattering amplitude in an arbitrary direction is expressed in terms of a two-dimensional angular integration involving scattering amplitudes. The integral may be reduced to a one-dimensional integration only in special cases such as backscattering along a symmetry axis. In new research an analogous theorem for two-dimensional scattering is found. The theorem for two-dimensional scattering allows (for an arbitrary direction) a projection of the complex scattering amplitude to be expressed as a one-dimensional integral. The integral is a circular-autocorrelation of complex scattering amplitudes from -180 deg to $+180$ deg. This was numerically evaluated for simple targets to confirm the theorem. Potential applications of generalized optical theorems include the verification of the consistency of scattering approximations and investigations of multiple-scattering by inversion-symmetric sets of scatterers. [Work supported by ONR.]

10:30

5aUW8. Characterization of underwater explosions by cepstral analysis, modeling and inversion. Douglas R. Baumgardt and Angelina M. Freeman (ENSCO, Inc., 5400 Port Royal Rd., Springfield, VA 22151)

This paper describes a cepstral simulation and inversion algorithm for characterizing regional seismic recordings of underwater explosions to determine the yield and depth of the explosions. The spectra of seismic recordings of underwater blasts have strong time-independent spectral scalloping, produced by the correlated bubble pulses and echoes of acoustic reflections between the water surface and the sea bottom. Signed cepstra, computed from the Fourier transform of the log-amplitude spectra of regional seismic phases (Ph, Pg, Sn, and Lg), contain positive peaks produced by the bubble pulse correlation, and negative peaks from the water-surface reflection. Theoretical cepstra of underwater explosions with assumed depths and yields are computed from simulated time sequences of bubble pressure pulses and water column reflections. The prototyped inversion algorithm determines explosion depth and yield by matching the synthetic model cepstra to data cepstra, using exhaustive and optimized search methods. This algorithm has been applied to presumed underwater explosions in the North Sea, tested successfully on three calibration explosions in the Dead Sea, and utilized for the analysis of the Kursk submarine event, a presumed underwater explosion. [Work supported by DTRA.]

10:45

5aUW9. An approach to array element localization using a far-field source. Craig S. MacInnes (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029) and David P. Knobles (Appl. Res. Labs., Univ. of Texas, Austin, TX 78713)

An approach is introduced for the towed array element localization problem. The method uses a dominant acoustic far-field source to obtain the true array shape. Constraints are introduced that significantly reduce the dimensionality of the problem. The array shape is expanded in terms of Chebyshev type II polynomials. The array elements are constrained to be in sequential order, and the arc length between elements is conserved as the array transforms from a straight to a curved configuration. A simulated annealing (SA) algorithm is used to find the expansion coefficients that minimize the inverse of the magnitude of the plane-wave beamformer response evaluated at the bearing to the source. The minimum of the cost function is obtained when the array elements are in their correct positions. Simulated studies are made for several array shapes. In all cases examined the expansion method with SA is able to find an accurate solution, independent of the initial array configuration. The analysis suggests that the use of broadband as opposed to single-frequency cost functions has the effect of reducing the number of local minima in the cost function that have values near the global minimum. [Research supported by the U.S. Navy.]

11:00

5aUW10. Effect of seawater salt on compressional wave in dry beach sand. Jacques R. Chamuel (Sonoquest Adv. Ultrason. Res., P.O. Box 81153, Wellesley Hills, MA 02481-0001)

At the previous ASA meeting, quantitative results were presented revealing the effect of water content on compressional wave velocity in beach sand. New ultrasonic compressional wave results are presented demonstrating the effect of salt crystals bonding beach sand grains as seawater is evaporated. Wave velocity and attenuation results are compared with beach sand rinsed with distilled water. Although the salinity of

seawater is only about 35 parts per thousand, the salt concentration increases rapidly as the seawater is evaporated. The sand grains become covered with a glaze and solid salt bonds are formed connecting the grains. The acoustic properties of salt crystals (compressional wave velocity 4.53 km/s, and density 2.165×10^3 kg/m^3) are close to those of a sand grain. Results from controlled experiments using spherical particles and plates bonded with evaporated salt water are also discussed. The author is not aware of any publication describing the adhesion of sea salt to beach sand. The new findings contribute to understanding the propagation of high-frequency compressional waves in granular materials leading to reliable seismoacoustic detection of buried objects in beach sand. [Work supported by ONR.]

11:15

5aUW11. Solution of the nonhomogeneous Sturm–Liouville problem and an inverse solution to an inclusion imbedded in a waveguide. M. F. Werby (NRL Code 7180, Stennis Space Center, MS 39529 and Dept. of Phys., The Catholic Univ. of America, Washington, DC 20064) and H. Uberall (The Catholic Univ. of America, Washington, DC 20064)

The homogeneous Sturm–Liouville problem is the usual starting point of a normal mode theory provided one adds absorption as a perturbation after the nonlossy solution in a waveguide. In this development we assume that prior to an inclusion the solution is known for the waveguide problem for any given source. When an inhomogeneity is introduced into the waveguide we assume that it may be treated as a source term. Using completeness, we expand the source term as a sum of eigenfunctions with unknown coefficients. Under reasonable conditions measurements of the field of the waveguide before and after the presence of the inhomogeneity allows one to determine the expansion coefficient of the source term and thus many relevant properties of the inclusion itself. Thus, determination of the source term allows one to determine features of an inclusion.

11:30

5aUW12. Ambient noise and source tracking in very shallow water. Stewart Glegg and Antony LaVigne (Ctr. for Acoust. and Vib., Dept. of Oceanogr. Eng., Florida Atlantic Univ., Boca Raton, FL 33431)

This paper will describe recent experiments on ambient noise levels and source tracking in shallow water. The measurements were made using a sparse acoustic array in approximately 20 m deep water at a location just south of the Port Everglades inlet. The ambient noise was dominated by local boat traffic events and was found to be very intermittent especially during daylight hours. Between events the noise was dominated by biological activity inshore of the array. It will be shown that the most important issue that adversely affects the acoustic tracking of AUVs is boat traffic noise. It was found that tracking accuracy at larger ranges is good, but the results are more inconsistent due to lower received signal levels. The algorithm used in this analysis was not necessarily optimized for noise rejection, but the results clearly indicate that signal fading and signal to noise are the main problems that have to be addressed to improve tracking performance. [Work supported by ONR.]

FRIDAY AFTERNOON, 8 JUNE 2001 ADAMS ROOM, 1:30 TO 2:50 P.M.

Session 5pAA

Architectural Acoustics, Noise, Physical Acoustics and Engineering Acoustics: Sound of Baseball

Joseph Pope, Chair
Pope Engineering Company, P.O. Box 590236, Newton, Massachusetts 02459-0002

Invited Papers

1:30

5pAA1. The crack-of-the-bat: The acoustics of the bat hitting the ball. Robert K. Adair (Dept. of Phys., Yale Univ., P.O. Box 208121, New Haven, CT 06520-8121, adair@hepmail.physics.yale.edu)

When a baseball is hit straight at an outfielder he cannot quickly judge the angle of ascent and the distance the ball will travel. If his first step is wrong (in for a long fly ball or back for a short fly ball) the turnaround time reduces his range and he will miss catchable balls. To help his judgment the professional ball player listens to the sound of the wooden bat hitting the ball. If he hears a ''crack'' he runs out, if he hears a ''clunk'' he runs in. The crack occurs if the ball is hit near the sweet spot, little collision energy is lost to bat vibration but is returned to the ball which then sails a long distance. The collision time is then near 0.5 ms generating a singular acoustic disturbance of about that time duration—thus a ''crack.'' If the ball is hit off of the sweet spot, much of the collision energy will be expended in bit vibrations near 170, 560, and 1100 Hz. With that energy diverted to the collision energy, the ball will not go very far and the vibrations generate persistent sound waves that add to a diminished singularity to sound a ''clunk.''

1:50

5pAA2. The sounds of baseball: The bat–ball collision and the crack of the bat. Robert D. Collier (Thayer School of Eng., Dartmouth College, Hanover, NH 03755)

The crack of the bat is one of the most exciting aspects of the game of baseball. This sound provides significant information to both player and spectator. The impact force between a fast ball pitch and a rotating bat provides the energy for the exit velocity of the ball. The relative efficiency of the energy transfer to the batted ball is related to the location of the impact area on the bat, e.g., the sweet spot, and the duration of the period in which the ball is in contact with the bat. Energy losses in the bat involve the vibratory modal response of the bat and its associated sound radiation. This paper discusses the modal response functions of both solid wood and tubular metal in relation to the location and nature of the impact forces generated by collisions with baseballs. Differences in amplitude, spectral characteristics and decay rates of the resulting radiated sound fields are analyzed for a range of bat–ball collision scenarios. Experimental results from both controlled laboratory and field hitting tests illustrate the relationships between measured bat radiated sounds and the subjective evaluation by players of batted ball trajectories.

2:10

5pAA3. Acoustical design of baseball stadiums. David E. Marsh (Pelton Marsh Kinsella, 1420 W. Mockingbird Ln., Ste. 400, Dallas, TX 75247, marshde@c-b.com)

Baseball team owners and stadium managers alike wish to have good acoustics in their venues. To achieve this, many issues must be addressed early in the design of a new stadium. For example, if community noise is a concern for a new open-air stadium, addressing this will require construction budgeting for certain architectural and sound system design features. Minimizing echoes is an acoustical design challenge for both open-air and enclosed stadiums. This requires a general awareness of likely echo paths, knowing what acoustical treatment options are generally acceptable for this building type, and use of predictive modeling for the sound system design. Enclosed stadiums pose additional design challenges such as the quieting of very large air-handing units, minimizing reverberation, and making the sound system intelligible. These challenges extend beyond the seating bowl to the concourses and a host of ancillary areas. This paper reviews the authors' experience in the overall acoustical design of modern baseball stadiums.

2:30

5pAA4. Acoustics sound systems for baseball. Jim Brown (Audio Systems Group, Inc., 4875 N. Ravenswood, Chicago, IL 60640, jim@audiosystemsgroup.com)

Acoustic considerations impact the design of sound systems for baseball stadiums in several important ways. Intrusion by the sound system into adjacent residential communities must often be minimized. Excess attenuation greatly increases the required high frequency component of acoustic power for sound travelling long distances, while wind and temperature gradients modify the effective coverage patterns of loudspeaker systems, and exacerbate community noise problems. These atmospheric effects can vary radically from moment to moment. These considerations tend to favor use of distributed systems, at the same time architectural considerations make them more difficult to implement. And especially with distributed systems, the timing of sound arrivals at each listener's ear is often more important than the overall amplitude of the sound field. In modern stadiums, other auxiliary systems are required. Crowd mics fed to distributed loudspeakers must bring the crack of the bat and the ''smell of the crowd'' into private suites which are often enclosed. That crowd noise, along with the public address announcer, must also be fed to broadcast media for their use. Timing is also an issue with these systems. This paper addresses both design challenges and practical solutions.

FRIDAY AFTERNOON, 8 JUNE 2001 PDR 18, 1:15 TO 2:30 P.M.

Session 5pBB

Biomedical Ultrasound/Bioresponse to Vibration and Physical Acoustics: Brad Sturtevant Memorial Session in Lithotripsy II

E. Carr Everbach, Chair

Department of Engineering, Swarthmore College, 500 College Avenue, Swarthmore, Pennsylvania 19081-1397

Contributed Papers

1:15

5pBB1. Effect of a rigid boundary on bubble dynamics in ESWL. Claus-Dieter Ohl (Dept. of Appl. Phys., Univ. of Twente, The Netherlands), Reinhard Geisler, Werner Lauterborn (Third Phys. Inst., Univ. of Goettingen, Germany), and Andrea Prosperetti (Johns Hopkins Univ., Baltimore, MD)

Recently it has been observed that two countertraveling lithotripter pulses form a cavitation bubble cloud with band-like structure of bubble activity [D. Sokolov *et al.*, J. Acoust. Soc. Am. **107**, 2838 (2000)]. A similar structure becomes apparent after focusing a single lithotripter pulse onto a rigid boundary. Here, the growing cavitation bubbles interact with the reflected pulse and the band-like structure appears near the boundary. Experiments will be presented to compare the nucleation, explosive growth and shock wave interaction of cavitation bubbles near the reflector to bubbles in a free liquid. The experiments have been conducted by means of high speed photography (up to 10 million frames per second) employing microscopic imaging. The observed radial and translational motion will be compared with a force balance model, taking into account the pressure radiation force.

1:30

5pBB2. Bubble dynamics in a lithotripter shock wave. Brian Storey (Olin College, 1735 Great Plain Ave., Needham, MA 02492, brian.storey@olin.edu), Andrew Szeri (Univ. of California, Berkeley, CA 94720), and Tom Matula (Univ. of Washington, Seattle, WA 98105)

Typically, a pulse in shock wave lithotripsy has an intense shock front followed by a long negative pressure tail. A micron sized bubble subjected to this pulse is compressed rapidly by the shock, before growing to millimeter size in response to the negative pressure tail. As this tail passes, the bubble will then undergo an essentially free collapse and subsequent afterbounces. Calculations of the bubble dynamics using a model that takes phase change, heat transfer, mass transfer, and chemical reactions into account provides excellent agreement with experimental measurements. The role of phase change and mass transfer is dominant and the bubble is composed primarily of vapor throughout the free collapse and afterbounces. When the vapor bubble collapses there are significant chemical reactions and predictions on the amount of radical species produced and peak temperature achieved will be presented.

1:45

5pBB3. Lithotripter shock waves are capable of damage at subcellular dimensions in the absence of cavitation. James C. Williams, Jr., Drew L. Rietjens, Chad A. Zarse, and James A. McAteer (Dept. of Anatomy and Cell Biol., Indiana Univ. School of Medicine, 635 Barnhill Dr., Indianapolis, IN 46202-5120, williams@anatomy.iupui.edu)

Lithotripter shock waves (SW's) cause tissue damage, but the size dimension of the SW-tissue interaction is not known. Unilamellar membrane vesicles were used as a model for subcellular organelles, and SW-induced damage was quantitated with and without cavitation. Membrane vesicles ($\approx$100 nm diameter) were made by extrusion from egg phosphatidylcholine, and exposed to SW's (1 Hz and 20 kV) in a research lithotripter that yields output similar to an unmodified Dornier HM3. SW-dependent vesicle lysis was linearly dependent on SW number ($p < 0.001$), with a mean lysis of $0.12 \pm 0.02\%$ at 400 SW's ($n = 49$ vials). This lysis value is >100 times less than that seen in red blood cells exposed to a similar SW dose. Vesicle lysis was independent of kV in the range tested (17–23 kV, $p > 0.9$, $n = 14$ vials). When vesicles were exposed to SW's in the presence of high overpressure (>130 atm) to eliminate cavitation effects, SW's still induced lysis not different from that at atmospheric pressure, $0.10 \pm 0.02\%$ ($n = 10$). This implies that lithotripter shock waves may be capable of causing intracellular damage to tissue even in regions where cavitation does not occur. [Work supported by NIH Grants P01 DK43881 and R01 DK55674.]

2:00

5pBB4. Combined shock wave and immunotherapy of mouse melanoma tumors. Douglas Miller, Jianming Song, Lang Li, Jeremy Taylor (Univ. of Michigan Medical Ctr., Ann Arbor, MI 48109), Shiping Bao (Washington State Unversity, Richland, WA 99352), and Darrell Tata (Southern Illinois Univ., Carbondale, IL 62901)

The effect of lithotripter shock waves (SWs) and interleukin-12 (IL12) treatment on the growth of B16 melanoma was investigated. Tumor cells were implanted and grown for 10 days on the hind legs of C57/Bl mice. Prior to treatment, mice were anesthetized, the tumor region was shaved, and air and phosphate buffered saline (PBS) were injected into the tumor at 10% of tumor volume each. Shock-wave treatment consisted of 500 shock waves (7.4-MPa peak negative pressure) from a spark-gap lithotripter. IL12 treatment consisted of 500 ng in the initial PBS injection, and additional injections 2 days and 4 days later. Tumor volume was measured every other day and tumor growth was statistically modeled. After 10 days, modeled volume increases, relative to the initial volumes in groups of six mice, were a factor of 17.8 for sham treatment, 7.8 for SWs, 3.6 for IL12, and 1.0 for combined SWs and IL12 ($P < 0.0001$ relative to shams). The combined treatment significantly prolonged survival (0.09 relative risk of death versus sham) and one mouse had complete tumor remission. These results are encouraging for combined SW and immunotherapy treatment of tumors. [Work supported by NIH CA83904.]

2:15

5pBB5. Renal vasoconstriction following shock wave lithotripsy (swl) is mediated by renal nerves. Bret Connors, Andrew Evan, Lynn Willis, Jay Simon (School of Medicine, Indiana Univ., Indianapolis, IN), Arieh Shalav, Ryan Paterson, and James Lingeman (Methodist Hospital, Indianapolis, IN)

High-energy shock waves (SWs) to one kidney (2000 SWs at 24 kV) reduces blood flow in both kidneys (65% in shocked kidneys, 33% in contralateral unshocked kidneys). We determined if unilateral renal denervation of the unshocked kidney alters renal blood vessel constriction after SWL. Six-week-old pigs underwent unilateral kidney denervation. Visible renal nerves along the renal artery of one kidney were isolated, cut, and painted with 10% phenol. Two weeks later, the pigs were anesthetized and kidney function was followed with inulin and PAH clearance. 2000 SWs (at 24 kV, Dornier HM3) were applied to the lower pole of the innervated kidney. Bilateral glomerular filtration rate (inulin clearance) and renal plasma flow (PAH clearance and extraction) were measured 1 h before and 1 h after SWL. Both kidneys were then removed to determine norepinephrine (NE) content in the tissue. Denervation reduced NE levels to 8.7% of innervated kidneys. While blood flow fell as expected in the innervated shocked kidney, there was no change in blood flow in the denervated kidney. Results suggest that renal nerves mediate part of the vasoconstrictor response of unshocked kidneys to SWL. [Work supported by NIH P01 DK43881.]

FRIDAY AFTERNOON, 8 JUNE 2001 CRYSTAL ROOM, 1:00 TO 2:30 P.M.

Session 5pSA

Structural Acoustics and Vibration in Physical Acoustics: David G. Crighton Memorial Session III

Paul E. Barbone, Chair

Department of Aerospace and Mechanical Engineering, Boston University, 110 Cummington Street, Boston, Massachusetts 02215

Contributed Papers

1:00

5pSA1. Effects of relative motion on linear and nonlinear waves in a bubbly medium. Jerome J. Cartmell, Ali Nadim (Dept. of Aerosp. and Mech. Eng., Boston Univ., Boston, MA 02215), and Paul E. Barbone (Boston Univ., Boston, MA 02215)

We describe a continuum formulation to model flow of a bubbly liquid. In our treatment, the bubbles are allowed to move relative to the host liquid. The continuum theory consists of conservation equations for bubble (number) density, mixture density, and momentum. These are closed via a constitutive equation which relates the instantaneous mixture pressure to the mass and number densities. The resulting continuum description of shock waves in a bubbly medium exhibits two distinct time scales. In addition to the usual time scale associated with bubble oscillation, there is a relaxation time scale associated with the velocity difference between the bubbles and their host liquid. Examples are presented.

1:15

5pSA2. Instability of a vortex sheet leaving a right-angled corner. Anthony M. J. Davis (Math Dept., Univ. of Alabama, Tuscaloosa, AL 35487) and Edward J. Kerschen (Univ. of Arizona, Tucson, AZ 85721)

David Crighton had a long-standing interest in how the results of Orszag and Crow [Stud. Appl. Math. **49**, 167–181 (1970)], for the instability of a vortex sheet leaving a zero-thickness trailing edge which separates a uniform stream from a fliud at rest, would be modified by creating a solid right-angled corner that restricts the stationary unperturbed flow to

a quadrant. The Wiener–Hopf technique cannot be applied to this problem which, having wedge-shaped regions, is amenable to Mellin transforms. The matching conditions at the vortex sheet yield a pair of functional difference equations whose solution in terms of the Barnes double gamma function is achieved after doubling their order. Then, local series expansions valid near and far from the corner enable the appropriate pole structure to be identified and the arbitrary periodic functions to be determined. The results are applied to instability wave excitation by an acoustic wave incident on the trailing edge region.

1:30

5pSA3. Sound interaction with moving vortices. Samuel A. Rybak, Andrew G. Semenov, and Vadim N. Alekseev (N. N. Andreev Acoust. Inst., Russian Acad. of Sci., 4 Shvernik St., Moscow, 117036 Russia)

This report is devoted to the nonlinear acoustic pressure of sound waves propagating in the vicinity of a rather general class of localized flows induced by the motion of spherically symmetric structures of the type of vortices in the liquid, say, biological, atmosphere, or ocean inhomogeneous environment. As a result expressions for cyclic and mean forces exerted on a vortex are derived. Conventional general expression for transport cross section derived in nonlinear theory of interaction of sound with rigid particles is proven to be incorrect for the vortex-sound interaction case. The absence of mean force is demonstrated for specific case of Hills vortex both in the traveling and standing sound wave fields. It is underlined, however, that the force could be observable if the distribution of mean sound energy in the incident sound field in the frames of vortex dimensions could be characterized by a nonzero spatial gradient. Instead of the previously used expression for mean radiation force exerted on a moving inhomogeneity valid for a solid particle based on its transport cross section, the correct generalized expression for force exerted on vortex structure is derived.

1:45

5pSA4. *Q*-factor limit of nonlinear systems. Claes M. Hedberg and Oleg V. Rudenko (Blekinge Inst. of Tech., 371 79 Karlskrona, Sweden, claes.hedberg@ima.hk-r.se)

The *Q* factor of a resonator is a constant, which gives an estimation of its ''quality.'' A system with a high *Q* can accumulate high energy and provide markedly nonlinear phenomena even in the case of weak power sources. At small amplitudes *Q* is limited by linear absorption caused by both the medium inside the resonator and its boundaries. As the amplitude increases, nonlinear absorption and frequency response distortion come into play, and *Q* is determined by both system design and the acoustic field strength. In general, nonlinearity leads a to significant decrease in *Q*. One can enhance *Q* in different ways using impedance boundaries whose frequency-dependent response destroys shock fronts, a medium containing selective absorbers suppressing shock formation, resonators of special shape, or special types of excitation to counter the distortion. Data on nonlinear *Q*-factors and response taking into account dissipative and frequency-dependent properties of systems are presented.

2:00

5pSA5. Waves in fluid and solid media. Włodzimierz Domański (Inst. of Fundamental Technol. Res., Polish Acad. of Sci., Swietokrzyska 21, 00-049 Warsaw, Poland)

''Waves in Fluid and Solid Media'' is the title of the last unpublished lecture notes of David G. Crighton. The lectures were given at the University of Cambridge and have the date Lent 1998. In these notes Crighton introduces basic notions connected with wave motion, analyzing in particular propagation of sound waves in a compressible fluid as well as dilatational and shear waves in an elastic solid. A part of these lectures is devoted to nonlinear acoustics including the discussion of wave steepening, formation of shocks, Rankine–Hugoniot conditions, Hugoniot adiabatic etc. In my talk the above mentioned ideas will be extended and recent developments in the theory of propagation and interaction of weakly nonlinear waves both in fluid and solid media will be presented [W. Domański, Contemp. Math. **255**, 45–61 (2000)].

2:15

5pSA6. Old-age stage evolution of intense cylindrical and spherical acoustic noise. Sergey N. Gurbatov, Valery V. Cherepennikov (Dept. of Radiophysics, Univ. of Nizhny Novgorod, 23 Gagarin Ave., 603600, Russia, gurb@rf.unn.runnet.ru), and Bengt O. Enflo (KTH, S-100 44 Stockholm, Sweden)

The investigation of the old-age behavior of an intense periodic spreading wave was one of the significant contributions of Crighton in nonlinear acoustics. By him and his co-workers a classification of different regimes of evolution was done and it was shown at which condition the saturation of the amplitude takes place. Here the evolution of intense spreading acoustic noise is considered on the base of generalized Burgers' equations. It is shown that at the old-age (linear) stage the energy spectrum of the noise has a universal structure: it is proportional to the second power of the frequency at small frequencies and decreases exponentially at high frequencies. The proportionality coefficient does not depend on the distance and is determined by the nonlinear effects on the initial stage of the propagation. It is shown that even for a small acoustical Reynolds number the energy of the noise at the old-age stage is proportional to the third power of the initial energy for cylindrical waves and to the power 5/2 with some logarithmic correction for spherical waves. The analytical estimations are supported by numerical simulation. [Work supported by RFBR and INTAS grants.]

Session 5pSC

Speech Communication: Speech Perception and Production in Special Populations: Cochlear Implants *et al.* (Poster Session)

Diane Kewley-Port, Chair
Department of Speech and Hearing Science, Indiana University, 200 South Jordan Avenue, Bloomington, Indiana 47405

Contributed Papers

All posters will be on display from 1:30 p.m. to 5:00 p.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 1:30 p.m. to 3:15 p.m. and contributors of even-numbered papers will be at their posters from 3:15 p.m. to 5:00 p.m.

5pSC1. Speech perception among good and poor readers. Erin Eckhouse and Ratree Wayland (Prog. in Linguist., Univ. of Florida, P.O. Box 115454, Gainesville, FL 32611-5454, erinlou789@aol.com)

A relationship between phonological awareness and reading ability among children has been suggested by some researchers. Several studies also suggested that phonological deficits among children are related to impairments in speech perception. The goal of this study is to examine the ability to discriminate and categorize various speech sounds by school-aged good and poor readers. Subjects were 30 third and fourth graders recruited from a local elementary school in Gainesville, Florida. Based on the school's, as well as, the State of Florida's standardized test, half of these subjects will be classified as good readers and the other half as poor readers. All subjects are native speakers of English. They were asked to perform five speech perception tasks. The first task ''tug-dug,'' examined subjects' categorization of a VOT continuum ranging from 10–80 ms long. The second task, ''spy-sky'' was designed to test subjects' ability to detect onset frequency of formant transitions. The last three tasks, ''shop-chop,'' examined the combined effect of a gap and fricative noise duration in the categorization of fricative–affricate continuum. A relationship between subjects' speech perception and their reading ability as measured by several standardized tests will be explored and discussed.

5pSC2. Discrimination and identification of vowels in hearing-impaired and masked normal-hearing listeners. Carrie Davis, Diane Kewley-Port, and Maureen Coughlin (Dept. of Speech and Hearing Sci., Indiana Univ., Bloomington, IN 47405, carodavi@indiana.edu)

This study examined the effects of mild-to-moderate sloping sensorineural hearing loss on vowel discrimination and identification. Five young, hearing-impaired (YHI) adults listened to vowel stimuli in three conditions: (a) at a conversational level with flat frequency response; and in two gain conditions, (b) at a high-sound level with flat frequency response; and (c) with frequency shaped gain according to the subject's hearing loss. Listeners discriminated changes in the synthetic vowels /ɪ ɛ e æ ʌ/ when $F1$ or $F2$ varied, and later identified the five vowels in a closed-set task. The results indicated that neither gain condition restored performance to that of young, normal-hearing listeners and that the potential upward spread of masking in condition (b) did not affect performance. Young, normal-hearing (YNH) listeners were age matched to YHI subjects. Hearing thresholds for the YNH listener of each YHI–YNH pair were precisely matched to those of the YHI listener by masking with spectrally shaped noise. YNH listeners with ''simulated-hearing loss'' were tested on the same discrimination and identification tasks. This design yields performance comparisons between YHI–YNH pairs when conditions of audibility are similar. The results will explain whether differences in performance between YHI–YNH pairs may be due to audibility or pathology. [Supported by NIHDCD–02229.]

5pSC3. Learning to recognize spectrally reduced speech: A cochlear implant simulation study. C.-Y. Peter Chiu (Dept. of Psych. and Dept. of Commun. Sci. and Disord., 401A Dyer Hall, Univ. of Cincinnati, Cincinnati, OH 45229-0376, peter.chiu@uc.edu), Bethany Strange, and James Brockman (Univ. of Cincinnati, Cincinnati, OH 45229-0376)

This study investigated training-related improvement in open-set identification of sentences that were spectrally reduced to four channels using the CIS processing strategy [Shannon *et al.*, J. Acoust. Soc. Am. **104**, 2467–2476 (1998)]. Over the course of 11 training blocks, listeners with normal hearing identified Harvard IEEE [IEEE Trans. Audio Electroacoust. **17**, 225–246 (1969)] sentences with feedback. Half of the listeners listened to a male talker during training; the other half to a female talker. Listeners were tested at different points with blocks of IEEE sentences spoken by the familiar talker, blocks of IEEE sentences by the novel talker, blocks of IEEE sentences by the two talkers randomly intermixed, and blocks of HINT sentences [Nilsson *et al.*, J. Acoust. Soc. Am. **95**, 1085–1099 (1994)]. Preliminary data showed an improvement in sentence identification over the course of training. The data also showed a trend of talker specificity for such improvement, in that sentences produced by the novel speaker were harder to identify. In some instances, random intermixing of talkers led to poorer performance than having just one talker within a testing block. Implications for theories of auditory memory and for practical aspects of training will be discussed. [Work supported by URC, Univ. of Cincinnati.]

5pSC4. Effect of spectrally and temporally altered auditory feedback on the speech of hard of hearing persons. Dragana Barac-Cikoja, Chizuko Tamaki, Maureen O'Brien, and Nancy McIntosh (Dept. of Audiol. and Speech-Lang. Pathol., Gallaudet Univ., Washington, DC)

Nine participants with severe to profound hearing loss read a six-sentence passage under different auditory feedback conditions that included spectrally altered and/or delayed speech. Spectral smearing was implemented by filtering the speech signal into either one, two, or four frequency bands, extracting respective amplitude envelope(s), and amplitude-modulating the corresponding noise band(s). As the number of the filter bands decreased, the spectral information, and therefore the intelligibility of the resulting noise replicas of the speech signal, diminished. The three noise conditions and the unaltered speech were each tested under the simultaneous and four delayed (25, 50, 100, and 200 ms) feedback conditions. Auditory feedback was presented via insert earphones at

the participant's most comfortable listening level. Different measures of duration, intensity, and fundamental frequency ($f0$) were computed on individual sentences, and subjected to statistical analyses. Results indicate a differential effect of the two feedback manipulations on different acoustic parameters. Although speech delays affected duration and intensity measures the most, spectral smearing showed the most robust effect on $f0$ variability. Individual differences were significant. Auditory self-monitoring will be discussed as a function of hearing loss.

5pSC5. Quick SIN and BKB-SIN, two new speech-in-noise tests permitting SNR-50 estimates in 1 to 2 min. Mead C. Killion, Patricia A. Niquette (Etymotic Res., 61 Martin Ln., Elk Grove Village, IL 60007, abonso@aol.com), Lawrence J. Revit (Revitronix, Brownsville, VT 05037), and Margaret W. Skinner (Washington Univ. School of Medicine, St. Louis, MO 63110)

The evolution of two new speech-in-noise tests is described, with particular emphasis on obtaining equivalency across lists. The Quick SIN Test uses IEEE sentences with female talker (L. Braida, personal communication on Harvard Phonetically Balanced Sentences developed at Harvard University during WWII, 2000); the BKB Test uses British Childrens sentences with male talker (Jon Shallop) [J. Bench and J. Bamford (eds.), *Speech-Hearing Tests and the Spoken Language of Hearing-Impaired Children* (Academic, London, 1979)]. In both cases, the Auditec of St. Louis four-talker babble recording was used to provide competing noise, and SNR-50 (signal-to-noise ratio for 50% correct key words in sentences) was estimated using the Tillman–Olsen recommendation [T. W. Tillman and W. O. Olsen, *Modern Developments in Audiology*, 2nd ed. (Academic, New York, 1973), pp. 37–74]. Normal-subject testing as well as hearing-impaired and cochlear-implant subject testing was used to check for equivalency across a range of signal-to-noise ratio performances. The standard deviation for single and multiple lists will be reported.

5pSC6. Case study: Speech production following cochlear implantation in a postlingually deafened adult. Christine H. Shadle (Dept. of Electron. and Computer Sci., Univ. of Southampton, Southampton SO17 1BJ, UK, chs@ecs.soton.ac.uk), Julie Brinton, Carl Verschuur, and Oliver S. Blacklock (Univ. of Southampton, Southampton SO17 1BJ, UK)

While cochlear implants are designed primarily to aid speech perception, some aspects of speech production may improve post-implant in postlingually deafened speakers. Postural parameters such as average SPL may change immediately; phonemic parameters such as place changes allowing [s, ʃ] to be distinguished may continue to change for two years or more post-implant [Matthies *et al.*, J. Acoust. Soc. Am. **96**, 1367–1373 (1994)]. In this paper we describe a female subject who received a cochlear implant following a progressive hearing loss from age 5. She became profoundly deaf at age 34, and was implanted at age 35 with a Nucleus CI24M. She was recorded saying an extensive speech corpus pre-implant and at intervals post-implant. Pre-implant, she had a very weak and creaky voice quality, and /s,z/ were consistently produced as silence, or one of [t, d, θ, ð, tʃ]. Defining parameters to characterize the errors in manner is not straightforward, however. She received some speech therapy post-implant; her voice quality gradually improved, but the /s,z/ errors persisted. Her aided hearing levels and sentence perception results at 18 months are excellent. However, her Speech Perception Audiometry and consonant confusion results demonstrate difficulties and may help explain the continued [s,z] production problems. [Research funded in part by EPSRC.]

5pSC7. Comparing perception and production vowel spaces of cochlear implant users. Richard Wright (Dept. of Linguist., Univ. of Washington, P.O. Box 354340, Seattle, WA 98195-4340), James D. Harnsberger, David B. Pisoni (Speech Res. Lab., Indiana Univ., Bloomington, IN 47401), Mario A. Svirsky, and Adam R. Kaiser (Indiana Univ. School of Medicine, Indianapolis, IN 46202)

This study investigates the relationship between speech production and auditory feedback by comparing vowel production and perception in normal hearing and cochlear implant (CI) using populations. CI users vary greatly in their improvement in speech production after implantation, indicating varying benefit from the CI device. This study examines the degree to which distortions in CI users' speech can be attributed to perceptual distortions. Three predictions are tested in this study: (1) perceptual distortions will correlate with distortions in speech, (2) there will be a tendency to collapse towards the middle of the vowel space, (3) there will be hyperarticulation of vowels analogous to the Lombard effect. Three measures were used in relation to gender and dialect matched normal hearing subjects: (1) overlap between vowel categories, (2) the distortions of the vowel space centers, (3) vowel space compression. Results indicate that there is little correlation between individual CI users' spoken vowel spaces and the range of variability in the perceptual vowel space. However, there is a tendency to collapse towards the middle of the vowel space. This behavior is unlike the normal hearing response to noise in which production vowel categories tend to expand away from the middle of the space.

5pSC8. Speech processing and memory throughout the lifespan. Aimee M. Surprenant (Purdue Univ., West Lafayette, IN 47905)

Research on cognitive aging has converged on the concept of speed of processing to explain certain age-related declines in cognitive capabilities. According to this explanation, older adults perform worse on a variety of tasks because they process information at a slower rate than younger adults. This view, however, does not address the question of why older adults might process information more slowly. This study investigated the hypothesis that noise and/or mild hearing loss can cause difficulty in perceptual processing which, in turn, affects cognitive processing. The extra effort required to identify each stimulus draws resources normally devoted to higher-level processes. Forty normal or near-normal hearing subjects, ranging in age from 30–75 years participated in a variety of hearing and cognitive tasks. The main finding was that performance on a memory for syllables in noise task was predicted by performance on an identification of words in noise task, regardless of the age of the listener. These results suggest that one possible factor underlying slower cognitive processing as we age is more effortful sensory processing. [Work supported by a Gerontology Pilot Grant from Purdue University.]

5pSC9. Predictors of visual enhancement and lipreading ability in older and younger adults. Mitchell S. Sommers and Debra Sawyer (Dept. of Psych., Washington Univ., St. Louis, MO 63130, msommers@artsci.wustl.edu)

The present study was designed to examine the importance of several factors in predicting visual enhancement (auditory-visual compared with auditory-only performance) and lipreading abilities in older and younger adults. Thirteen consonants were presented in an aCa context and participants responded using a closed-set response format. Stimuli were presented in auditory-only, visual-only, and auditory-visual conditions to each participant. Predictor variables included measures of education, vocabulary, verbal working memory, spatial working memory, and processing speed. For both younger and older adults, the two measures of working memory (verbal and spatial) and overall processing speed accounted for a significant percentage of the variance in visual-only performance. None of the predictor variables accounted for significant variance in visual enhancement for young adults. Preliminary results for older participants indicated that speed of processing and verbal working memory were reliable predictors of visual enhancement. Neither group exhibited a significant

relationship between visual-only performance and visual enhancement, suggesting that the ability to integrate across modalities, rather than lipreading ability, is the principal factor contributing to individual differences in visual enhancement. Implications of these findings for improving the speech perception abilities of hearing-impaired older adults are discussed. [Work supported by NIA and the Brookdale Foundation.]

5pSC10. The effects of symmetrical and asymmetrical sensorineural hearing loss on speech perception in noise. Dale A. Ostler (U.S. Army Aeromedial Res. Lab., P.O. Box 620577, Ft. Rucker, AL 36362-0577, dale.ostler@se.amedd.army.mil) and Carl C. Crandell (Univ. of Florida, Gainesville, FL 32611)

The effect of symmetrical and asymmetrical sensorineural hearing loss (SNHL) on the speech reception threshold (SRT) in speech-spectrum-shaped noise was assessed with the hearing in noise test (HINT) using an adaptive procedure. Subjects consisted of 16 listeners with symmetrical SNHL and 16 listeners with asymmetrical SNHL. Speech and noise were presented in a sound field from spatially separated loudspeakers at various azimuths (speech at 0/noise at 180, speech at 45/noise at 225, speech at 315/noise at 135). Results suggest that individuals with mild to moderately sloping asymmetrical SNHL (defined as a 15 dB or greater difference in hearing thresholds between ears at two or more frequencies) have speech-perception ability equivalent to that of individuals with an equal amount of symmetrical SNHL as long as the speech signal is presented to the better ear (BE) or from directly in front of the listener. Individuals with asymmetrical SNHL, however, exhibited more difficulty understanding speech in noise than did individuals with symmetrical SNHL if the speech signal was presented toward the poorer ear (PE). Implications are discussed as well as limitations and future directions of research.

5pSC11. Acoustic analysis of the speech of adults with cochlear implants: A case study. Joanna H. Lowenstein (Dept. of Linguist., Univ. of Chicago, 1010 E. 59th St., Chicago, IL 60637, jhlowens@uchicago.edu)

Most investigations that use post-lingually deafened adults with cochlear implants as their subjects focus on post-implant changes in perception. The primary acoustic research on the speech production of adult CI users consists of seven papers by Joseph Perkell's research group at MIT (1992–1997). The authors argue that the Ineraid analog cochlear implant provides a sufficient degree of auditory feedback to affect vowel placement and duration, sibilant contrast, VOT and syllable duration, and fundamental frequency. Currently, most adults are implanted with multichannel digital implants, which provide a more complex signal and utilize more advanced speech processors. The greater frequency range and increased spectral detail should particularly affect vowel placement, stops, and intonation. This paper presents acoustic analysis of the speech of a post-lingually deafened adult with a Nucleus cochlear implant (pre-implant, 1-, 3-, and 6-month post-activation recordings). Articulatory changes that follow experience with a CI are characterized via acoustic analysis of vowel formants and duration; stop VOT, duration, and burst spectra; and F0 peaks and contours for sentences. Subject's data are analyzed for individual longitudinal trends and implant effects. Subject trends are compared to those found in the MIT studies.

5pSC12. Use of partial stimulus information in spoken word recognition without auditory stimulation. Lorin Lachs and David B. Pisoni (Speech Res. Lab., Indiana Univ., Bloomington, IN 47405)

The identification of isolated words in speech-reading environments is extremely prone to error. Many of these errors are due to the impoverished nature of spoken stimuli when the only perceptual information available is visual; some estimates place the number of visually discriminable segments at just over 25% of the number discriminable in auditory-only environments. Previous research has shown that the confusion of lipreaders is patterned with respect to the perceptual discriminability of phonetic segments in visual-only environments. In addition, other sources of information such as phonotactic, lexical, and semantic constraints can play a role in speech-reading performance. The current study examined the speech-reading responses generated by 200 normal hearing participants observing 300 isolated English words spoken by 10 talkers each. The responses were analyzed to determine if errors were random or patterned in a way that belies the use of partial information and lexical constraints during the process of visual-only spoken word recognition.

5pSC13. Word recognition and phonetic perception by adult cochlear implant users: An examination of /d/–/t/. Paul Iverson (Dept. of Phonet. and Linguist., Univ. College London, 4 Stephenson Way, London NW1 2HE, UK)

Speech perception by cochlear implant users is marked by large individual differences in word recognition accuracy. The aim of the present study was to examine whether these individual differences in speech understanding are correlated with measures of phonetic sensitivity and categorization. Thirty adult post-lingually deafened cochlear implant patients, who were heterogeneous in terms of their implants and processing strategies, were tested in four experiments. Experiments 1 and 2 tested open-set word recognition and forced-choice consonant recognition for words and syllables spoken by ten different talkers. Experiments 3 and 4 tested forced-choice phoneme identification and discrimination along a synthetic /d/–/t/ continuum. The results demonstrated that open-set word recognition accuracy was highly correlated with the location of sensitivity peaks along the /d/–/t/ continuum; word recognition accuracy was higher for cochlear implant users who had sensitivity peaks near the typical /d/–/t/ category boundary location of normal-hearing individuals. In contrast, word recognition accuracy was unrelated to mean levels of discrimination performance or to measures of phoneme identification. The results suggest that at least some aspects of the common phonetic perceptual phenomena found in normal-hearing listeners have a functional role in speech understanding via cochlear implants.

5pSC14. The resolution of spectral peaks by cochlear implant and normal-hearing listeners. Belinda A. Henry and Christopher W. Turner (Dept. of Speech Pathol. and Audiol., Univ. of Iowa, Iowa City, IA 52242, belinda-henry@uiowa.edu)

The ability to resolve spectral peaks is important in the discrimination of speech. This study investigated the differences in spectral resolution abilities between cochlear implant (CI) and normal-hearing (NH) listeners, and among CI listeners having different speech perception abilities. CI listeners using the Cochlear Ltd. CI-24M implant and SPrint processor, and NH listeners participated. The stimuli were rippled-spectrum noise signals, with ripple densities of approximately 1, 1.5, 2, 3, 4, 6, 8, 16, 32, 64, 126, and 252 peaks between 0 and 8 kHz. The stimuli were presented to the CI subjects (CIS) using a 12-channel CIS speech processing strategy via the direct processor input, and to NH subjects via headphones through a 12-channel CI simulation. The highest resolvable ripple density was determined for each listener as the ripple density at which an interchange of peak and trough positions in the rippled spectrum was discriminated at the 70.7% correct level. This efficient method of measuring ripple density resolution indicates that CI listeners have poorer spectral resolution abilities than NH listeners when listening with 12 channels. The relationship between ripple density resolution and vowel recognition in CI listeners will be discussed. [Work supported by NIDCD.]

5pSC15. Perception of dynamic speech sounds by normal hearing listeners and cochlear implant recipients. John W. Hawks (School of Speech Pathol. and Audiol., Kent State Univ., P.O. Box 5190, Kent, OH 44242) and Marios S. Fourakis (Ohio State Univ., Columbus, OH 43210)

In a continuing series of experiments aimed at identifying factors that influence speech perception by cochlear implant (CI) users, we are currently investigating the reception and utilization of relevant perceptual information available for place of articulation of stop consonants in synthetic stop-vowel sequences. Normal hearing subjects and adults fitted with clarion and nucleus cochlear implant (CI) devices provided identification responses for synthetic CVs which varied in (1) initial onset frequencies of $F2$ and $F3$, (2) duration of frequency transition from onset to steady-state target, and (3) steady-state target $F2$ and $F3$ frequencies of the following vowel. While resembling stop-vowel sequences, no burst or aspiration parameters were included, thus forcing listeners to use only transition information as the primary cue to place of articulation. The results demonstrate the potential diagnostic value afforded from mapping dynamic speech sounds in an acoustic space using synthetic tokens.

5pSC16. On recognizing spectrally reduced speech by multiple talkers: A cochlear implant simulation study. Kristen Fishbeck (Dept. of Psych., Univ. of Cincinnati, Cincinnati, OH 45229-0376, kfishbec@hotmail.com) and C.-Y. Peter Chiu (Univ. of Cincinnati, Cincinnati, OH 45229-0376)

In a previous study [Loizou *et al.,* J. Acoust. Soc. Am. **106**, 2097–2103 (1999)], normal-hearing listeners identified sentences from the multitalker sentence set TIMIT that were spectrally reduced using the CIS processing strategy [Shannon *et al.,* J. Acoust. Soc. Am. **104**, 2467–2476 (1998)]. It was reported that listeners could achieve open-set sentence recognition (90%) with as few as five channels, with asymptotic performance achieved with eight channels. Such results were compared with another study [Dorman *et al.,* J. Acoust. Soc. Am. **102**, 2403–2411 (1997)] using different sentences from the single-talker HINT set, where asymptotic performance was achieved with four channels. The present study directly compares identification for spectrally reduced TIMIT sentences spoken by a male talker, by a female talker, and by multiple talkers. Separate groups of subjects with normal hearing served in these three conditions. The design therefore controlled potential differences in the lexical, syntactic, and semantic aspects of the materials in TIMIT versus HINT that could have accounted for the above findings. Results will be discussed in terms of theories of auditory memory and speech perception.

5pSC17. Final consonant discrimination in children with phonological disorders. Jan Edwards, Robert A. Fox, and Bridgett Isermann (Dept. of Speech and Hearing Sci., Ohio State Univ., 1070 Carmack Rd., Columbus, OH 43210-1002)

A number of researchers have shown that children with phonological disorders, relative to typically developing age peers, have difficulty discriminating the particular contrasts that they do not produce correctly. It is less clear whether children with phonological disorders, as a group, have greater difficulty with more general measures of speech perception than their peers. In this study, 35 preschool-aged children with phonological disorders and a group of 35 typically developing controls (matched for age, gender, and nonverbal IQ) were asked to discriminate between two CVC words which differed only in the identity of the final consonant. The stimuli were digitized natural speech including both whole word and backward-gated versions. The children with phonological disorders performed significantly more poorly than their age peers. Significant correlations (with age partialed out) were observed between discrimination scores (d-primes) and raw scores on both the Goldman–Fristoe test of articulation and the PPVT-III. Since the cues to final consonants in CVC words (VC transition and release burst) are not integrated in time and the burst is not reliably present, it is not surprising that representations of final consonants for children with phonological disorders are particularly vulnerable to diminished redundancy in the acoustic signal. [Work supported by NIDCD.]